Irish Journal of Agricultural and Food Research

Forecasting prices of dairy commodities – a comparison of linear and nonlinear models

B.G. Hansen[1][†]

[1]TINE SA, Department of Research and Development, Farm Advisory Services, TINE SA, Postbox 58, N-1431, Ås, Norway

Abstract

Dairy commodity prices have become more volatile over the last 10–11 yr. The aim of this paper was to produce reliable price forecasts for the most frequently traded dairy commodities. Altogether five linear and nonlinear time series models were applied. The analysis reveals that prices of dairy commodities reached a structural breakpoint in 2006/2007. The results also show that a combination of linear and nonlinear models is useful in forecasting commodity prices. In this study, the price of cheese is the most difficult to forecast, but a simple autoregressive (AR) model performs reasonably well after 12 mo. Similarly, for butter the AR model performs the best, while for skimmed milk powder (Smp), whole milk powder (Wmp) and whey powder (Whp) the nonlinear methods are the most accurate. However, few of the differences between models are significant according to the Diebold–Mariano (DM) test. The findings could be of interest to the whole dairy industry.

Keywords

Dairy commodity prices • forecasting • linear models • nonlinear models • volatility

Introduction

Milk products account for about 6% of agricultural exports FAO (2016). Although the global dairy trade increases every year, only 14.3% of all milk delivered to dairy and 20% of all tradable dairy products were traded internationally in 2013 (IFCN, 2014). The four main traded dairy products on the world market are butter, cheese, skimmed milk powder (Smp) and whole milk powder (Wmp) (FAO, 2016). The prices of these four commodities and whey powder (Whp) from 2002 to 2016 are shown in Figure 1. All prices have a general upward trend until 2014, with increasing volatility from 2007 on. The variation in the length and amplitude of the price cycles challenges forecasting and decision making. Prices declined sharply in 2015 due to- decreased demand from China and the Russian Federation's import embargo for several dairy products. Increased production from key exporters as the European Union (EU) also played a role. Increased production in the EU was associated with abolishment of milk quotas. The price turned upward again in the second half of 2016, triggered by a slowdown in milk supply.

Given the high volatility of commodity prices and the importance of raw materials in production, accurate forecasts are of great interest for various purposes. For business management of farm businesses, agribusinesses, wholesalers and retailers, forecasting commodity prices is important to marketing or procurement. Reliable price forecasts can aid cash flow management and improve farm production decision planning: what and how much feed to grow, what time of year to produce the milk, etc. A price forecast is also useful for planning annual cash flow and loan requests. Good price forecasts will also give dairy companies better tools to plan their activities, for example, when to sell their products, what produce and when. Therefore, research on agricultural price forecasting is important (Martin-Rodriguez & Cáceres-Hernández, 2012).

The causes of extreme price volatility in dairy commodity markets are well established in economics literature. Even small changes in supply can cause very large changes in price (O'Connor & Keane, 2011; Bolotova, 2016). Economic theory suggests that seasonality and cycles are common features in agricultural commodity prices (Tomek & Robinson, 2003; Piot-Lepetit & M'Barek, 2011). Such cycles may be the result of the lag between the decision to change milk supply be based on current price and the actual availability of this milk on the market because of the time it takes to expand or contract supply (Bergmann *et al*., 2015). While price cycles

[†]Corresponding author: B.G. Hansen
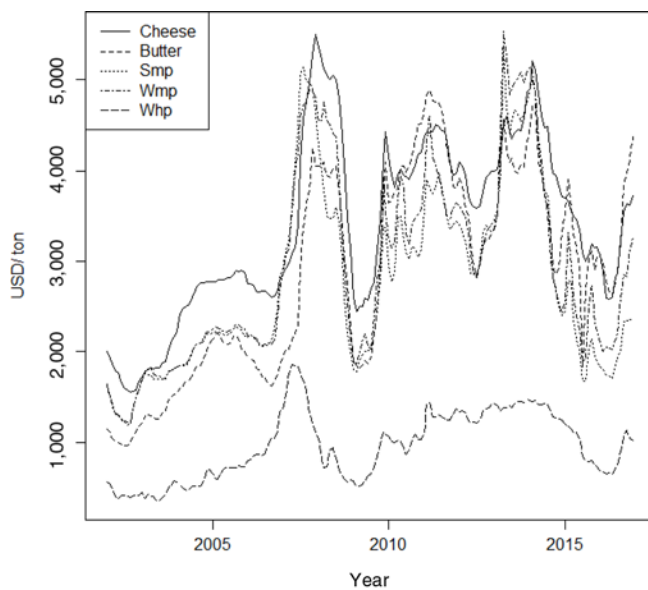E-mail: bjorn.gunnar.hansen@tine.no

Figure 1. World market prices of cheese, butter, Smp, Wmp and Whp from 2002 to 2016 in USD/ton.

have attracted a relatively strong attention in the literature, there are relatively few studies focusing on forecasting of dairy commodity prices over the last 15 yrs. A finding from using autoregressive integrated moving average (ARIMA) models is that unit root behaviour is common in commodity prices (Myers *et al.*, 2010). To cope with this problem, both cointegration and vector error-correction models (VECM) gained popularity. Among other models used to analyse agricultural commodity price series are structural time series (Labys & Kouassi, 1996; Durbin & Koopman, 2001; Nicholson & Stephenson, 2015), multi-resolution analysis (Hansen & Li, 2016) and state space models (Aoki & Havenner, 1991; Foster *et al.*, 1995; Walburger & Foster, 1998).

An interesting question is whether futures have the potential to make traditional price forecast methods redundant. A relatively large body of literature has explored the predictive performance of futures prices for different commodities, ranging from oil and metals to cattle and dairy products. Yang *et al.* (2001), Bowman & Husain (2004), Coppola (2008), Reichsfeld & Roache (2009), Reeve & Vigfusson (2011) and Chinn & Coibion (2013) found evidence which support the predictive performance of futures. Contrary, Moosa and Al-Loughani (1994), Fortenbery & Zapata (1997), Chernenko *et al.* (2004), Ahlquist & Kilian (2010) and Ahlquist *et al.* (2013) found little support for futures as the best forecast. To sum up, literature has not yet reached a consensus on the predictive performance of futures. Thus, it is fair to claim that the introduction of futures has not deemed traditional price forecasts redundant.

In recent years, nonlinear time series models have become increasingly popular in fields such as macroeconomics and finance (Teräsvirta *et al.*, 2011). However, in forecasts of dairy commodity prices, there are relatively few applications of nonlinear models. The aim of this article was to explore the usefulness of nonlinear time series models as compared to linear models in forecasting the world's five most traded dairy commodity prices. The remainder of the paper is organized as follows: First materials and methods are presented, then follows results, discussion and conclusion.

Material and preliminary statistics

For all commodities except Whp, monthly prices from the United States Department of Agriculture agricultural marketing service (USDA, 2019) were used. For Whp, the prices from 2010 on were collected from Süddeutsche Butterbörse (SB, 2019), and earlier from the USDA. Summary statistics of the five dairy commodities from 2002 to 2016 are provided in Table 1. The cheese price refers to the price of cheddar.

In Table 1, one can see that there is considerable volatility in all prices, with butter, Smp and Wmp showing the largest fluctuations.

We now study some statistical properties of the price series, shown in Tables A1 and A2. Tables A1 and A2 report the Elliott–Rothenberg–Stock (ERS) (Elliott *et al.*, 1996) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) (Kwiatkowski *et al.*, 1992) tests for nonstationarity. The ERS test takes unit root as the null hypothesis, while the KPSS test takes the null hypothesis as a stationary process. While the ERS test statistics are mixed with respect to a unit root, the KPSS test results show that most series are nonstationary. However, when dealing with nonlinear time series, one should keep in mind that many unit root tests have a low power against nonlinearity. Thus, tests of unit root may confuse nonlinearity with a unit root. Generally, if the data has a unit root, we differentiate the series to make it stationary. However, when the

Table 1: Summary statistics for the prices of cheese, butter, Smp, Wmp and Whp from 2002 to 2016 in USD/ton

Commodity	Average	s.d.	Min.	Max.
Cheese	3,381	992	1,550	5,500
Butter	2,861	1,127	963	4,890
Smp	2,775	1,017	1,196	5,348
Wmp	2,935	1,080	1,229	5,538
Whp	957	376	363	1,856

Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.

data is nonlinear, taking differences will change the nonlinear structure of the data, and therefore it is not recommended (see, e.g., Teräsvirta *et al*., 2011). Thus, under the heading "stationarity and non-stationarity" on page 3, Teräsvirta *et al*. (2011). state that "it is necessary to develop new tools, and well-tried, familiar tools have to be discarded as being no longer appropriate".

Methods

Testing for linearity and structural breaks

In this paper, the parametric Tsay test (Tsay, 1986) against linearity together with the Teräsvirta test (Teräsvirta *et al*., 1993) and the Brock, Dechert and Scheinkman (BDS) test (Brock *et al*., 1987) are applied. Originally designed to test for independence and identical distribution (iid), the BDS test is also shown to have power against nonlinear alternatives (Brock *et al*., 1991). In fact, simulations have shown that it exhibits the highest power in detecting nonlinearity, and for this reason it should be the first to be used (Bisaglia & Gerolimetto, 2014). One advantage of the BDS test is that it is a statistic which requires no distributional assumption of the data to be tested.

Testing whether the parameters in a model are constant over time is another option. The Chow test proposed by Chow (1960) and the cumulative sum (CUSUM) test (Brown *et al*., 1975) were applied. Unfortunately, there is no good theory about how to forecast in the presence of breaks, and the simple rule is to split the sample at the estimated break (Hansen, 2012). One obvious disadvantage with this procedure is loss of data. Others recommend forecasting models with time varying parameters (Hyndman, 2014), for example, the self-exciting threshold autoregressive (SETAR) model in (2). In a preliminary analysis, the author tried to split the sample at the point of the structural break. However, this did not improve forecasts significantly, and it was therefore decided to include the whole sample in the analysis.

Time series models and forecasting methods

The linear model

The primary argument for linearity is simplicity in estimation, interpretation and forecasting, and the analysis of nonlinear models is reasonably straightforward.

The autoregressive (AR) model

As AR models are well known and make up important building blocks in all subsequent models in this paper, it is reasonable to include AR models in the analysis. An AR model is based on the idea that the current value of a series y_t can be explained as a function of p past values, y_{t-1}, y_{t-2}, y_{t-p}, where p determines the number of steps into the past needed to forecast the current value. An AR model for y_t of order p is of the form

$$y_t = \phi_1(y_{t-1}) + \phi_2(y_{t-2}) + \dots + \phi_p(y_{t-p}) + \varepsilon_t, \quad (1)$$

where y_t is stationary, $\phi_1, \phi_2, \dots, \phi_p$ are constants ($\phi_p \neq 0$) and ε_t is a Gaussian white noise process (Shumway & Stoffer, 2010).

Nonlinear models

To model nonlinear behaviour, it is natural to allow for the existence of different states of the world or regimes, and to allow the dynamics to be different in different regimes. By extending AR models to allow for nonlinear behaviour, the resulting nonlinear models are easy to understand and due to their variety and flexibility, the so-called regime switching models have become popular in the class of nonlinear models.

1. The SETAR model

The SETAR model is a piecewise linear autoregression which has been widely applied in econometrics. A popular choice in practice is the two-regime model (Teräsvirta *et al*., 2011):

$$y_t = (\boldsymbol{\phi}_1 \boldsymbol{w}_t + \varepsilon_{1t}) \, I(y_{t-d} < c_1) + (\boldsymbol{\phi}_2 \boldsymbol{w}_t + \varepsilon_{2t}) \, \{1 - I(y_{t-d} < c_1)\} \quad (2)$$

Here $\boldsymbol{w}_t = (1, y_{t-1}, \dots y_{t-p})$, and y_{t-d}, $d > 0$, is an observable switch-variable. I is an indicator function, c_1 a threshold parameter to be estimated and $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are parameter vectors. Further, $\varepsilon_{1t} = \sigma_1 \varepsilon_t$ and $\varepsilon_{2t} = \sigma_2 \varepsilon_t$, with $\{\varepsilon_t\} \sim$ iid $(0, 1)$, $\sigma > 0$. Finally, d is a positive integer indicating the time delay. The regime of y_t is determined by its own lagged value y_{t-d}, thus the term "self-exciting", and d determines with how many lags does y_{t-d} influence the regime at time t. The switch-points are generally unknown. The observations y_t are generated either from the first regime when y_{t-d} is smaller than the threshold, or from the second regime when y_{t-d} is greater than the threshold. Estimation of the model can be carried out by conditional least squares.

Before developing an SETAR model, the time series were tested for the existence of possible threshold-type nonlinearity and the number of such thresholds. The test proposed by Hansen (1999) implemented in the R procedure SETARTest was applied. To select an appropriate SETAR model, the selectSETAR function in R was used. The time series, the embedding parameters and a vector of values for each provided hyper-parameter are passed to this function. The routine then tries to fit the model for the full grid of hyper-parameter values and gives as output a list of the best combinations with respect to the pooled Akaike information criterion (AIC). The pooled AIC sums the AICs in

the different regimes. All SETAR models in this paper have only two regimes.

2. The logistic smooth transition autoregressive (LSTAR) model

Sometimes it is reasonable to assume that the regime switch happens gradually in a smooth fashion. The step function $I(y_{t-d} \leq c_1)$ in (2) is replaced by a transition function. Here, the focus is on the logistic transition function. An LSTAR model of order p can be defined as (Teräsvirta et al., 2011):

$$y_t = \phi_{10} + \boldsymbol{\phi}_1 \boldsymbol{w}_t + (\phi_{20} + \boldsymbol{\phi}_2 \boldsymbol{w}_t) \left[(1 + \exp\{-\gamma(y_{t-d} - c)\})^{-1} \right] + \varepsilon_t \tag{3}$$

where $\varepsilon_t \sim$ iid(0, σ_ε^2) and γ determines the speed and smoothness of the transition, $\gamma > 0$. The remaining parameters are similar to (4). Now the observations y_t switch between two regimes smoothly in the sense that one regime has more impacts in some times, and the other regime in other times. If, for example, $\gamma = 0$ holds, then model (3) is a linear AR (p) model. When $\gamma = 0$ then ϕ_{20}, ϕ_2 and c in (3) can take any values, and the model is not defined. To choose parameters in the LSTAR models, we used the selectLSTAR function in the package tsDyn.

3. The additive autoregressive model (AAR)

In contrast to the SETAR and LSTAR models, the AAR model is nonparametric. Additive models are a flexible class of estimators which combine many methods as building blocks for fitting an additive model. AAR models avoid forcing the data into a given structure. Thus, they maintain a lot of the nice properties of linear models but are much more flexible.

A nonlinear AAR model can be written as:

$$y_t = f_1(y_{t-1i}) + f_2(y_{t-2i}) + \ldots + f_p(y_{t-ip}) + \varepsilon_t, \tag{4}$$

where $\varepsilon_t \sim$ iid(0, σ_ε^2), is are positive integers, and $f_i(\cdot)$s are smooth functions of the lagged variables to be estimated. The key feature of the model is additivity. Each input feature makes a separate contribution to the response, and these just add up. The smooth functions (splines) f are composed by sums of base functions and their corresponding regression coefficients. The places where the polynomial pieces connect are called knots. The mgcv package in R represents the smooth functions using penalised regression splines to regularise the smoothness. By default, the mgcv procedure uses basis functions for these splines that are designed to be optimal, given the number of functions used. The smoothing parameters are chosen by minimising the generalised cross validation (GCV) score of the whole model (Hastie et al., 2009). All AAR models in this paper are estimated using cubic regression splines.

Forecasting the prices of different dairy commodities in 2016

Four different forecasting models were fitted with data from January 2002 to December 2015 as the training set: the AR, the SETAR, the LSTAR and the AAR. The year 2016 was used as the test set. Due to computational problems, the LSTAR models were dropped for Smp and Wmp. Further, for cheese, the LSTAR, AAR and SETAR models did not perform well, therefore they are not shown.

Forecasts based on the AR models

At time T, the optimal j-step ahead forecast C_{T+j} is the conditional expectation $E(C_{T+j}|Y_T)$, where Y_T denotes the information till time T. For each commodity models with different time lags were tried, and the AIC (Akaike, 1969) and the Bayesian information criterion (BIC) (Schwarz, 1978) were used to choose between models. Model parameters for each commodity are given in the Appendix.

Forecasts based on the nonlinear models

The forecasts were obtained recursively from the estimated model. Ignoring the residuals in the second and more steps ahead forecasts leads to biased forecasts, so-called naive. Therefore, the bootstrap resampling method was applied, where residuals are resampled from the empirical residuals from the model (Franses & van Dijk, 2000). The bootstrap does not require knowledge of the distribution of the residuals. It is therefore more robust and considered more satisfactory on the purely pragmatic grounds of producing better forecasts (Teräsvirta et al., 2011). Let us consider the simple regression model

$$y_{t+1} = g(x_t) + \varepsilon_{t+1} \tag{5}$$

for example,

$$y_{t+1} = x_t^2 + \varepsilon_{t+1} \tag{6}$$

where

$$x_{t+1} = \phi x_t + \eta_{t+1}, \tag{7}$$

and $\{\eta_t\} \sim$ iid (0, σ_η^2).

Suppose we have observations until T and want to forecast y_{T+1}. The one-step forecast becomes

$$f_{T,1}^y = g(x_T), \tag{8}$$

as $E\{\varepsilon_{T+1}\} = 0$, given the information until T. The two-step forecast is not as easy. The optimum two-step forecast is

$$f_{T,2}^y = E\{y_{T+2} \mid \mathcal{F}_T\} = E\{g(x_{T+1}) \mid \mathcal{F}_T\}, \tag{9}$$

where $\mathcal{F}_T$ is the information set. As x_{T+1} is usually not known at time T, it must be forecasted from (7). The one-step OLS forecast becomes

$$f_{T,1}^x = \phi x_T. \tag{10}$$

Expressing (6) in the form

$$f_{T,2}^y = E\{g(f_{T,1}^x + \eta_{T+1}) \mid \mathcal{F}_T\}, \tag{11}$$

the bootstrap method yields the following two-step forecast:

$$fb_{T,2}^y = \frac{1}{N_B} \sum_{j=1}^{N_B} g(f_{T,1}^x + \widehat{\eta}_t^{(j)}), \tag{12}$$

where $\widehat{\eta}_t^{(j)}$, $j = 1, \ldots, N_B$, are the N_B independent draws with replacement from the set of residuals $\{\widehat{\eta}_t\}_{t=2}^T$ estimated from (9) over the sample period. The forecast will be

$$f_{t,2}^y = \phi^2 x_t^2 + \widehat{\sigma}_\eta^2 \text{ , assuming that} \tag{13}$$

$$N_B^{-1} \sum_{j=1}^{N_B} \widehat{\eta}_t^{(j)} \text{ is near zero.} \tag{14}$$

In practice the function $g(f_{T,1}^x + \widehat{\eta}_t^{(j)})$ is not known and must be estimated. Thus, g must be replaced with ĝ in the forecast above. In this paper, the number of bootstrap replications was set to 200.

Testing the models and evaluating the forecasts
A good forecast is one that generates low expected loss when used in economic decisions. The costs of different mistakes – typically different magnitudes of over- and underpredictions of the outcome – must therefore be considered in selecting a forecasting model, estimating its parameters and generating forecasts. A model which fits the training data well will not necessarily forecast well (Hyndman & Athanasopoulos, 2019). Thus, the ultimate test of the models is their predictive accuracy on a test set unknown to the model, a common principle, for example, when comparing different machine learning models. In this study, five different commodities with quite different prices were compared. Percentage errors have the advantage of being unit-free, and so are frequently used to compare forecast performances between datasets. Thus, mean absolute percentage error (MAPE) is the most commonly used measure of forecast accuracy (Hyndman & Athanasopoulos, 2019).

$$\text{MAPE} = 100 \sum_{t=1}^{T} \frac{\left| \frac{y_t - \hat{y}_{t|t-1}}{y_t} \right|}{T}.$$

Root mean squared error (RMSE) was also applied, as it is one of the most commonly used scale-dependent accuracy measures (Hyndman & Athanasopolous, 2019). The RMSE is:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T} e_t^2}{T}}.$$

A difference between the two measures is that because the RMSE squares the errors, it penalises large errors heavier than the MAPE. Similar evaluation metrics as applied in this study are also used in other comparisons of forecasts, see, for example, Guo & Tseng (1997). Plots of the residuals and the BDS test were used to evaluate the models.

The most versatile and widely used forecast accuracy test to compare the quality of forecast between competing models in empirical studies is the Diebold–Mariano (DM) test (Diebold & Mariano, 2002). To compare forecasts, we apply the modified DM version proposed by Harvey *et al*. (1997). The DM test is a formal test to compare predictive accuracy. Thus, the fact that estimated MSE of forecast A in one sample is lower than that of forecast B does not necessarily mean that method A is the best in the population. The DM test aims at answering the question: Is method A truly superior in the population or is it merely lucky given the sample? If we denote the forecast error $L(e_t)$, DM calculates the time t forecast error loss $(\widehat{\text{MSE}})$ differential between two forecasts 1 and 2 as: $d_{12t} = L(e_{1t}) - L(e_{2t})$. The DM test tests the hypothesis of equal expected loss, that is, $d_{12t} = 0$. If two methods are equally good $E(d_{12t}) = 0$, which means that $DM_{12} = \overline{d_{12}} / \widehat{\sigma}_{\overline{d_{12}}} \to N(0, 1)$, where $\overline{d_{12}} = (1/T) \sum_{t=1}^{T} d_{12t}$, $\overline{d_{12}}$ is the sample mean loss differential, and $\widehat{\sigma}_{\overline{d_{12}}}$ is a consistent estimate of the s.d. of $\overline{d_{12}}$.

In this study, we apply the DM test to test out of sample forecast accuracy. Unfortunately, we had problems with negative long-run variance in the DM tests, which is common when dealing with multi-step-ahead predictions in small samples (Harvey *et al*., 1997). Thus, we did not manage to perform the DM tests for a longer forecast period than 6 mo ahead.

All models were estimated using the free statistical software package R, and the libraries *flinear, forecast, mgcv, urca, tsDyn, strucchange* and *carjols*.

Results

Linearity and structural breaks
The test results of the null hypothesis that the time series are linear are given in Tables 2 and 3.
Following the recommendations in Teräsvirta *et al*. (2011), a significance level of 10% was applied. In Tables 2 and 3 we

Table 2: The test statistics and *P*-values for the Tsay test and Teräsvirta tests of H_0 that the price series are linear

Commodity	Tsay test (order = 3)	Teräsvirta test (lag = 3)
Cheese	1.793	2.537**
Butter	7.511***	7.195***
Smp	12.66***	5.648***
Wmp	5.161**	2.833***
Whp	1.605	3.225***

Significance code: **P = 0.01, ***P = 0.001.
Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.

Table 4: The Chow test with the aveF test statistic for a possible structural break in the prices between 2006 and 2008

Commodity	Breakpoint	Test statistic
Whp	February 2006	170.6***
Smp	November 2006	135.49***
Wmp	December 2006	168.69***
Cheese	May 2007	271.56***
Butter	June 2007	346.04***

Significance code: ***P = 0.001.
Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.

Table 3: The BDS test of linearity for the different series: test statistics and *P*-values for different combinations of *m* and *ε*

Commodity	m	ε = 1	ε = 4
Cheese	2	185.404***	41.946***
	5	1,535.229***	41.224***
Butter	2	186.801***	43.667***
	5	1,311.1105***	42.907***
Smp	2	102.250***	33.019***
	5	653.794***	30.774***
Wmp	2	112.147***	36.032***
	5	735.973***	34.281***
Whp	2	360.293***	43.794***
	5	2,519.527***	43.156***

Significance code: ***P = 0.001.
BDS = Brock, Dechert and Scheinkman; Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.

can see that all three tests suggest that the prices of butter, Smp and Wmp are nonlinear. However, for cheese and Whp, the Tsay test does not reject the null hypothesis of linearity. From Figure 1, one can see a possible structural break around 2007. Therefore, the series were tested for a potential breakpoint between 2006 and 2008, and the results of the Chow test are given in Table 4.

The results suggest that the dry products first reached the breakpoint. It was first half a year later butter and cheese followed. Similarly, the CUSUM test for structural change confirms the existence of a structural break in all commodity prices, with test statistics 4.959*** for cheese, 4.404*** for Wmp, 5.188*** for butter, 4.300*** for Whp and 4.136*** for Smp. The essence of linear systems theory is that any stochastic process can be separated into the sum of two processes – a deterministic one that is a linear function of its past values and a stochastic one that is a linear function of previous values of an uncorrelated random variable. However, based on the preliminary tests, the test results of linearity and structural breaks, the main conclusion is that the dairy commodity series may be best predicted using nonlinear models.

The model estimates are provided in Tables A3–A6. The BDS tests applied to the residuals of the best performing models according to MAPE and RMSE are provided in Table A7. The tests show that the hypothesis that they are iid is rejected for most combinations of *m* and *ε* at conventional significance levels. However, plots of the autocorrelation function (acf) and partial autocorrelation function (pacf) show that there is little correlation among them. Plots of the residuals also show that several models have problems with heteroscedasticity from the time of the breakpoints on, with clearly increased absolute values. The quantile plots also show mixed results as to whether the residuals are approximately normally distributed or not.

Commodity price forecasts for 2016

The MAPE and the RMSE of the different forecasting models for 2016 are shown in Figures 2–6.

From Figure 2 we can see that for cheese the forecasts become inaccurate after only a few months. However, after 11–12 mo the AR recovers and yields an MAPE of around 12, which is reasonable.

From Figure 3 we can see that the AR yields the best price forecast for butter, with a reasonable MAPE up to 8–9 mo ahead. After 9 mo the SETAR and the LSTAR produce equally good forecasts.

The Smp price can be forecasted 12 mo ahead with an MAPE of 8 or below (Figure 4), which is pretty good. The flexible AAR produces the best forecast according to the MAPE and RMSE criteria and manages to capture the fluctuations in the Smp price quite well. The SETAR model also yields relatively good forecasts. Similarly, the AAR manages to forecast the Wmp price 8 mo ahead with an MAPE below 8, which is acceptable
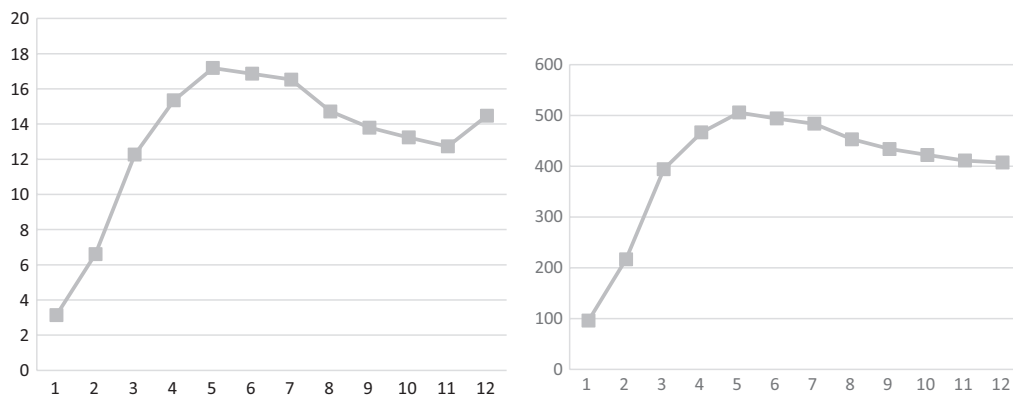
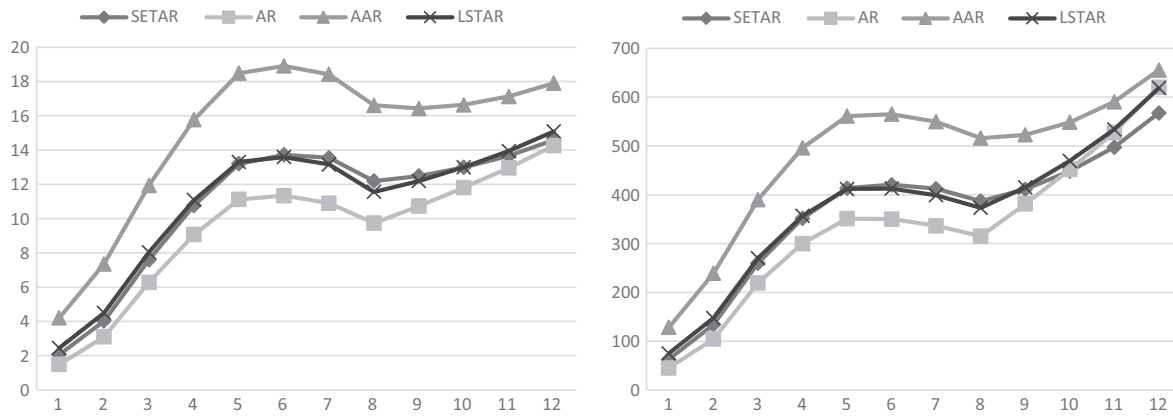**Figure 2.** MAPE (left panel) and RMSE (right panel) for cheese for the AR model.

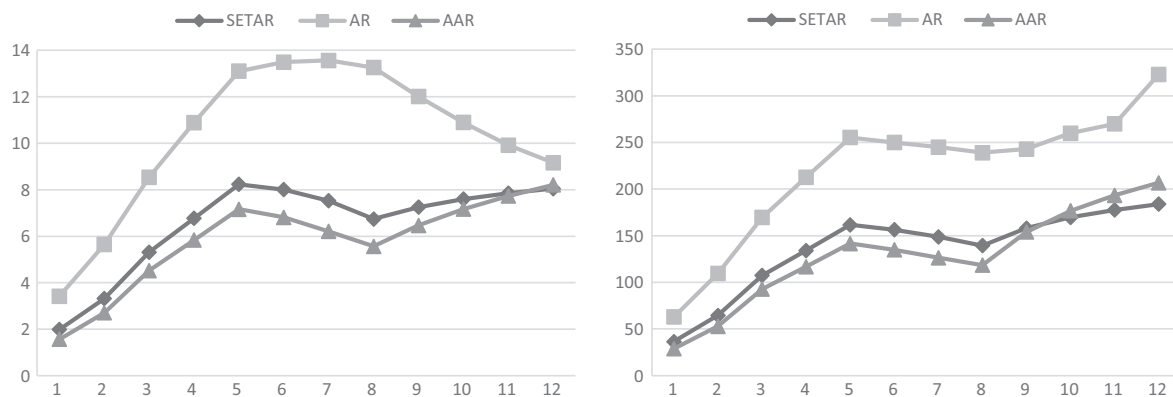**Figure 3.** MAPE (left panel) and RMSE (right panel) for butter for the different models.

Figure 4. MAPE (upper panel) and RMSE (lower panel) for Smp for the different models.
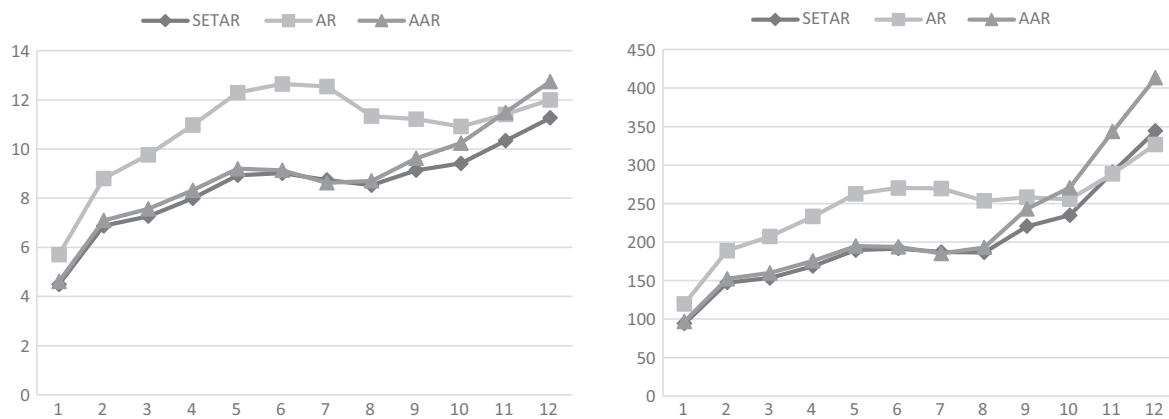
Figure 5. MAPE (left panel) and RMSE (right panel) for Wmp for the different models.
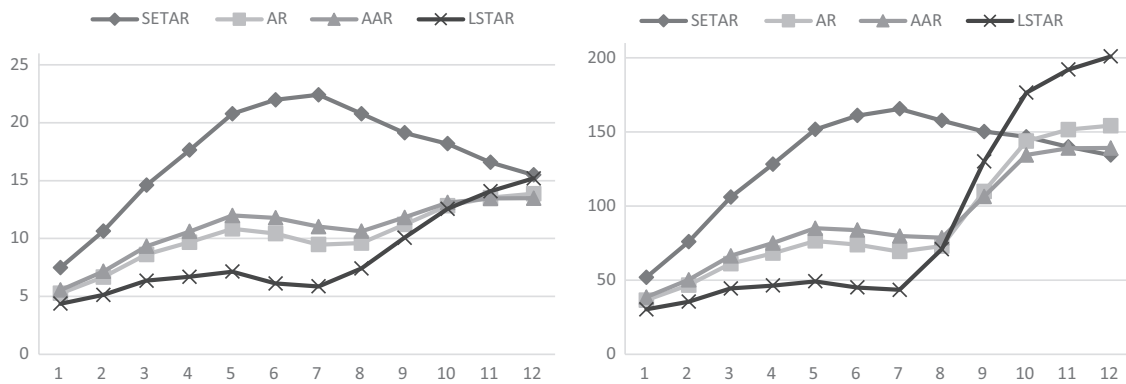

Figure 6. MAPE (left panel) and RMSE (right panel) for Whp for the different models.

(Figure 5). Up to 8 mo the SETAR and the AAR produce equally good forecasts, but then the SETAR takes the lead, with an MAPE below 10 up to 10 mo ahead. However, one should keep in mind that the differences are small, and not significant according to the DM tests.

Finally, the LSTAR produces very good forecasts of the Whp price up to 7 mo ahead (Figure 6), with a MAPE of 6–7. After 8–9 mo the AR and the AAR perform almost equally well.

In Table 5 we show the results from the DM tests of model comparisons.

In Table 5 we can see that the only significant difference between forecasts is the one between the AAR and AR forecasts for Smp. Thus, we can conclude from Figure 4 that for Smp the AAR performs significantly better than the AR.

The prediction intervals for the best models according to MAPE and RMSE are provided in Table A8. Wide intervals reflect a high uncertainty.

Discussion and conclusion

Taken together, the forecast results show that no single model produces by far the best forecast for all products throughout the whole forecasting period. Measured as the MAPE or RMSE several models may perform almost equally well for all products. A model which yields reasonably well in one period may be outperformed by other models in another. Thus, our findings are in line with Elliott & Timmermann (2016) and Bergmann *et al*. (2018) that there is almost never a single forecasting approach that uniformly dominates all other alternatives to forecasting. As the forecast horizon increases the forecast interval increases substantially, and what can be considered a reasonable prediction interval varies between products. Thus, the environment is highly dynamic, and all models struggle to catch up with rapid price changes.

Table 5: The Diebold–Mariano test statistic for comparisons of the different models for out of sample forecast 6 mo ahead

Models compared	DM test statistic			
	B	Smp	Wmp	Whp
AR-LSTAR	1.077			
LSTAR-AAR	−0.236			
LSTAR-SETAR	−0.121			
AR-AAR	−0.088			
AR-SETAR	−0.043			
AAR-SETAR	0.396			
AAR-SETAR		−0.183		
AAR-AR		−2.410*		
SETAR-AR		−1.172		
AAR-SETAR			−0.504	
AR-SETAR			−0.704	
AR-AAR			−0.946	
AR-LSTAR				−0.794
LSTAR-AAR				0.623
LSTAR-SETAR				0.353
AR-SETAR				0.522
SETAR-AAR				0.213
AAR-LSTAR				−0.753

Significance code: *P = 0.05.
AR = autoregressive; AAR = additive autoregressive model; DM = Diebold–Mariano; LSTAR = logistic smooth transition autoregressive; MAPE = mean absolute percentage error; RMSE = root mean squared error; SETAR = self-exciting threshold autoregressive; Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.

Therefore, the forecaster should have several tools available in the forecasting toolbox.

An interesting question is what the prices and forecasts for both Smp and butter would have looked like without the use of market instruments, such as the intervention scheme in the Common Agricultural Policy (CAP). Although the prices analysed here are world market prices, intervention is likely to have had a stabilising effect on the prices of both butter and Smp. The results suggest that dry commodities or powders are relatively easy to forecast as compared to butter and particularly cheese. Of all prices, forecasting the Smp price seems to be the easiest task. Powders have a short production time and are easily tradable. Thus, changes in market prices are quickly transmitted to producer prices. Further, while cheese is a high-end product, powders and butter are both commonly used as regulatory products. This may explain why their prices show larger s.d. as compared to the world cheese price (Table 1). Cheese can be stored for

several years until it reaches the market, and stock levels of cheese are among the highest for dairy products. Thus, for cheese price signals from the market have a significant time lag, which makes it more difficult for both farmers and dairies to adjust the production to current prices. The significant time lag may possibly explain why most models applied in this paper failed to forecast the cheese price. Further, changing consumption patterns to more processed food, and the fact that we use the price of cheddar only as representative for cheese may also make forecasting challenging. Thus, future studies could use, for example, weighted US and EU prices of the most traded cheeses instead. Finally, the fact that major cheese producing countries also import cheese may also influence prices. However, if one accepts a relatively high error rate, a simple AR model can be applied to forecast the cheese price 12 mo ahead.

For butter, according to the MAPE and RSME, the AR model performs the best. However, according to the DM test, the differences between models are not significant. Similarly, for Smp, the AAR model performs the best, and significantly better than the AR. However, the difference between the SETAR and the AAR models is small and not significant. The findings in our study supports the statement by Teräsvirta *et al*. (2011:362): "In some cases the nonlinear models clearly outperform the linear ones, but in other occasions they may be strongly inferior to the latter". In this study, the simple AR (2) model works for all commodities and performs surprisingly well. Thus, the findings also support the statement in (Teräsvirta *et al*., 2011) that sometimes, even when the data-generating process is nonlinear, a linear model can yield more accurate forecasts than the correctly specified nonlinear one. As is argued in Granger & Teräsvirta (1993), the prediction errors generated by a nonlinear model will be smaller only when the nonlinear feature modelled in-sample is also present in the forecasting sample. The success of nonlinear time series in producing better forecasts than linear models depends on how persistent the nonlinearities are in the data. In this dataset inspection of the price series shows many signs of nonlinearity. Nevertheless, the results show that applying an ensemble of different nonlinear and linear models is an efficient way to analyse statistical data, in line with the view held in the community of statistical learning (James *et al*., 2017).

In contrast to the view held by Sumner & Matthews (2016), the findings reported here suggest that forecasting dairy commodity prices can be useful. For large dairy exporting regions or countries like the EU-28, the US and New Zealand, it is crucial to have an idea of how the world market prices of different commodities will develop. Similarly, large importers of dairy products like China, the Russian Federation, Mexico and Japan also have an interest in monitoring the world market prices to make optimal buying decisions. The world market

prices are characterized by large fluctuations and the degree and timing of changes are different. Due to these changes, both sellers and buyers can suffer great losses.

A much wider disparity between the largest and smallest values after the breakpoints in 2006/2007 (Figure 1) is one possible reason why several models face problems with heteroscedasticity. Omitted explanatory variables in the models might be another reason. The dairy sector, from production to final use, including trade and inventory balance, is complex. For example, the issue of storage of dairy commodities raises the question of whether a new variable measuring stocks or stocks relative to demand could improve the forecasts. This could be particularly interesting for cheese. The growing importance of price cycles (Bergmann *et al*., 2015) could engage scholars to improve the forecasting models. Wavelet analysis could be one option worth exploring. Further, linear and nonlinear state space models, which accommodate the treatment of possible inter-relationships between multiple time series, could be another. In an agricultural context, forecast accuracy can also be improved by combining forecasts of individual models (Colino *et al*., 2012; Bergmann *et al*., 2018). When combining forecasts weights based on inverse past forecast errors often outperform more complex methods (Timmermann, 2006). Finally, the predictive capacity of dairy commodity forecast models might also be improved by including the price information inherent in futures. Nevertheless, one should never forget that forecasting models that are simple enough to lend themselves to empirical estimation must be strongly condensed representations of a far more complex, and possibly changing, data-generating process. The correct perspective is therefore to regard all forecasting models as mis-specified.

The underlying presumption behind time series models that correlation between adjacent points in time is best explained in terms of a dependence of the current values on past values means that the models depend heavily on the time periods analysed. China's role as a key importer of many traded dairy products is a key uncertainty in the future developments of world dairy markets. China's milk production has increased, along with investments in processing capabilities. Further, environmental legislation can have strong impacts on the future development of dairy production. Water access and manure management are other areas where policy changes could have an impact. Major outbreaks of animal diseases or unusual weather events during the forecast period could also alter the setting. Finally, dairy demand and export opportunities could also be affected by the outcome of various free trade agreements currently under discussion. These considerations make it necessary to recalibrate and update the models regularly, in line with the recommendations of Stock & Watson (2003).

In conclusion, prices of dairy commodities reached a structural breakpoint in 2006/2007. A combination of linear and nonlinear models is useful in forecasting dairy commodity prices. The cheese price seems to be the most difficult to forecast, but the AR performs reasonably well after 1 yr. When evaluated by the MAPE and the RMSE, the butter price is best forecasted with a simple AR model. Similarly, an AAR model can be applied for Smp, an SETAR model for Wmp and an LSTAR model for Whp. However, an important finding is that several models can produce almost equally good results. Although some models outperform others according to the MAPE and the RMSE, one should keep in mind that the DM tests show that only one of the differences between models is significant. A drawback is of course that the DM tests are limited to the first 6 mo ahead. The Smp price is relatively easy to forecast 10–12 mo ahead. For the other commodities forecast errors are acceptable over a period of 6–8 mo. The findings presented here could be of interest to the dairy industry.

Acknowledgement

References

Ahlquist, R. and Kilian, L. 2010. What do we learn from the price of crude oil futures? *Journal of Applied Econometrics* **25**: 539–573.

Ahlquist, R., Kilian, L. and Vigfusson, R.J. 2013. Forecasting the price of oil. In: "Handbook of Economics Forecasting", Volume 2 (eds. G. Elliot and A. Timmerman), North-Holland, Amsterdam, pages 427–507.

Akaike, H. 1969. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* **21**: 243–247.

Aoki, M. and Havenner, M. 1991. State space modeling of multiple time series. *Econometric Reviews* **10**: 1–59.

Bergmann, D., O'Connor, D.G. and Thümmel, A. 2015. Seasonal and cyclical behaviour of farm gate milk prices. *British Food Journal* **117**: 2899–2913.

Bergmann, D., O'Connor, D.G. and Thümmel, A. 2018. An evaluation of point and density forecasts for selected EU farm gate milk prices. *International Journal of Food and Agricultural Economics* **6**: 23–53.

Bisaglia, L. and Gerolimetto, N. 2014. "Testing for (Non)Linearity in Economic Time Series: a Monte Carlo Comparison". Working Paper Series, N.3. Available online: https://core.ac.uk/download/pdf/31144416.pdf [Accessed 10 September 2019].

Bolotova, Y. 2016. An analysis of milk pricing in the United States dairy industry. *Agribusiness* **33**: 194–208.

Bowman, C. and Husain, A.M. 2004. "Forecasting Commodity Prices; Futures versus Judgment". IMF Working Paper. Available online: https://www.imf.org/external/pubs/ft/wp/2004/wp0441.pdf [Accessed 15 November 2018].

Brock, W.A., Dechert, W.D. and Sheinkman, J.A. 1987. A Test of Independence Based on the Correlation Dimension, SSRI No. 8702. Department of Economics, University of Wisconsin, Madison.

Brock, W.A., Dechert, W.D., Scheinkman, J.A. and LeBaron, B. 1991. "Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence". MIT Press.

Brown, B.W., Durbin, J. and Evans, J.M. 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society Series B* **37**: 149–192.

Chernenko, S.V., Schwarz, K.B. and Wright, J.H. 2004. "The Information Content of Forward and Futures Prices: Market Expectations and the Price of Risk". *International Finance Discussion Papers* No. 808. Available online: https://www.federalreserve.gov/pubs/ifdp/2004/808/ifdp808.pdf [Accessed 13 November 2018].

Chinn, M.D. and Coibion, O. 2013. The predictive content of commodity futures. *The Journal of Futures Market* **34**: 607–636.

Chow, G.C. 1960. Test of equality between sets of coefficients in two linear regressions. *Econometrica* **28**: 591–605.

Colino, E.V., Irwin, S.H., Garcia, P. and Etienne, X. 2012. Composite and outlook forecast accuracy. *Journal of Agricultural and Resource Economics* **37**: 228–246.

Coppola, A. 2008. Forecasting oil price movements: Exploiting the information in the futures market. *Journal of Futures Markets* **28**: 34–56.

Diebold, F. and Mariano, R. 2002. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **20**: 134–144.

Durbin, J. and Koopman, S.J. 2001. "Time Series Analysis by State Space Methods". Oxford University Press, London, 270 pages.

Elliott, G. and Timmermann, A. 2016. "Forecasting in Economics and Finance". Available online: https://escholarship.org/content/qt6z55v472/qt6z55v472.pdf [Accessed 17 September 2019].

Elliott, G., Rothenberg, T.J. and Stock, J.H. 1996. Efficient tests for an autoregressive unit root. *Econometrica* **64**: 813–836.

FAO. 2016. OECD-FAO. "Agricultural Outlook". Available online: http://www.fao.org/3/a-BO101e.pdf [Accessed 13 May 2017].

Fortenbery, T.R. and Zapata, H.O. 1997. An evaluation of price linkages between futures and cash markets for cheddar cheese. *Journal of Futures Markets* **17**: 279–301.

Foster, K.A., Havenner, A.M. and Walburger, A.M. 1995. System theoretic forecasts of weekly live cattle prices. *American Journal of Agricultural Economics* **77**: 1012–1023.

Franses, P.H. and van Dijk, D. 2000. "Nonlinear time series models in empirical finance". Cambridge University Press, London, Cambridge, 277 pages.

Granger, C.W.J. and Teräsvirta, T. 1993. "Modelling Nonlinear Economic Relationships". Oxford University Press, New York, 187 pages.

Guo, M. and Tseng, Y.K. 1997. A comparison between linear and nonlinear forecasts for nonlinear AR models. *Journal of Forecasting* **16**: 491–508.

Hansen, B.E. 1999. Testing for linearity. *Journal of Economic Surveys* **13**: 551–576.

Hansen, B.E. 2012. "Structural Breaks". Presentation held at the Summer school in Economics and Econometrics, University of Crete, July 23–27. Available online: https://www.ssc.wisc.edu/~bhansen/crete/crete5.pdf [Accessed 14 January 2018].

Hansen, B.G. and Li, Y. 2016. An analysis of past world market prices of feed and milk and predictions for the future. *Agribusiness* **33**: 1–19.

Harvey, D, Leybourne, S. and Newbold, P. 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting* **13**: 281–291.

Hastie, T., Tibshirani, R. and Friedman, J. 2009. "The Elements of Statistical Learning", 2nd Edition, Springer, New York, 745 pages.

Hyndman, R. 2014. "Structural breaks". Available online: https://robjhyndman.com/hyndsight/structural-breaks/ [Accessed 17 December 2017].

Hyndman, R. and Athanasopoulos, G. 2019. "Forecasting: Principles and Practice". Available online: https://otexts.com/fpp2/ [Accessed 18 September 2019].

IFCN. 2014. Dairy Report 2014. International Farm Comparison Network Dairy Research Center. Kiel, 215 pages.

James, G., Witten, D., Hastie, T. and Tibshirani, R. 2017. "An introduction to statistical learning". Springer, New York, 426 pages.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics* **54**: 159–178.

Labys, W.C. and Kouassi, E. 1996. Structural time series modeling of commodity price cycles. Research Paper 9602, Regional Research Institute, West Virginia University.

Martin-Rodriguez, G. and Cáceres-Hernàndez, J.J. 2012. Forecasting pseudo-periodic seasonal patterns in agricultural prices. *Agricultural Economics* **43**: 531–544.

Moosa, I. and Al-Loughani, N. 1994. Unbiasedness and time varying risk premia in the crude oil futures market. *Energy Economics* **16**: 99–105.

Myers, R.J., Sexton, R.J. and Tomek, W.G. 2010. A century of research on agricultural markets. *American Journal of Agricultural Economics* **92**: 376–403.

Nicholson, C.F. and Stephenson, M.W. 2015. Milk price cycles in the U.S. dairy supply chain and their management implications. *Agribusiness* **31**: 507–520.

O'Connor, D. and Keane, M. 2011. Empirical issues relating to dairy commodity price volatility. In: "Methods to Analyse Agricultural

Price Volatility" (eds. I. Piot-Lepetit and R. M'Barek), Springer, New York.

Piot-Lepetit, I. and M'Barek, R. 2011. "Methods to Analyze Agricultural Price Volatility". Springer, New York.

Reeve, T. and Vigfusson, R.J. 2011. "Evaluating the Forecasting Performance of Commodity Futures Prices". International Finance Discussion Papers No. 1025. Available online: https://www.federalreserve.gov/pubs/ifdp/2011/1025/ifdp1025.pdf [Accessed 12 November 2018].

Reichsfeld, D.A. and Roache, S.K. 2009. Do commodity futures help forecast spot prices? IMF Working Paper.

SB. 2019. Süddeutsche butter- und käse-börse e.V. Kempten. Available online: https://www.butterkaeseboerse.de/ [Accessed 5 September 2019].

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* **6**: 461–464.

Shumway, R. and Stoffer, D. 2010. "Time Series Analysis and Its Applications", 3rd Edition, Springer, New York, 589 pages.

Stock, J.H. and Watson, M.W. 2003. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature* **41**: 788–829.

Sumner, D.A. and Matthews, W.A. 2016. "When Someone Claims to Know Where Commodity Prices Are Really Heading…Grab Your Wallet and Run". Available online: http://alfalfa.ucdavis.edu/+symposium/2016/PDFfiles/2%20Sumner%20Dan.pdf [Accessed 7 June 2017].

Teräsvirta, T., Lin, C.F. and Granger, C.W.J. 1993. Power of the neural network linearity test. *Journal of Time Series Analysis* **14**: 209–220.

Teräsvirta, T., Tjøstheim, D. and Granger, C.W. 2011. "Modelling Nonlinear Time Series", Oxford University Press, London, 597 pages.

Timmermann, A. 2006. Forecast combinations. In: "Handbook of Economic Forecasting" (eds. G. Elliott, C. Granger and A. Timmermann), North Holland, 1010 pages.

Tomek, W.G. and Robinson, K.L. 2003. "Agricultural Product Prices", 4th Edition. Ithaca, Cornell University Press, 389 pages.

Tsay, R.S. 1986. Nonlinearity tests for time series. *Biometrika* **73**: 461–466.

USDA. 2019. "United States Department of Agriculture Agricultural Marketing Service". Available online: https://www.ams.usda.gov/mnreports/dybintprytd.pdf [Accessed 10 September 2019].

Walburger, A.M. and Foster, K. 1998. Determination of focal pricing regions for U.S. fed cattle. *American Journal of Agricultural Economics* **80**: 84–95.

Yang, J., Bessler, D.A. and Leatham, D.J. 2001. Asset storability and price discovery in commodity futures markets: a new look. *The Journal of Futures Markets* **21**: 279–300.

Appendix Table A1: The Elliott–Rothenberg–Stock point optimal test P_T^τ of unit root with eight lags

Test and significance level	C	B	Smp	Wmp	Whp
			Test statistics		
Level	4.922	5.741	1.669	1.821	5.771
5%			3.17		
10%			4.33		
Trend	4.81	3.903	5.771	2.686	8.708
5%			5.66		
10%			6.86		

Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.

Appendix Table A2: The KPSS test statistic of stationarity with eight lags

Test and significance level	C	B	Smp	Wmp	Whp
			Test statistics		
Level-stationarity	1.175	1.323	0.773	0.867	1.011
5% level			0.463		
10% level			0.347		
Trend stationarity	0.197	0.189	0.142	0.210	0.115
5% level			0.146		
10% level			0.119		

KPSS = Kwiatkowski–Phillips–Schmidt–Shin; Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.

Appendix Table A3: The parameters in the AR models

	Constant	ϕ_1	ϕ_2
Cheese	85.386*	1.621***	−0.645***
(s.e.)	33.785	0.059	0.059
Butter	93.947*	1.312***	−0.342***
(s.e.)	43.463	0.073	0.073
Smp	105.262*	1.372***	−0.408***
(s.e.)	47.614	0.071	0.071
Wmp	118.124*	1.403***	−0.442***
(s.e.)	49.600	0.070	0.070
Whp	25.795*	1.501***	−0.527***
(s.e.)	12.378	0.066	0.066

Significance codes: *P = 0.05, ***P = 0.001.
AR = autoregressive; Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.

Appendix Table A4: The parameters in the SETAR models

	Threshold	Low regime			High regime		
		Constant	ϕ_1	ϕ_2	Constant	ϕ_1	ϕ_2
Butter	1,825	2.168	1.013***		185.947***	1.293***	−0.349***
(s.e.)		166.483	0.116		68.322	0.074	0.073
Smp	2,400	81.721	1.478***	−0.513*	327.469***	1.317***	−0.410***
(s.e.)		151.810	0.223	0.207	110.369	0.077	0.075
Wmp	2,550	77.182	1.561***	−0.587***	298.920*	1.345***	−0.428***
(s.e.)		152.743	0.194	0.180	120.004	0.078	0.076
Whp	735	10.649	1.409***	−0.415*	62.656*	1.497***	−0.551***
(s.e.)		37.413	0.182	0.175	26.670	0.072	0.071

Significance codes: *P = 0.05, ***P = 0.001.
SETAR = self-exciting threshold autoregressive; Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.

Appendix Table A5: The parameters in the LSTAR models

	Smoothing parameter	Low regime			High regime		
		Constant	ϕ_1	ϕ_2	Constant	ϕ_1	ϕ_2
Butter	14.206	−8.058	0.373*	0.655***	137.936	1.121***	−1.191***
(s.e.)		145.719	0.208	0.162	156.188	0.220	0.176
Whp	13.3	35.625***	1.509***	−0.556***	170.166***	−0.105*	
(s.e.)		18.903	0.067	0.064	60.399	0.049	

Significance codes: *P = 0.05, ***P = 0.001.
LSTAR = logistic smooth transition autoregressive; Whp = whey powder.

Appendix Table A6: The parameters in the AAR models, with coefficients for the intercept, together with estimated degrees of freedom (E df) for the two smoothing terms s1 and s2, and reference degrees of freedom (Ref df) for the F- test

	Parametric intercept term	Edf s1	Ref df s1	Edf s2	Ref df s2
Butter	2,852.542***	4.004***	4.970	4.955***	6.010
(s.e.)	15.818				
Smp	2,845.34***	6.547***	7.574	2.593***	3.368
(s.e.)	14.44				
Wmp	2,987***	5.144***	6.238	5.179***	8.717
(s.e.)	16.448				
Whp	970.831***	1.000***	1.000	1.692***	2.128
(s.e.)	4.563				

Significance code: ***P = 0.001.
Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.

Appendix Table A7: BDS tests of independence of the residuals from the best performing models according to the MAPE and RMSE: test statistics and *P*-values for different combinations of *m* and ε

Commodity	m	ε = 1	ε = 4
AR-cheese	2	1.121	0.973
	3	2.297*	3.244**
AR-butter	2	6.448***	2.080*
	3	10.228***	2.643*
AAR-Smp	2	8.286***	4.180***
	3	14.025***	4.247***
SETAR-Wmp	2	6.255***	2.152*
	3	9.326***	2.667**
LSTAR-Whp	2	3.889***	4.799***
	3	4.754***	5.217***

Significance code: *P = 0.05, **P = 0.01, ***P = 0.001.
AR = autoregressive; AAR = additive autoregressive model; BDS = Brock, Dechert and Scheinkman; LSTAR = logistic smooth transition autoregressive; MAPE = mean absolute percentage error; RMSE = root mean squared error; SETAR = self-exciting threshold autoregressive; Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.

Appendix Table A8: Predicted prices with prediction intervals 6 and 12 mo ahead for the best performing models according to the MAPE and RMSE

Commodity	Months ahead	Predicted	Lower 5%	Upper 95%
AR-cheese	3	3,196	2,759	3,616
	6	3,245	2,315	4,113
	12	3,307	2,187	4,456
AR-butter	3	3,170	2,743	3,674
	6	3,135	2,430	3,849
	12	3,094	2,084	4,286
AAR-Smp	3	1,870	1,294	2,350
	6	1,981	832	3,084
	12	2,286	427	4,433
SETAR-Wmp	3	2,175	1,603	2,860
	6	2,324	1,232	3,424
	12	2,459	930	4,103
LSTAR-Whp	3	715	559	846
	6	712	402	975
	12	721	237	1,242

AR = autoregressive; AAR = additive autoregressive model; LSTAR = logistic smooth transition autoregressive; MAPE = mean absolute percentage error; RMSE = root mean squared error; SETAR = self-exciting threshold autoregressive; Smp = skimmed milk powder; Whp = whey powder; Wmp = whole milk powder.