REFEREED ARTICLE

DOI: 10.5836/ijam/2015-03-132

Different methods to forecast milk delivery to dairy: a comparison for forecasting

BJØRN GUNNAR HANSEN[1]

ABSTRACT

To estimate future sales and to ensure customer deliverability, the dairy industry needs reliable forecasts for milk delivery from the farmers. In light of the shortage of milk in Norway in fall 2011, the dairy industry recognized that it needed better tools for forecasting milk delivery. Therefore I developed models which can help the industry avoiding similar situations in the future. I analysed the monthly milk deliveries to Norwegian dairy companies from January 2001 to December 2010 and fitted two time series models. I tested a multiplicative Holt Winters' exponential smoothing model (HW) and a multiplicative seasonal autoregressive integrated moving average model (SARIMA) for forecasting monthly milk delivery. The two time series models were compared with a model based on expert opinions, and a model based on historic monthly quantities. The test showed that a combination of the Expert model and the two time series models give reliable forecasts for a period of up to two years.

KEYWORDS: dairy; milk production; time series; Holt Winters' exponential method; SARIMA

1. Introduction

Organizations rely extensively on forecasts in making strategic decisions, and forecasting ability appears to be a distinctive organizational capability (Makadok and Walker, 2000). Thus forecast accuracy is essential to a firms success and performance (Barney, 1986; Makadok and Walker, 2000). Businesses often make decisions under acute shortages of information. Given these constraints, managers estimate future sales volume to make budgets, predict operating expenses, cash flows, pricing, advertising outlays, etc. Actual turnover and profitability depend on the accuracy of these decisions. Norway made the news headlines across the world in fall 2011 because of the so-called butter crisis. The inland milk production did not meet the total demand for dairy products, particularly for milk fat. One reason for this was the low-carbohydrate diets which became popular in 2011. This increased the demand for milk fat at a moment of time where the inland milk production was declining due to a bad roughage harvest. In practice the dairy companies had too little milk to produce enough butter before Christmas 2011, and therefore Norway had to import 1922 tons of butter in December 2011 and January 2012. For decades Norway had had a supply surplus as compared to inland demand. Therefore the shortage of milk fat came as a total surprise to the whole dairy industry. To increase inland production farmers were allowed to produce over quota in the quota year 2011/ 2012. The lesson learned from the butter crisis was that the dairy industry needed better tools for forecasting milk production, and this motivated the study. The present study aimed at improving the prediction of milk supply from Norwegian dairy farmers to the only two Norwegian dairy companies. In 2013 the 10,700 Norwegian dairy farmers produced 1,525 million litres of milk (TINE, 2013). The dairy companies have accurate records only of historical data of monthly milk supply from individual members over the years. These data represent a good starting point for developing time series models. This paper tests different time series models for forecasting monthly milk delivery to dairy and compare them with the traditional forecasting model, which I denote the Expert model. The paper proceeds as follows. First I briefly discuss different ways of forecasting milk delivery before I present historic data on monthly milk delivery and the different models I will test. In the result section I show how the different forecasting models perform. Finally I discuss the results and conclude.

Forecasting milk delivery

Forecasting can be done in different ways. Qualitative forecasting techniques rely on experience that has not been captured in the form of hard data, and can e.g. rest on expert opinions. Quantitative techniques can be divided in causal models and time series models. Causal models are based on finding a cause and effect relationship, e.g. the milk yield per cow and the total milk

Original submitted May 2014; revision received September 2014; accepted November 2014.
[1] TINE SA, 1432 Ås, Norway. bjorn.gunnar.hansen@tine.no.

delivery. It is important that the cause variable is a leading indicator, i.e. that it can be measured in advance of the production it is assumed to cause. The Expert model represents a mixture of qualitative and quantitative techniques, and I return to a more detailed description below. Time series models attempt to identify patterns that have been present in the past and assume they will continue in the future. A time series is a set of observations measured at successive points in time or over successive periods. We typically search for three major components in past milk production: An average, a trend and seasonal fluctuations. A time series that has no trend is called stationary; if a trend is present the time series is non-stationary. The seasonal component refers to a cyclical pattern that repeats over time. Forecasting methods based on time series proceed in three separate steps as follows:

1. Use past data (a training set) to estimate the parameters of the model.
2. Use estimated parameters to determine how well the time series model would have done in predicting past milk production.
3. Use estimated parameters to forecast production in the future.

Time series models are widely applied to forecast milk production and autoregressive integrated moving average (ARIMA) models (Box and Jenkins, 1970) are perhaps the most common. Sataya *et al.* (2007) tested several time-series models and concluded that ARIMA-models gave the best forecasts for milk production in India. Sankar and Prabakaran (2012) found that the most appropriate model for forecasting milk production in Tamilnadu in India was an ARIMA (1,1,0) model. A limitation with ARIMA models is that they do not take seasonal fluctuations into account. In many countries milk production is characterized by seasonal fluctuations within years, and relatively stable patterns between years (See e.g. IFCN, 2012). In such time series it is fair to assume constant variance of the disturbance term. Therefore seasonal autoregressive integrated moving average (SARIMA) models (Shumway and Stoffer, 2010) are well suited for time series of milk production. SARIMA models, which are similar to ARIMA models except that they take seasonal effects into account, are known as flexible tools for the analysis of time series. There are few applications of SARIMA models in dairy farming. An exception is Akter and Rahman (2010). They used a dataset from England and compared a SARIMA model with a Holt Winters' exponential smoothing model (Holt, 1959; Winters, 1960; Chatfield and Yar, 1988) together with several less advanced smoothing models. According to their findings the Holt-Winters' exponential smoothing technique, which I will denote the HW model, and the SARIMA model were the most accurate, and generated forecasts with errors less than 3 percent. They also concluded that forecasts for periods of more than a year could be used with caution. A weakness of their study was that they only had data for eight years, which gave a small training set to fit their models. Such a short period leads to relatively high errors and makes it difficult to generalize, as commented on by Akter and Rahman (2010). I support their conclusion that the question of how long we can forecast beyond the sampling period can be more precisely investigated when a longer data series is available. This study builds on the study of Akter and Rahman (2010) by applying the HW model and the SARIMA model as the preferred time series model candidates. However, I complement their study in two ways. First I follow their recommendations and apply a longer time series (13 years), which makes the findings more robust. Second, I compare the HW model and the SARIMA model with a model based on expert opinions. In addition I use the milk delivery each month in the last year of the training set as a benchmark to evaluate the other models. The rest of the paper is organized as follows: First I present the dataset. Then I present and test the forecasting models and the two criteria I use to evaluate them. Finally I compare the forecasts from the different models, discuss the results and conclude.

2. Materials and methods

Both the Norwegian dairy companies report how much milk they collect from the farmers each month to the authorities, and I use these figures. I apply these monthly milk delivery data from January 2001 to December 2013 in my analysis, and divide it in two datasets. The time series from January 2001 to December 2010 is the training set, which I use to fit the time series models. The time series from January 2010 to December 2013 is the test set, which I use to make forecasts. In Figure 1 shows the whole dataset.

From Figure 1 we can see a relatively stable level over the years, but with large repetitive seasonal fluctuations. The volume peaks during late autumn and winter and reaches the bottom level in summer. The overall picture is that there is no clear trend in the data. However, if we look more closely at Figure 1 we notice an increasing trend until 2008, then a decreasing trend from 2008, and finally an increasing trend from 2012 on. The volume peaks in winter 2008 due to the change of the quota year from January 1 to March 1. We also notice an increasing variance to the right in the figure. To simplify the interpretation of the time series smoothing was introduced. Kernel smoothing is a moving average smoother that uses a weight function, or kernel, to average the observations. I apply the Nadaraya-Watson kernel weighted average (Watson, 1966) which can be

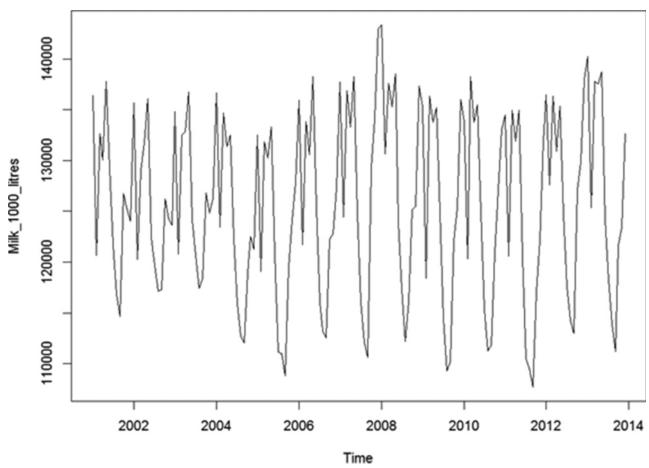

Figure 1: Monthly milk delivery in 1000 litres in Norway from January 2001 to December 2013

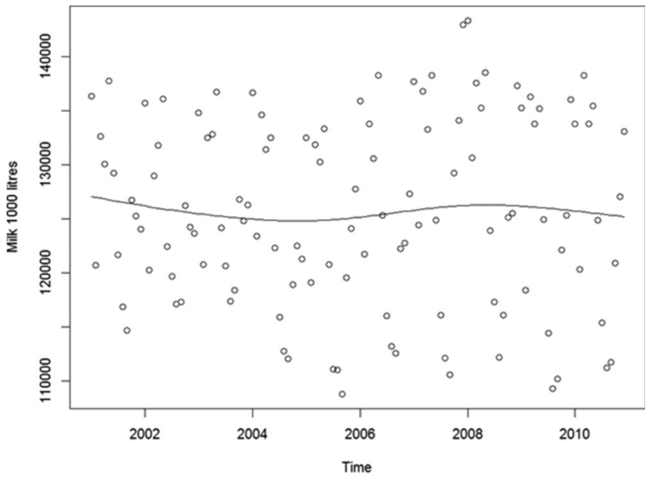

Figure 2: Monthly milk deliveries in 1000 litres in Norway from January 2001 to December 2010, smoothed with the Nadaraya-Watson kernel weighted average

introduced in R by the 'ksmooth' function. Figure 2 shows a smoothed curve of the training set where the level of smoothing is based on cross-validation, the simplest and most widely used method for estimating prediction error (See e.g. Hastie *et al*., 2009).

Smoothing the time series eases the interpretation. We notice a decreasing trend from 2001 to 2005, then an increasing trend until 2008, and finally a decreasing trend from 2008 to 2010. Decreasing milk production combined with increasing popularity of diets low in carbohydrate can explain the shortage in milk fat in Norway in 2011.

To explore the dominant seasonal components in the time series in more detail I apply the periodogram (See e.g. Shumway and Stoffer, 2010) which is given by:

$$I(\omega_j) = \Sigma_{h=-(n-1)}^{n-1}\gamma(h)e^{-2\pi i\omega jh} \quad (1)$$

Here $\gamma(h)$ is the auto covariance function, h is the number of time lags and the frequency is given by $\omega_j = j/n$ for the number of cycles j in n time points, $j=0,1,,,n-1$. Thus ω_j is the frequency measured in cycles per unit time. For $\omega = 1$ the time series makes one cycle per time unit or month. One cycle every twelve months corresponds to 0,083 cycles per monthly observation. In Figure 3 I present the raw periodogram for the training set.

In Figure 3 the frequency axis is labelled in multiples of 1/12. We notice the dominant spectrum occurring at $\omega = 1/12$, or one cycle per year. This corresponds to the regular seasonal pattern in Figure 1.

Description of the models

In this section I present the different models which I will fit to the training set and compare in the result section. To analyse the time series I apply the statistical package R (http://CRAN.R-project.org/). First I briefly comment on the smoothing technique used in Figure 2. Smoothing techniques use observations close to the target point to fit a simple model to the dataset in such a way that the resulting estimated function is smooth. This is achieved by a weighting function or kernel, which assigns weights to observations close to the target point. The Nadaraya-Watson kernel weighted average assigns weights that die

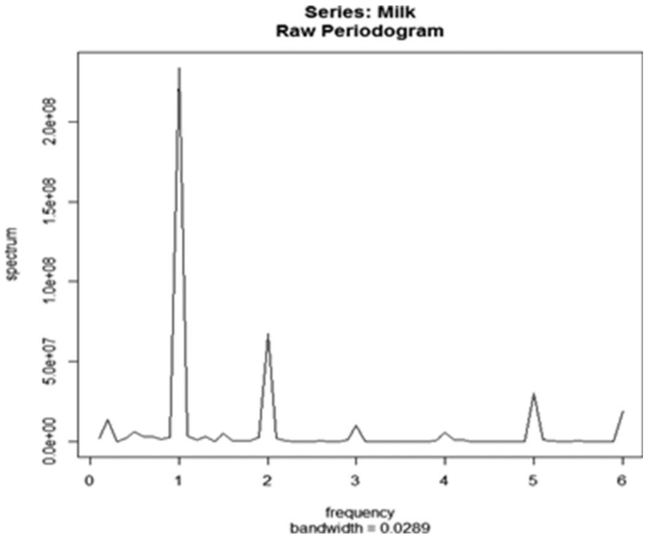

Figure 3: Periodogram for monthly milk delivery from January 2001 to December 2010

off smoothly with distance from the target point, making the fitted function continuous. For further details see e.g. Hastie *et al*., (2009).

The Naive model

In analysis of time series analysis with yearly seasonal cycles is common to use values from the last known year as a benchmark to evaluate forecasting models against. In the Naïve model I therefore set all forecasts equal to the value from the same month in the last year of the training set. For example, the forecast for all future February values is set equal to the last observed February value (2010), and so on. Here I use the Naive model more as a benchmark rather than the method of choice. If the other methods do not outperform the naïve model, they are not worth considering.

The SARIMA model

In general the multiplicative SARIMA model is given by

$$\Phi_p(B^s)\varphi(B)\nabla_s^D\nabla_s^d y_t = \delta + \Theta_Q(B^s)\theta(B)w_t, \quad (2)$$

(Shumway and Stoffer, 2010), where w_t is Gaussian white noise processes, or simply uncorrelated random variables. The general model is denoted ARIMA $(p, d, q) \times (P,D,Q)_s$. The ordinary autoregressive and moving average components are represented by the polynomials $\Phi(B)$ and $\varphi(B)$ of orders p and q respectively. The seasonal autoregressive and moving average components are represented by $\Phi_p(B^s)$ and $\Phi_Q(B^s)$ of orders P and Q, and the ordinary and seasonal difference components by

$$\nabla^d = (1-B)^d, \text{and} \quad (3)$$

$$\nabla_s^D = (1-B)^D. \quad (4)$$

B is the backshift operator:

$$B \cdot y_t = y_{t-1}. \quad (5)$$

There are a few basic steps to fitting SARIMA models to time series data. These steps involve plotting the data, possibly transforming the data, identifying the dependence orders of the model, parameter estimation, diagnostics, and model choice. First I construct a time plot of the data, and inspect the graph for any anomalies. In the example the variability in the data grows with time, and therefore it is necessary to transform the data to stabilize the variance. Here I apply first differencing,

$$\nabla y_t = y_t - y_{t-1}, \quad (6)$$

to stabilize the variance. This means that we simply look at the difference between two adjacent months.

The next step is to identify preliminary values of the autoregressive order, *p*, the order of differencing, *d*, and the moving average order, *q*. When preliminary values of *d* have been settled, the next step is to look at the sample ACF (Autocorrelation function) and PACF (Partial autocorrelation function) for whatever values of *d* have been chosen. With monthly milk production data, there is a strong yearly component occurring at seasonal lags s that are multiples of $s=12$, because of the strong connections of all biological activities to the calendar year. Because of this, it is appropriate to introduce autoregressive and moving average polynomials that identify with the seasonal lags. For diagnosis and fit of the SARIMA model I refer to the Appendix.

I tried models with different time lags, and to compare the models I use Akaike's Information Criterion (Akaike, 1969; 1973; 1974) and the Bayesian Information Criterion (Schwarz, 1978). Once I had fitted a suitable time series model to the historic data, I used the model to forecast future milk delivery. To assess the precision of the forecasts, prediction intervals are calculated along with the forecast. I choose the model with the lowest AIC- and BIC- values, AIC = 2016.31, and BIC = 2051.06. The variance σ_w^2 is estimated to 4932115, and the log likelihood to -995.16. The chosen model has the form:

(2,1,4)*(2,1,4)

The three values in the first bracket are related to year effects, and the last three to seasonal effects. The value two means that we apply the milk quantity in the same month as the one we are predicting for two years back in time. The value four means that for white noise we use the noise from the actual month we are predicting for four years back in time. Similarly, for the seasonal effect we use the milk quantity two months back in time in the actual year, and the noise four months back in time. The two number 1's mean that we differentiate by one month both between years and between months. Once it is developed, the SARIMA model is very easy to use and update. In practice it takes ten minutes to update it with new delivery figures every month.

The Holt Winters' exponential smoothing model

When a time series can be described using a model with increasing or decreasing trend and seasonality, Holt-Winters exponential smoothing (HW) can be applied to make short-term forecasts (Holt, 1959; Winters, 1960). The HW is an exponential smoothing approach for handling seasonal data. It is a widely used tool for forecasting business data that contain seasonality, changing trends and seasonal correlation. A weakness of the HW is that it can be sensitive to unusual events and outliers. Exponential smoothing methods give larger weights to more recent observations, and the weights decrease exponentially as the observations become more distant. HW is based on three smoothing equations —one for the level, one for trend, and one for seasonality. The HW forecast is determined using three smoothing constants, α, β and γ, with values between 0 and 1, and the following four equations:

$$\text{Level:} \quad \ell_t = \alpha y_t / s_{t-m} + (1-\alpha)(l_{t-1} + b_{t-1}), \quad (7)$$

$$\text{Growth:} \quad b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta^*) b_{t-1}, \quad (8)$$

$$\text{Seasonal:} \quad s_t = \gamma y_t / \ell_t + (1-\gamma) s_{t-m}, \quad (9)$$

$$\text{Forecast:} \quad \hat{y}_{t+h|t} = (\ell_t + b_t h) s_{t-m+h} \quad (10)$$

Here m is the length of the seasonal cycle (e.g., the number of months), ℓ_t represents the level of the series, b_t denotes the growth or trend, s_t is the seasonal component and $\hat{y}_{t+h|t}$ is the forecast for h periods ahead. With a monthly time series s_{t-m+h} becomes S_{t-11} when forecasting one step ahead. For a more detailed description of the HW model I refer to e.g. Hyndman *et al.* (2008).

There are two versions of the HW model, the additive and the multiplicative. I apply the multiplicative version which uses seasonal factors as multipliers rather than additive constants, because the multiplicative model gives the best fit to the training set. In practice this also seems to be the most commonly used (Hyndman *et al.*, 2008). The HW model allows trend and seasonal pattern to change over time. Values of the smoothing parameters that are close to 0 mean that relatively little weight is placed on the most recent observations when making forecasts.

Every exponential smoothing method requires initialization of the smoothing process. A robust and objective way to obtain values for the unknown parameters included in any exponential smoothing method is to estimate them from the observed data. The unknown parameters and the initial values for any exponential smoothing method can be estimated by minimizing the sum of squared prediction errors (SSE) over the training set, where the one-step-ahead within sample prediction error is specified as

$$e_t = y_t - \hat{y}_{t|t-1} \; for \; t = 1, \ldots, T. \quad (11)$$

This procedure involves a non-linear minimization problem. The optimizing function 'optim' in R tries to find the optimal values of α and/or β and/or γ by minimizing the squared one-step prediction error. Further, in R the start values for level, trend and season are inferred by performing a simple decomposition in trend and seasonal component using moving averages on the first periods of the training set. A simple linear regression on the trend component is used for starting level and trend.

The estimated HW model yields a SSE of 807928649. The estimated values of alpha, beta and gamma are 0.399, 0.00, and 1.0, respectively. The value of alpha is relatively low, indicating that the estimate of the level at the current time point is based upon both recent observations and some observations in the more distant past. The value of beta is 0.00, indicating that the estimate of the slope of the trend component is not updated over the time series, and instead is set equal to its initial value. This makes good intuitive sense, as the level changes over the time series, but the slope b of the trend component remains roughly the same. In contrast, the value of gamma (1.0) is high, indicating that the estimate of the seasonal component at the current time point is just based upon very recent observations. For diagnosis of the HW model I refer to the Appendix. So far I conclude that the HW model provides an adequate predictive model for monthly milk delivery. Similar to the SARIMA model the HW model is easy to use and update with new monthly figures.

The Expert model

This is the model which TINE SA cooperative dairy company uses today. To construct forecasts, dairy experts give their opinion on the most probable number of dairy cows per month, the daily milk yield per cow per month and the milk quota filling up to 14 months ahead in time. To judge the future number of cows the experts estimate the number of first calving heifers based on historic figures of inseminated and slaughtered heifers. Similarly, they estimate the number of cows that will be slaughtered based on historic figures. To make an assumption of the future milk yield per cow the dairy experts use historic figures and supplemental information on the forage harvest with respect to both quantity and quality. Finally, to make an assumption of the future milk quota filling the experts use historic data and also consider possible adjustments in the quota regulations. Thus in practice the Expert model is a combination of historic figures and expert opinions. The model is quite time consuming, and therefore the forecasts are made only every other month.

I now present the measures I use to compare the forecast from the Expert model with the forecast from the Naïve model, the SARIMA model and the HW model.

Choice of accuracy measures for the forecasting models

There are many ways to evaluate the accuracy of forecasting methods. They all involve looking at past data and comparing the value that would have been forecasted using the model and the estimated parameters, $\hat{y}_{t|t-1}$, with the actual observation, y_t. Different accuracy measures often give different results (Hyndman *et al.*, 2008). Therefore the choice of accuracy measure must be adapted to the problem at hand. In this study there are two main goals of forecasting. For the dairy industry it is important to maximize exploitation of capacities. Therefore it is important to know the milk quantity in each month ahead. To measure the forecast accuracy of the different models in each month I apply a widely used measure of variability, the mean absolute percentage error (MAPE) (Hyndman *et al.*, 2008). Percentage errors have the advantage of being scale-independent, and so are frequently used to compare forecast performance between different data sets.

$$MAPE = 100/n \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_{t|t-1}}{y_t} \right| \quad (12)$$

In MAPE a positive forecast error in one month is not outweighed by a similar negative error in another month. Thus all monthly forecast errors contribute to increase the MAPE- value, and therefore the MAPE is advantageous when the main interest is to measure the milk quantity of each month. However, measures based on percentage error have the disadvantage of having extreme values when y_t is close to zero, and they also assume a scale with a meaningful zero.

The dairy industry also needs to know the total milk quantity over a longer period, e.g. two years ahead. To measure the forecast accuracy over a longer period I will also apply the sum of forecasting errors (SFE), which is simply calculated by

$$SFE = \sum_{t=1}^{n} (y_t - \hat{y}_{t|t-1}) \quad (13)$$

The SFE provides an indication of bias, i.e. the overestimation or under-estimation in the model. In SFE a positive forecast error in one month can be outweighed by a negative error in another month, since we do not take the absolute values of the errors. Both accuracy measures have in common that the lower the values, the better the forecast.

3. Results

In this section I present the results from the comparison of the four models.

Figure 4 shows the MAPE values for the four different forecasting models over different forecasting horizons in months.

From Figure 4 we can see that for the first six months the Expert model outperforms the other models, with very low values of MAPE. After six months the Expert model is outperformed by the HW model, and the Expert model does not give forecasts beyond 14 months. The HW model continues to give the most reliable forecast up to 18 months. After 21 months the HW model gives poorer forecasts than both the Naïve model and the SARIMA model. From 18 months on the SARIMA model performs better than the others up to 24 months, although the difference compared with the Naïve model is small both at 18 and 21 months. The SARIMA model produces the least accurate forecast of the models for the first six months, but has the best long term performance.

Figure 5 compares the different forecasting models according to the SFE. From Figure 5 we notice that the first three months the Expert model has the lowest SFE, which is in line with the finding in Figure 4. After six months the SARIMA model performs similarly to the Expert model, which in practice loses its' predictive power after six months. The HW model performs best after 12 months, with SARIMA second. They both perform significantly better than the Naïve model. From 12 months to 21 months the SARIMA model has by far the lowest SFE, and significantly lower than the Naïve model. However, the difference between the SARIMA model and the Naïve model declines sharply when we reach 24 months. Thus there seems to be less to gain from applying time series models for forecasting horizons beyond 24 months.

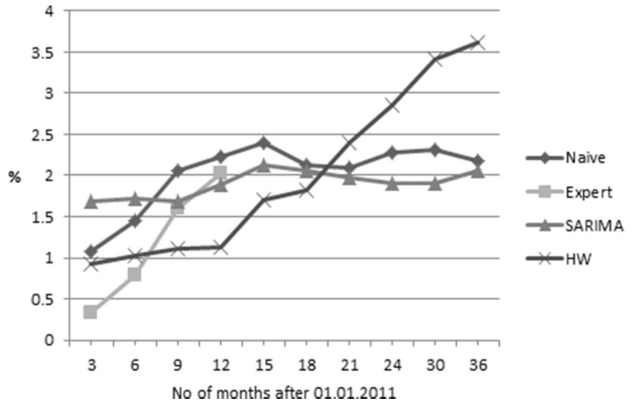

Figure 4: MAPE over different forecasting horizons in months for the forecasting models

4. Discussion and conclusion

Milk production in Norway first declined and then increased significantly during the forecast period. In spite of this the forecasting models fit the data quite well, at least in the first 18 months, with MAPE values equal to or less than approximately two. Thus the findings support the findings of Sataya *et al*. (2007) and Sankar and Prabakaran (2012) with respect to the SARIMA model, except that these authors did not take seasonality into account. The findings reported here show that the Expert model loses much of its predictive power after six months. This is noteworthy since its main application today is for forecasts 12 to 14 months ahead, and the findings show that for this purpose the time series models are preferable. Further investigation has revealed that the main cause of the prediction error is the misjudgement of future milk yield per cow.

If we look at the MAPE, the HW model performs best from 12 to 18 months, but if we look at the SFE the SARIMA model performs best. The findings reported here illustrate that the two accuracy measures serve slightly different purposes. If the purpose is to forecast the milk quantity in each month accurately, one should use the Expert model for the first six months, the HW model from six to 18 months, and the SARIMA model from 18 to 24 months. However, if the main interest is to measure the total milk quantity over a period of several months, one can use the SARIMA model the first six months, the HW model the next six months, and then the SARIMA model again from 15 to 24 months.

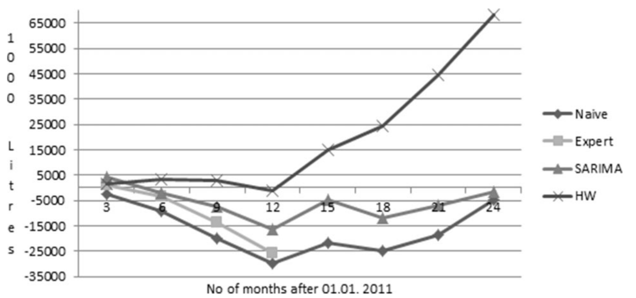

Figure 5: SFE in 1000 litres for the four models over different forecasting horizons in months

The finding that the HW model loses much of its predictive power between 15 and 18 months is somewhat contrary to the finding of Akter and Rahman (2010), who claimed that the HW model could be used for forecasts up to two years. When interpreting the differences between the two findings one should keep in mind that Akter and Rahman (2010) used a much shorter time series, which give higher forecast errors. In general the MAPE values in this study are lower than the ones reported by Akter and Rahman (2010). However, conflicting results like this are not uncommon when performing forecasting competitions between methods (Hyndman *et al*., 2008). As forecasting tasks can vary by many dimensions considering the length of forecast horizon, the size of test set, the forecast error measures and the interval of data etc., it is unlikely that e.g. time series models will be better than all other models for all forecasting scenarios. What we require from a forecasting method are consistently sensible forecasts, and these should be frequently evaluated against the task at hand.

According to the results forecasts beyond 24 months should be dealt with caution, and here the findings are in line with the findings of Akter and Rahman (2010). However, contrary to their advice I recommend the SARIMA model instead of the HW model when forecasts beyond two years are necessary. Even after two and a half years the SARIMA model still has a low MAPE, but after three years the difference compared with the Naïve model is negligible. If we take all three years together the Naïve model performs remarkably well.

Taken together the findings that time series models should be combined are in line with Hyndman *et al.* (2008), who claim that the SARIMA models and the HW models overlap and are complementary. They both have their strengths and weaknesses. The underlying presumption that correlation between adjacent points in time is best explained in terms of a dependence of the current values on past values represents means that both models depend heavily on the time period analysed. Thus analysis of other periods could produce other models. This dependence makes it necessary to recalibrate the time series models regularly. Unlike most prior studies this study compares time series models of milk delivery with a model based on expert opinions. I think the findings reported here show that time series models can make the dairy sector more proactive and capable of responding more quickly changes in milk production like e.g. crop failure, and at a lower cost. A possible avenue for future research could be to try to improve the performance of the Expert model by combining it with time series models. For example one could use time series models to forecast the number of cows and the milk yield per cow.

In conclusion the Expert model performs well for the first six months, but has the disadvantage that it is much more time consuming than the times series models. A combination of the Expert model, the HW model and the SARIMA model gives reliable forecasts of monthly milk delivery for a period of up to two years. Forecasts beyond two years should be dealt with caution. However, the SARIMA model still performs better than the Naïve model up to three years ahead.

About the author

Bjørn Gunnar Hansen holds Master degrees in agricultural science and economics and business management, and a Ph.D. in business management. He currently works as a dairy scientist in TINE SA, the Norwegian cooperative dairy company, where the research was carried out.

REFERENCES

Akaike, H., (1969). Fitting autoregressive models for prediction. Ann. Inst. Stat. Math. 21: 243–247.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principal. In *2nd Int. Symp. Inform. Theory, 267–281. B.N. Petrov and F. Csake, eds*. Budapest: Akademia Kiado.

Akaike, H. (1974). A new look at statistical model identification. *IEEE Trans. Automat. Contr., AC*, 19, 716–723.

Akter, S. and Rahman, S. (2010). Agribusiness forecasting with univariate time series modelling techniques: The case of a dairy cooperative in the UK. *Journal of Farm Management,*13 (11), 747–764.

Barney, J. (1986). Strategic factor markets: expectations, luck, and business strategy. *Management Science*, 42, 1231–1241. DOI: 10.1287/mnsc.32.10.1231

Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis, Forecasting, and Control*. Revised ed., Holden-Day.

Chatfield, C. and Yar, M. (1988). Holt-Winters Forecasting: Some Practical Issues. *The Statistician*, 37, 129–140. DOI: 10.2307/2348687

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning.* 2nd ed. New York. Springer. DOI: 10.1007/978-0-387-84858-7

Holt, C.C. (1959). Forecasting seasons and trends by exponentially weighted moving averages. *ONR Research Memorandum,* 52. http://CRAN.R-project.org/ Retrieved 16.05.14.

Hyndman, R., Kohler, A.B., Ord, J.K. and Schneider, R.D. (2008). *Forecasting with exponential smoothing. The State Space Approach.* New York. 1st ed. Springer. DOI: 10.1007/978-3-540-71918-2

IFCN (2012). *Dairy Report 2012*. IFCN Dairy Research Center, Kiel, Germany.

Makadok, R. and Walker, G. (2000). Identifying a distinctive competence: forecasting ability in the money fund industry. *Strategic Management Journal*, 21, 853–864. DOI: 10.1002/1097-0266(200008)

Sankar, T.J. and Prabakaran, R. (2012). Forecasting milk production in Tamilnadu. *International Multidisciplinary Research Journal*, 2, 10–15.

Sataya, P., Ramasubramanian, V. and Mehta, S.C. (2007). Statistical models for forecasting milk production in India. *J. Ind. Soc. Agril. Statist.* 61, 80–83.

Schwarz, F. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. DOI: 10.1214/aos/1176344136

Shumway, R. and Stoffer, D. (2010). *Time Series Analysis and Its Applications.* 3rd ed. New York. Springer. DOI: 10.1007/978-1-4757-3261-0

TINE, (2013). *Annual report*. TINE cooperative dairy company, Oslo.

Watson, G.S. (1966). Smooth regression analysis. *Sankhya,*26, 359–378.

Winters, P.R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, 324–342.

Table 1: The coefficients in the fitted time series model

ar1	ar2	ma1	ma2	ma3	ma4	sar1	sar2	sma1	sma2	sma3	sma4
-0.064	0.5755	-0.2688	-0.8171	0.1239	-0.0380	0.8111	-0.9815	-1.2806	1.7292	-0.8294	0.3242

Appendix: diagnosis of the SARIMA and HW models

The SARIMA model

The coefficients in the fitted model are given in Table 1.

In Table 1 the notions 'ar' and 'ma' refer to the autoregressive and moving average coefficients for year. Correspondingly, the notions 'sar' and 'sma' refer to the coefficients for months within year. Diagnosis of the model involves inspection of the residuals. Investigation of marginal normality can be accomplished visually by looking at a histogram of the residuals. In addition to this, a normal probability plot or a Q-Q plot can help in identifying departures from normality. I also inspect the sample autocorrelations of the residuals for any patterns or large values. Finally I check the Ljung Box Pierce Q-statistic to reveal possible accumulated autocorrelation between the residuals. In Figure 6 displays the diagnostic tools for the chosen model.

The standardized residuals show no obvious patterns. Notice that there are outliers, however, with a few values exceedingL 3 standard deviations in magnitude. The outliers that occur in 2008 are due to change of the quota year from January 1 to March 1. The ACF of the standardized residuals are low and show no apparent departure from the model assumptions. The normal Q-Q plot of the residuals shows some departure from normality at the tails. However, the model appears to fit well except for the fact that a distribution with heavier tails than the normal distribution could be employed. The Ljung-Box-Pierce Q- statistic uncovers no problems with autocorrelation between the residuals.

The HW model

If the predictive model cannot be improved upon, there should be no correlations between forecast errors for successive predictions. In other words, if there are correlations between forecast errors for successive predictions, it is likely that the simple exponential smoothing forecasts could be improved by another forecasting technique. To figure out whether this is the case, I obtain a correlogram of the in-sample forecast errors for lags 1–20 (Figure 7).

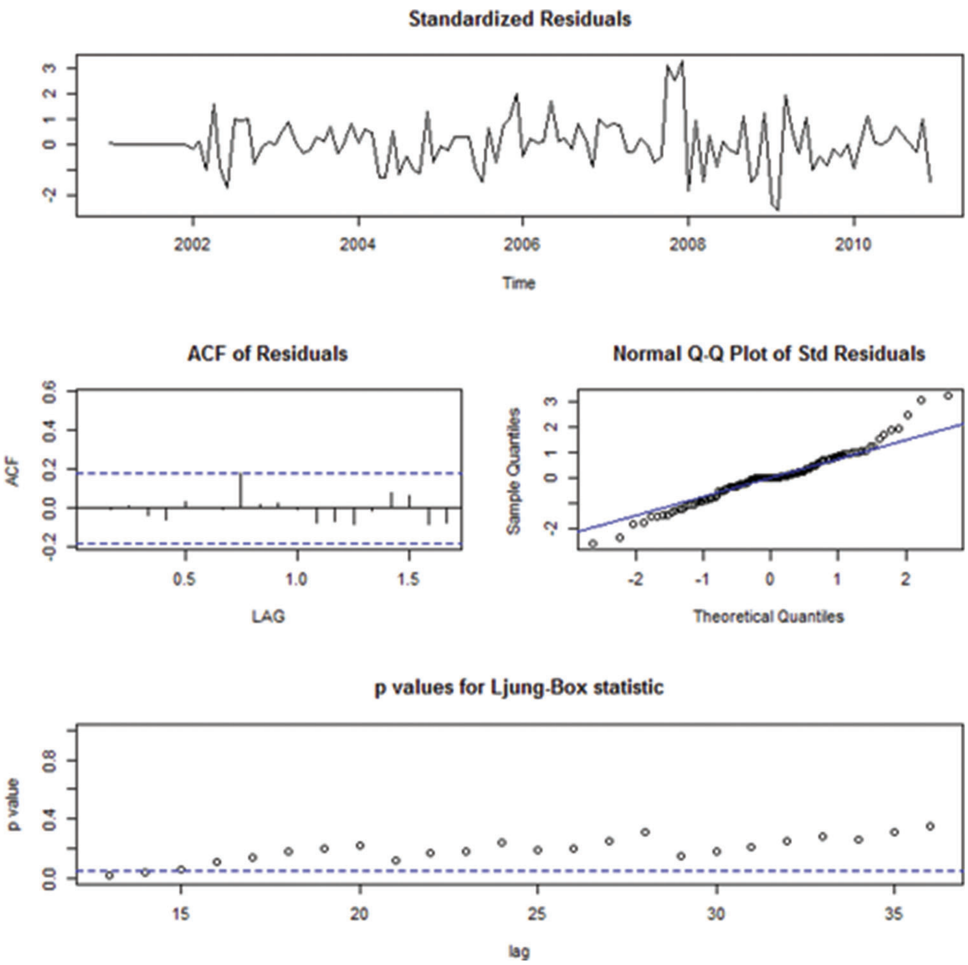

Figure 6: Analysis of residuals for the SARIMA model

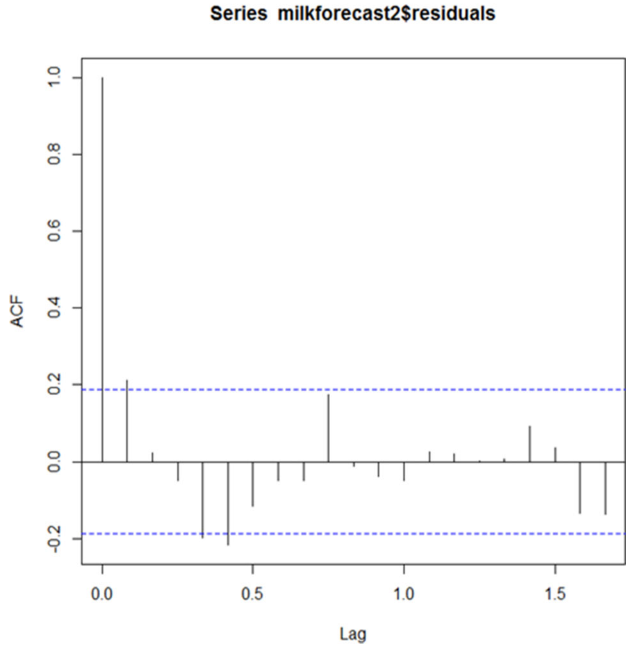

Figure 7: The autocorrelation function for in-sample forecast errors for the HW model

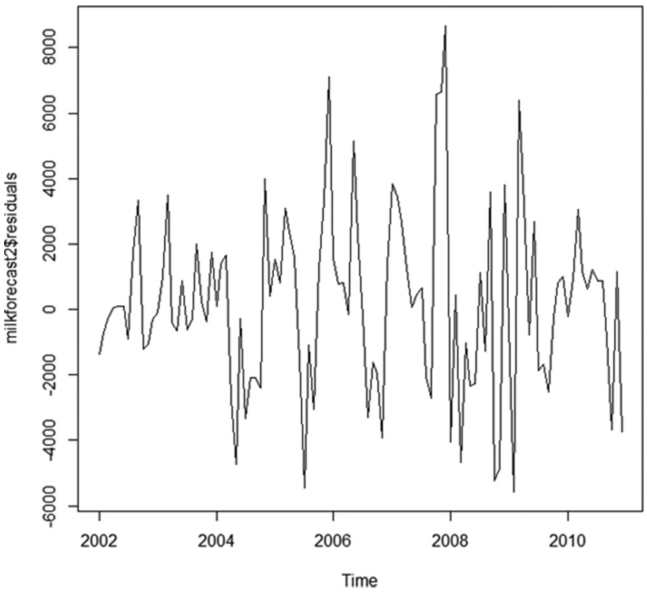

Figure 8: Residuals for the HW model

From Figure 7 we notice that there is very little autocorrelation between forecast errors between lags. To test whether there is significant evidence for non-zero correlations between residuals at lags 1–20, I carry out a Ljung-Box test. The Ljung-Box test statistic is 27.824 and the p-value is 0.11, so there is little evidence of non-zero autocorrelations in the in-sample forecast errors at lags 1 to 20.

I check whether the forecast errors have constant variance over time, and are normally distributed with mean zero, by making a time plot of the forecast errors (Figure 8) and a histogram with overlaid normal curve (Figure 9).

The plot shows that the in-sample forecast errors seem to have roughly constant variance in the middle of the time period. However, the fluctuations at the start and at the end of the time series are smaller than in the middle.

To check whether the forecast errors are normally distributed with mean zero, I plot a histogram of the forecast errors, with an overlaid normal curve that has mean zero and the same standard deviation as the distribution of forecast errors (Figure 9).

Figure 9 shows that the distribution of forecast errors is roughly cantered on zero, and is more or less normally distributed, although it seems to be slightly skewed to the left compared to a normal curve. However, the left skew is relatively small, and so it is plausible that the forecast errors are normally distributed with mean zero.

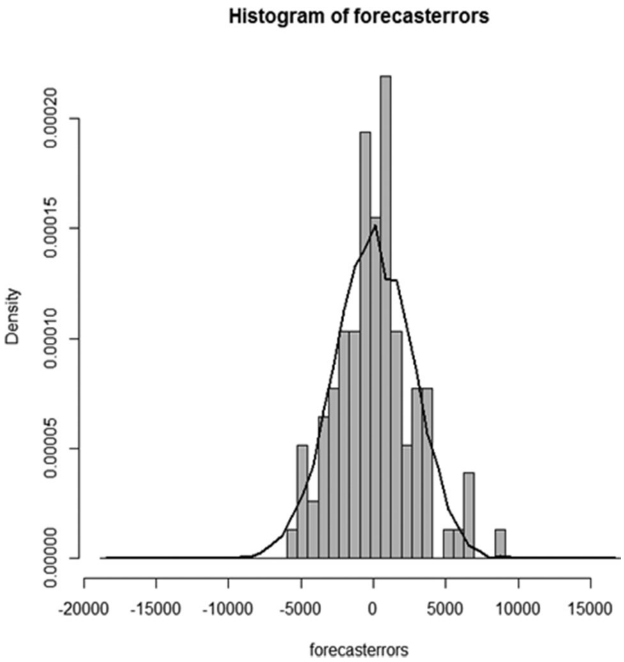

Figure 9: Histogram of forecast errors with overlaid normal curve from the HW model

