An exponential regression model to estimate daily milk yields

X.-L. Wu[1,2], G. Wiggans[1], H. D. Norman[1], A. M. Miles[3], C. Van Tassell[3], R. L. Baldwin VI[3], J. Burchard[1] and Joao Durr[1]

[1] Council on Dairy Cattle Breeding, Bowie, MD 20716, USA.

[2] Department of Animal and Dairy Sciences, University of Wisconsin, Madison, WI 53706, USA.

[3] USDA, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705-2350, USA.

Abstract

Accurate milking data are essential for herd management and genetic improvement in dairy cattle. Cows are typically milked two or more times on a test day, but not all these milkings are sampled and weighed. This practice started to supplement the standard supervised twice-daily monthly testing scheme in the 1960s, motivated by lowering the costs to the dairyman. The initial approach estimated a test-day yield by doubling the morning (AM) or evening (PM) yield in the AM-PM milking plans, assuming equal AM and PM milking intervals. However, AM and PM milking intervals can vary, and milk secretion rates may change between day and night. Statistical methods have been proposed afterwards, focusing on various forms of correction factors. Additive correction factors (ACF) are evaluated by the average differences between AM and PM milk yield for different milking interval classes (MIC), coupled with other categorical variables. Multiplicative correction factors (MCF) are ratios of daily yield to yield from single milkings, with varied statistical interpretations. MCF are now commonly used, but they have biological and statistical challenges. An exponential regression model was proposed as an alternative model for estimating daily milk yield, which was analogous to an exponential growth function with a partial yield as the initial state and the change of rate tuned by a linear function of milking interval. The results showed that the existing MCF model performed similarly. They all had substantially lower MSE and, therefore, greater accuracies than the initial approach of doubling AM or PM milk yields as the test-day milk yields. Two times AM or PM milk yields as the test-day milk yields were a reasonable approximation with equal AM and PM milking intervals but were subject to large errors with unequal AM and PM milking intervals. For computing MCF, discretizing the milking interval into categorical MIC led to a loss of accuracy. The exponential regression models had the smallest MAE and the greatest accuracies, representing a promising alternative for estimating daily milk yields. The statistical methods were explicitly described to estimate daily milk yield in AM and PM milking plans. Still, the principles generally apply to cows milked more than twice daily.

Key words: dairy cattle, milk yield, exponential function, milking interval, yield correction factors

Introduction

Accurate milking data are essential for herd management and genetic improvement in dairy cattle. Cost-effective milking plans started to supplement the standard supervised twice-daily monthly testing scheme in the 1960s, motivated by reducing the visits by a National Dairy Herd Information Association (**DHIA**) supervisor and thus lowering the costs to the dairyman (Putnam and Gilmore, 1968). Cows are typically milked two or more times on a test day, but not all these milkings are sampled and weighed. The initial AM-PM milking plan alternately sampled the morning (**AM**) or evening (**PM**) milking on test day throughout the lactation. Daily yield (milk, fat, and protein) was estimated by two times the yield from single milkings on each test day, assuming equal AM and PM milking intervals (Porzio, 1953). However, AM and PM milking intervals are different, and milk secretion rates can vary between days and nights.

Various methods have been proposed to adjust the estimated daily yields, mainly focused on correction factors to account for varied milking intervals. There are two broad categories of correction factors, additive (**ACF**) and multiplicative (**MCF**). In AM-PM plans, ACF provide additive adjustments to two times AM or PM milk yield as the estimated daily yield. ACF are evaluated by the population averages of the differences between the AM and PM milk yield, computed explicitly for each milking interval class (MIC) and other categorical variables (Everett and Wadell, 1970a, 1970b). Everett and Wadell (1970a) have shown that the difference between AM and PM yields is a function of milking interval and days in milk (**DIM**). Significant variables affecting such differences vary with cattle breeds, including months of lactation, herd production level, age classes, MIC, and their interactions (Everett and Wadell, 1970b). An ACF model is statistically equivalent to a regression model of daily yield on categorical regressor variables, and a continuous variable for AM or PM yield with a fixed regression coefficient of "2.0". Similarly, a linear regression (**LR**) model can be implemented as an ACF model with the regression coefficient for AM or PM yield estimated from the data; ACF are computed for discretized MIC. LR models can be defined with varying complexity (Liu et al., 2000).

MCF (also referred to as ratio factors) are ratios of daily yield to yield from single milkings computed for various MIC (e.g., Shook et al., 1980; DeLorenzo and Wiggans, 1986; Wiggans, 1986). Shook et al. (1980) empirically computed MCF from bulk AM and PM yields, subject to fitting a quadratic smoothing function for obtaining smoothed MCF. DeLorenzo and Wiggans (1986) proposed a linear regression model without intercept to derive MCF for cows milked twice daily, assuming heterogeneous means and variances, and fitted separate linear models for different MIC. They proposed linear smoothing by regressing the reciprocals of computed AM or PM factors on milking interval time to obtain smoothed MCF. Wiggans (1986) proposed deriving yield factors for cows milked three times a day through regressing AM or PM proportion of daily yield on milk interval. Arguably, the Wiggans (1986) model also applies to cows milked more than three times and twice daily. In the latter case, the model is subject to the violation of linearity with a longer milking interval (Schmidt, 1960).

MCF models are statistically challenged by the well-known "ratio problem" because each model has a ratio variable (i.e., AM or PM proportion of daily yield) as the dependent variable in the data density (Wiggans, 1986) or the smoothing functions (Shook et al., 1980; DeLorenzo and Wiggans, 1986). The consequences included possible biases in two aspects: omitted variable bias and measurement error bias (Lien et al., 2017). The former happens because the main model effects are missing if the model is re-arranged by multiplying both sides of the equation by the denominator variable. The latter occurs when there are measurement errors in the denominator variable of the response. Besides that, the MCF models postulated a rational function between daily milky yield and milking, in which the numerator is one, and the denominator is a linear function (DeLorenzo and Wiggans, 1986; Wiggans, 1986) or a quadratic function (Shook et al., 1980) of milking interval. Yet, no biological evidence has been available to support a rational function as a daily milk curve.

Early studies showed that daily milk (including fat and solid-not-fat) yield curves were not linear with intervals beyond 12 hr. (Ragsdale et al., 1924; Bailey et al., 1955; Elliott and Brumby, 1955; Schmidt,1960). For example, Brody (1945) showed milk yields and fat percentages for milking intervals between 1 and 36 hr., which empirically resembled an exponential function. Klopcic et al. (2012) proposed using a modified Michaelis-Menten function to predict the daily milk yields of dairy cows in relation to the interval of milkings. The modified Michaelis-Menten function (Klopcic et al., 2012) is a modified exponential function where the base is one plus the yield for an interval of 720 min (i.e., 12 hr.) and a non-linear function of milking time is the exponent. Biologically, the exponential behavior for milk production was assumed to result from cell degradation and milk in the udder (Neal and Thornley, 1983). In the present study, we proposed an exponential regression model for estimating daily milk yields. MCF can be derived from the expression regression model as well. The features of this new model and its performance for estimating daily milk yields were illustrated using simulated datasets, compared to the existing MCF models, with the approach of doubling AM or PM yield as the benchmark method.

Methods and Methods

Exponential regression model

Let x_{ij} be a partial yield for cow *i* from single milking *j*, for $j = 1$ (AM) or 2 (PM), and y_i be the corresponding test-day milk yield. First, we assume that the logarithm of daily to single milk yield ratio is a linear function of milking interval time:

$$log\left(\frac{y_i}{x_{ij}}\right) = \alpha_j + \beta t_{ij} + \epsilon_{ij} \quad (1)$$

where $\frac{y_i}{x_{ij}}$ is a ratio of daily yield to single milking (AM or PM) yield, α_j is the intercept pertaining to milking *j*, β is the regression coefficient, and ϵ_{ij} is the error. In the above model, for the sake of simplicity, we ignore other variables such as days in milking but note that they may be relevant in real applications.

With some re-arrangements, the above equation becomes:

$$log(y_i) = \alpha_j + \beta t_{ij} + blog(x_{ij}) + \epsilon_{ij} \quad (2)$$

Here, $log(y_i)$ is the response variable, $log(x_{ij})$ and t_{ij} are the dependent variables (i.e., main effects), $b = 1$ is a constant regression coefficient for log. For the model development, we relax the restriction for $b = 1$ in (2) and allow it to be estimated from the data. Now, taking the exponential on both sides of equation (2) gives:

$$y_i = x_{ij}^b e^{(\alpha_j + \beta t_{ij} + \epsilon_{ij})} \quad (3)$$

The above is recognized as an exponential regression model. This exponential regression model is non-linear, but its model parameters can be conveniently estimated by fitting the data to the linear logarithm model (2). Then, daily milk yield is calculated given the model parameter estimates ($\hat{b}$, $\hat{\alpha}_j$, and $\hat{\beta}$) and the observed partial (AM or PM) yield and milking interval time.

$$\hat{y}_i = x_{ij}^{\hat{b}} e^{(\hat{\alpha}_j + \hat{\beta} t_{ij})} \quad (4)$$

By noting $e \approx 2.718$, we show that the exponential function is analogous to an exponential growth function:

$$y = y_0(1 + 1.718)^{t^*} \quad (5)$$

where $y_0 = x^b$ is the initial value, $r = 1.718$ is the rate of change, tuned by a time function ($t^* = \alpha_j + \beta t_{ij} + \epsilon_{ij}$) as a linear function of milking interval.

This exponential regression model can be implemented as an ACF or MCF model. The former applies to the logarithm linear regression model (2), whereas the latter applies to the exponential regression model (3). Consider the MCF model based on equation (3). Taking expected values on both sides of equation (3) leads back to equation (2). Then, we applied the second order Taylor approximation (Spivak, 1994) by noting that $E(log(z)) \approx log(E(z)) - \frac{V(z)}{2E(z)^2}$, where z is a random variable. Hence, we have:

$$log(E(y_i)) = \alpha_j + \beta\left(E(t_{ij})\right) + blog\left(E(x_{ij})\right) + \left(\frac{V(y_i)}{2E(y_i)^2} - b\frac{V(x_{ij})}{2E(x_{ij})^2}\right) \quad (6)$$

Next, taking the exponential on both sides of equation (6), with some re-arrangements, gives:

$$E(y_i) = \rho E(x_{ij})^b e^{\{\alpha_j + \beta E(t_{ij})\}} \quad (7)$$

where $\rho = e^{\frac{1}{2}\left(V(y_i)E(y_i)^{-2} - bV(x_{ij})E(x_{ij})^{-2}\right)}$. Following Shook et al. (1980) and DeLorenzo and Wiggans (1986), MCF are defined as the ratios of daily to single milk yield. Thus, MCF are the ratio $E(y_i)$ over $E(x_{ij})$, evaluated by taking the expected values of (5) locally for each MIC, say k. That is,

$$F_j^{(k)} = \frac{E\left(y_i^{(k)}\right)}{E\left(x_{ij}^{(k)}\right)}$$

$$= \rho_j^{(k)} E\left(x_{ij}^{(k)}\right)^{b-1} e^{\alpha_j + \beta \bar{t}_j^{(k)}} \qquad (8)$$

where $\rho_j^{(k)} = e^{\frac{1}{2}\left(V\left(y_i^{(k)}\right)E\left(y_i^{(k)}\right)^{-2} - bV\left(x_{ij}^{(k)}\right)E\left(x_{ij}^{(k)}\right)^{-2}\right)}$, and $E\left(y_{ij}^{(k)}\right) = \bar{y}^{(k)}$ and $E\left(x_{ij}^{(k)}\right) = \bar{x}_j^{(k)}$ are the corresponding means for daily yield and AM (or PM) yield, respectively.

Confined to MIC k, we show the following relationship holds, assuming $E\left(y_i^{(k)}\right) = E\left(x_{i1}^{(k)}\right) + E\left(x_{i2}^{(k)}\right)$:

$$F_{1k}^{-1} + F_{2k}^{-1} = 1 \qquad (9)$$

The above brings convenience to computing MCF. MCF can be computed for AM and PM milkings jointly or separately. In the latter case, for example, given the computed PM MCF (F_{2k}), AM MCF can be obtained indirectly as follows:

$$F_{1k} = \left(1 - F_{2k}^{-1}\right)^{-1} = \frac{F_{2k}}{F_{2k}-1} \qquad (10)$$

A simulation study

Daily milk yields were simulated for 3,000 cows based on a modified Michael-Menten function (Klopcic et al., 2012), where the values for and k were simulated from truncated normal (TN) distributions: $\sim TN$ and $k \sim TN$. AM milking intervals were simulated following a truncated normal distribution with a mean equaling 12 hours and a standard deviation of 1.12 hours. PM milking intervals for the same cows were 24 hr. minus the AM milking intervals. As an example, the simulated daily milk yield curve (mean and 95% confidential intervals) and frequency distribution of AM and PM milking intervals from a replicate are shown in Figure 1. Approximately 98.6% of the cows had AM (PM) milking intervals between 9 hr. and 15 hr. The mean for AM (PM) daily milk yield was 12.10 kg, and the 95% was from 8.38 kg to 16.27 kg.

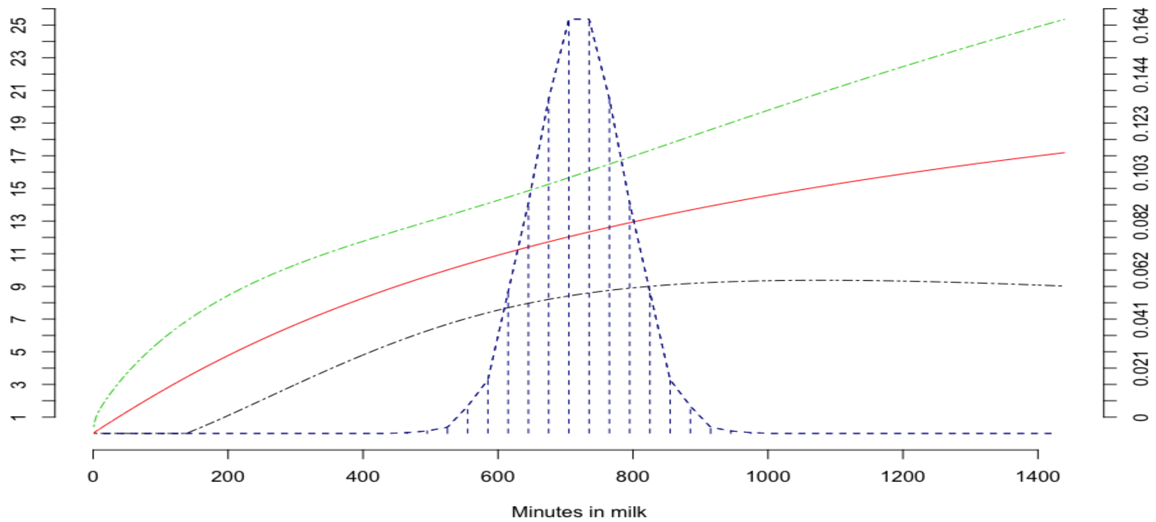

Figure 1. Mean (solid red line) and 95% confidential intervals (dotted green and grey lines) of simulated test-day milk curves and frequency distribution (dotted blue lines) of morning and evening milking intervals (min.)

Five MCF models were compared, which are: M2 = MCF model according Shook et al. (1980); M3 = MCF model according to DeLorenzo and Wiggans (1986); M4 = Wiggans (1986); M5A = exponential regression; M5B = MCF model derived from the exponential regression M5A. The approach of doubling AM or PM yield was included as a benchmark method (M1). These models were compared based on the computed MCF for M1, M2, M3, M4, and M5B and the mean absolute errors (MAE) for M1, M2, M3, M4, M5A, and M5B. MAE were evaluated from ten-fold cross-validation for all the models, each replicated $M = 30$ times. Briefly, the dataset was randomly split into ten equal subsets in each replicate. Nine subsets were pooled and used for training, and one subset was used for validation. The process was rotated ten times per replicate, with each subset used for validation only once. The ten testing sets were pooled to evaluate MAE per replicate. We included only MCF models in this simulation study, because they are the de facto standard methods for estimating daily yields (Liu et al., 2000).

Results and Discussion

All the MCF methods were roughly comparable MCF within MIC, except that M1 (doubling AM or PM yield as the estimated daily milk yield) had a fixed MCF of “2.0” for all MIC (Figure 2). All the computed MCF approximately equaled 2.0 when AM and PM milking intervals were both 12 hours. Hence, these results suggest that assuming a fixed multiplier of 2.0 was valid with equal AM and PM milking intervals. However, this assumption did not hold with unequal AM and PM milking intervals because computed MCF deviated from 2.0 with unequal AM and PM milking intervals. The larger the difference between AM and PM milkings, the more the computed MCF deviated from 2.0. The computed MCF were greater than 2.0 when AM milking intervals were less than 12 hours, and they were less than 2.0 when AM milking intervals were greater than 12 hours. An opposite trend was observed between MCF and PM milking intervals. For these methods (except M1), the differences in MCF between the methods became apparent when AM (PM) milking interval was less 10 hours or more than 14 hours.

Accuracies of these methods were measured by MAE of estimated test-day milk yields. As expected, all the methods had comparably low MAE with equal (12-12 hours) AM and PM milking intervals. Again, doubling AM or PM yield as the test-day milk yield with equal AM and PM milking intervals. Nevertheless, MAE increased substantially with unequal milking intervals (Figure 3). Overall, the more the AM (PM) milking interval deviated from 12 hours, the large MAE it generated. On average, MAE was 1.074 for M1, 0.424 for M2, 0.423 for M3, 0.425 for M4, 0.389 for M5A, and 0.418 for M5B. The exponential regression models (M5A and M5B) had the smallest MAE in all these methods. M1 had more than doubled the MAE compared to the other methods. Hence, two times AM or PM milk yield provided an approximate estimate of daily milk yield with equal AM and PM milking intervals but was subject to a large error with unequal AM and PM milking intervals. The use of MCF effectively reduced MAE by 60.4% to 61.1%. The exponential regression model reduced MAE by 63.8% compared to M0 (doubling AM or PM milk yields).

We note that model M5B had a larger average MAE than model M5A. The former estimated test-day milk yields directly through the estimated model parameter values. In contrast, the latter model computed MCF for discretized MIC and then estimated test-day milk yields through computed MCF. This result suggested that discretizing milking intervals into categorical MIC led to a loss of accuracy. This phenomenon generally holds for ACF or MCF models. For example, the MCF model can be implemented as a linear regression of AM or PM proportional daily milk yield on a continuous variable for milking interval time without the need to compute MCF. Then, daily milk yields can be estimated directly given the model parameter values. These directly estimated daily milk yields also had slightly higher accuracies than those computed directly through the computed MCF (data not presented).

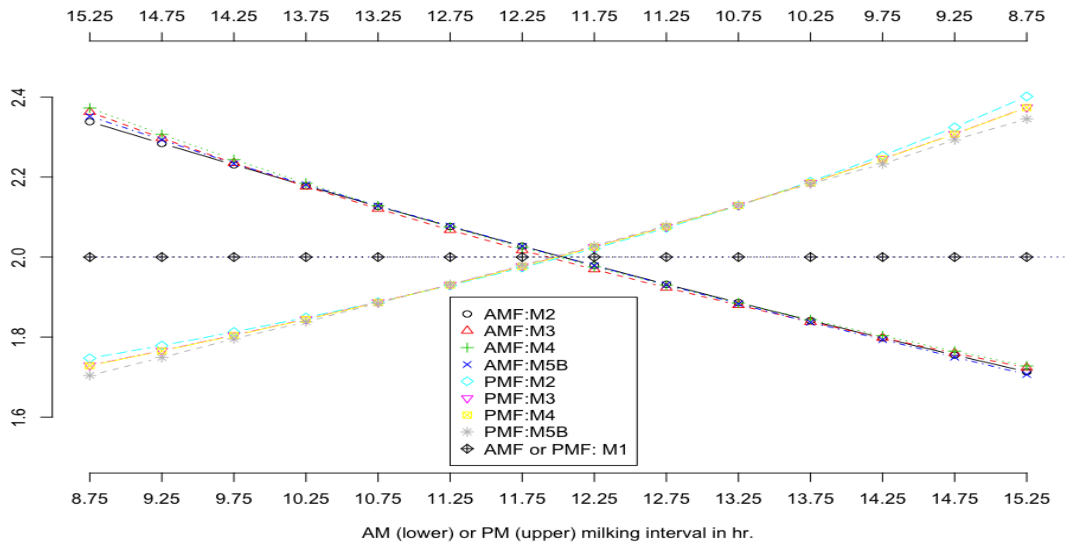

Figure 2. Comparing multiplicative correction factors computed using different methods. M1 = doubling AM or PM milk yield; M2 = MCF model according to Shook et al. (1980); M3 = MCF model according to DeLorenzo and Wiggans (1986); M4 = MCF model according to Wiggans (1986); M5B = MCF model derived from the exponential regression model M5A.

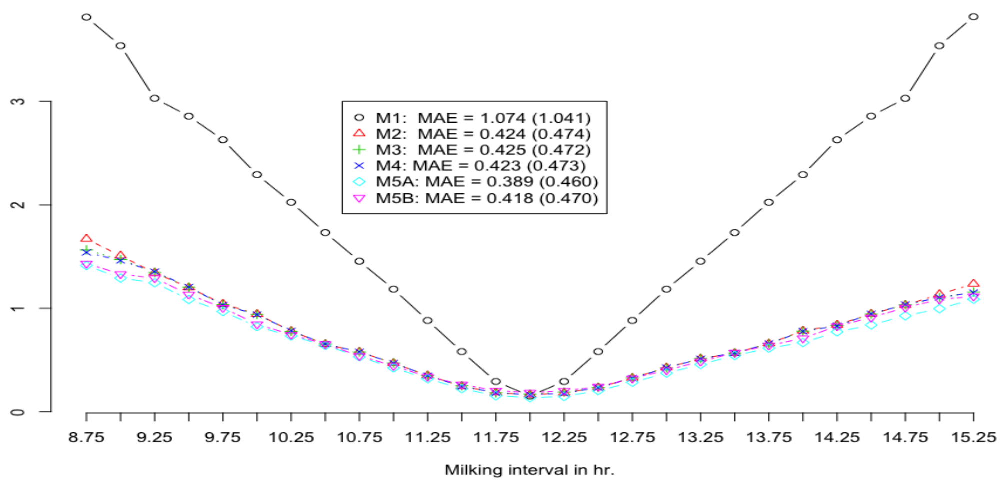

Figure 3. Comparing mean absolute errors (MAE) of estimated test-day milk yields using different methods. M1 = doubling AM or PM milk yield; M2 = MCF according to Shook et al. (1980); M3B = MCF according to DeLorenzo and Wiggans (1986); M4 = MCF according to Wiggans (1986); M5A = exponential regression model; M5B = MCF model derived from the exponential regression model M5A.

Conclusions

The performance of the existing MCF models and the newly proposed exponential regression models were evaluated by ten-fold cross-validations using simulated milking records, compared to the initial approach of taking two times AM or PM milk yields as the test-day milk yields. The results showed that the existing MCF model performed similarly, with substantially lower MSE and, therefore, greater accuracies over the initial approach of doubling AM or PM milk yields as the test-day milk yields. Two times AM or PM milk yields as the test-day milk yields were approximately taken with equal AM and PM milking intervals but were subject to large errors with unequal AM and PM milking intervals. Discretizing the milking interval into categorical MIC when computing MCF led to a loss of accuracy. The exponential regression models had the smallest MAE and the greatest accuracies, representing a promising alternative for estimating daily milk yields. Finally, the statistical methods were explicitly described to estimate daily milk yield in AM and PM milking plans. Still, the principles generally apply to cows milked more than twice daily.

References

Bailey, G. L., Clough, P. A. and Dodd, F. H. 1955. The rate of secretion of milk and fat. J. Dairy Res. 22, 22.

Brody, S. 1945. Bioenergetics and Growth. Reinhold Publishing Corporation, New York, NY.

DeLorenzo M. A. and Wiggans, G.R. 1986. Factors for estimating daily yield of milk, fat, and protein from a single milking for herds milked twice a day. J. Dairy Sci. 69, 2386–2394.

Elliott, G. M., and Brumby, P. J. 1955. Rate of milk secretion with increasing interval between milking. Nature, 176, 350.

Everett, R. W. and Wadell, L.H. 1970a. The relationship between milking intervals and individual milk weights. J. Dairy Sci. 53, 548.

Everett, R. W. and WadeH, L.H. 1970b. Sources of variation affecting the difference between morning and evening daily milk production. J. Dairy Sci. 53, 1424.

Klopcic, M., Koops, W. J. and Kuipers, A. 2012. A mathematical function to predict daily milk yield of dairy cows in relation to the interval between milkings. J. Dairy Sci. 96, 6080-90.

Lien, D., Hu, Y. and Liu, L. 2017. A note on using ratio variables in regression analysis. Econ. Lett. 150, 114–117.

Liu, Z. Reents, R., Reinhardt, F. and Kuwan, K. 2000. Approaches to estimating daily yield from single milk testing schemes and use of a.m.-p.m. records in test-day model genetic evaluation in dairy cattle. J. Dairy Sci. 83, 2672-82.

Neal, H. D. St. C. and Thornley, J. H. M. 1983. The lactation curve in cattle: A mathematical model of the mammary gland. J. Agric. Sci. 101:389–400.

Porzio, G. 1953. A new method of milk recording. An. Br. Abstr. 21:344.

Putnam, D. N. and Gilmore, H. C. 1968. The evaluation of an alternate AM-PM monthly testing pIan and its application for use in the DHIA program. J. Dairy Sci. 51, 985.

Ragsdale, A. C., Turner, C. W. and Prody, S. 1924. The Rate of milk Secretion as affected by an accumulation of milk in the mammary gland. J. Dairy Sci., 7, 249.

Schmidt, G. H. 1960. Effect of milking intervals on rate of milk and fat secretion. J. Dairy Sci. 43, 213.

Shook, G., Jensen, E. L. and Dickinson, F. N. 1980. Factors for estimating sample-day yield in am-pm sampling plans. Dairy Herd Improvement Letter, 56, 25-30.

Spivak, M, 1994. Calculus (3rd ed.). Houston, TX: Publish or Perish, p. 383.

Wiggans, G. R. 1986. Estimating daily yields of cows milked three times a day. J. Dairy Sci. 69, 2935-2940.