International Journal of Food and Agricultural Economics
ISSN 2147-8988, E-ISSN: 2149-3766
Vol. 6, No. 1, 2018, pp. 23-53

AN EVALUATION OF POINT AND DENSITY FORECASTS FOR SELECTED EU FARM GATE MILK PRICES

Dennis Bergmann
Cork Institute of Technology, Ireland
Email: dpbergmann@t-online.de

Declan O'Connor
Cork Institute of Technology, Ireland

Andreas Thümmel
University of Applied Science Darmstadt, Germany.

Abstract

Fundamental changes to the common agricultural policy (CAP) have led to greater market orientation which in turn has resulted in sharply increased variability of EU farm gate milk prices and thus farmers' income. In this market environment reliable forecasts of farm gate milk prices are extremely important as farmers can make improved decisions with regards to cash flow management and budget preparation. In addition these forecasts may be used in setting fixed priced contracts between dairy farmers and processors thus providing certainty and reducing risk. In this study both point and density forecasts from various time series models for farm gate milk prices in Germany, Ireland and for an average EU price series are evaluated using a rolling window framework. Additionally forecasts of the individual models are combined using different combination schemes. The results of the out of sample evaluation show that ARIMA type models perform well on short forecast horizons (1 to 3 month) while the structural time series approach performs well on longer forecast horizons (12 month). Finally combining individual forecasts of different models significantly improves the forecast performance for all forecast horizons.
Keywords: *Dairy industry, milk prices, forecasting, time series methods, density forecasts*.
JEl Codes: Q11, Q12, Q18

1. Introduction

Fundamental changes to the common agricultural policy (CAP) have led to greater market orientation which in turn has resulted in sharply increased variability of EU farm gate milk prices and thus farm income (Bergmann et al 2015). These changes had their origin in the Luxembourg 2003 agreement. Soon after the Luxembourg 2003 measures were fully implemented in 2007 the average EU farm gate milk prices started to rally from about 27 EUR/100kg to almost 40 EUR/100kg by the beginning of 2008. After this peak prices declined to near 25 EUR/100kg by mid-2009, marking the so called "milk crisis". Following this trough prices started to recover to peak again at about 40 EUR/100kg at the beginning of 2014. The years 2014 and 2015 saw farm gate milk prices falling to about 30EUR/100kg. This variability of milk prices causes serious problems for some farmers and processors as their decisions are usually based on prices in the future rather than current prices. For example farmers might wish to lower production output if prices are expected to decline and increase if they are expected to rise. Furthermore uncertainty may hinder investment and make it more expensive. Thus good forecasts of dairy prices are most valuable as they aid cash flow management and budget preparation. For example Gloy et al. (2002) found dairy farms undertaking cash flow analysis tended to be much more profitable than their peers who did not perform this analysis. In addition reliable price forecasts are of particular importance in the dairy industry as milk is a perishable product and where many decisions are based on a long horizon. For example at farm level there is a natural delay of three or more years between the decision to significantly expand production and the moment production actually starts to rise as a result[1].Likewise the expansion of processing capacity takes years from planning to commissioning.

Besides having good estimates of expected prices in the future it is also important to know the level of uncertainty around these estimates. Forecasting the whole price distribution gives information on the uncertainty of the point forecasts and can thus further improve decision making. For example potential worst case price scenarios[2] and their effect on solvency and liquidity can be identified from the distribution of prices in the future. Density forecasts would also be of great benefit to policymakers as they would allow them to quantify the probabilities associated with adverse market situations and assist in deciding whether to intervene or not. Likewise reliable forecasts could be of use in setting fixed priced contracts between dairy farmers and processors.

Forecasting has a long academic tradition. Allen (1994) provides a summary of agricultural forecasting studies. He founds that in general agricultural economists tend to "*overemphasis on explanation, and have little interest in the predictive power of models*". The importance of price forecasting and evaluation is also highlighted in Hamulczuk et al. (2013) who argue that "*agricultural price forecasting is also a way of gaining a competitive advantage*". In their study they provide a summary of econometric forecasting methods used for forecasting agricultural commodity prices. However despite the importance of price forecasting in agriculture, studies analysing milk price forecasting performance are relatively rare. An exception is the study of Lira (2013) who considers the forecast performance of both Winters' and seasonal autoregressive integrated moving average models (SARIMA) applied to procurement milk prices in Poland. A SARIMA model along with an error correction model (ECM) and wavelet methods are used in the study of Hansen and Li (2016) who forecast world milk prices based on data from the International Farm Comparison Network (IFCN) data. They evaluate the performance of these forecasts using the root mean square error (RMSE) and mean absolute error (MAE). The RMSE is also used to evaluate forecasts of ten different milk markets using single error correction models (ECM) and vector autoregressive models (ECM) in Glauco et al. (2015). They find the ECM performs better than the VAR model. Along with these academic forecasting studies involving dairy prices a number of public bodies, like the U.S. Department of Agriculture (USDA), Food and Agricultural Policy Research Institute (FAPRI) and the OECD/FAO, regularly provide long term forecasts of dairy prices, often up to ten years into the future. The USDA forecasts are based on expert opinion, although they also published forecasts from an econometric model which forecasts prices up to four quarters ahead (see Mosheim, 2012 and more recently MacDonald et al., 2016). FAPRI and OECD/FAO[3] on the other hand have developed partial equilibrium models to provide long term projections of agricultural prices in general and dairy prices in particular. In addition to these public bodies commercial organisations such as Rabobank regularly provide price forecasts of dairy commodity prices[4].

Although the value of interval and probabilistic forecasts in agriculture has long been acknowledged (Bottum, 1966; Timm, 1966; Bressler and King, 1989) most studies in this sector, and the forecasts given by the public bodies mentioned above, are usually based on point forecasts. Exceptions are Isengildina et al. (2004) where interval forecasts for corn and soybeans prices are evaluated, while livestock price interval forecasts are evaluated in Sanders and Manfredo (2003). In a recent study Trujillo-Barrera et al. (2012) extend beyond interval forecasts and evaluates density forecasts of hog futures prices. There are no applications of interval or density forecasts involving dairy prices to the authors' knowledge.

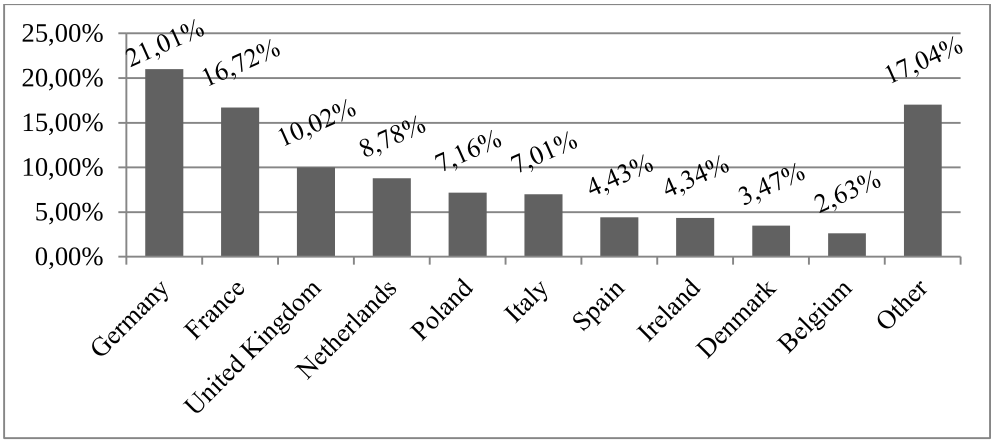
The analysis in this paper contributes to the existing literature as it provides both point and density forecasts of up to one year ahead for farm gate milk prices in Germany, Ireland and for an average EU milk price series. It thus can be seen as a complement to the existing long term projections regularly published by public bodies such as FAPRI and OECD/FAO. Out-of-sample evaluation is conducted using a rolling window framework. To derive these forecasts time series methods are used. These methods include the structural time series approach proposed by Harvey (1989), and applied to EU farm gate milk prices by Bergmann et al. (2015), and to milk prices in the US by Nicholson and Stephenson (2015). In addition SARIMA type models are used. Using SARIMA type models is common practice in forecasting agricultural commodity prices due to the presence of seasonality in many agricultural prices (Hamulczuk et al., 2013). In particular the X12-ARIMA procedure as used by the US census bureau and described in Findley et al. (1998) is applied.

Both the point forecasts and density forecasts are evaluated using a number of scores. For the point forecasts these scores include the root mean square forecast error (RMSE), mean absolute forecast error (MAE), root mean square percentage error (RMSPE) and mean absolute percentage forecast error (MAPE). Furthermore a directional forecast accuracy measure is applied. The ability to forecast the direction of price movements may be more important than minimizing the other scores above because directional accuracy can be directly linked to profits as is argued in Leitch and Tanner (1991). The continuous ranked probability score (CRPS) is used for evaluating density forecasts. Various forecast accuracy tests are also used to complement the scores above. In particular the Diebold and Mariano (1995) test is used to test whether the forecasts from a specific model perform significantly better compared to a naïve no-change benchmark. Directional forecast accuracy is evaluated using the Pesaran and Timmermann (2009) test. Forecast densities on the other hand are evaluated using the tests of Knüppel (2015) as well as Bai and Ng (2005). Both of these tests assess the correct calibration of the forecast density which ensures actual prices are consistent with the forecast distribution.

The remainder of this paper is structured as follows. The next section provides a background to the factors which influence the price formation of farm gate milk prices in Ireland and Germany. After this the models used to derive the forecasts and the methods used for evaluating them are presented. The following two sections present the data used and results obtained. The final section concludes.

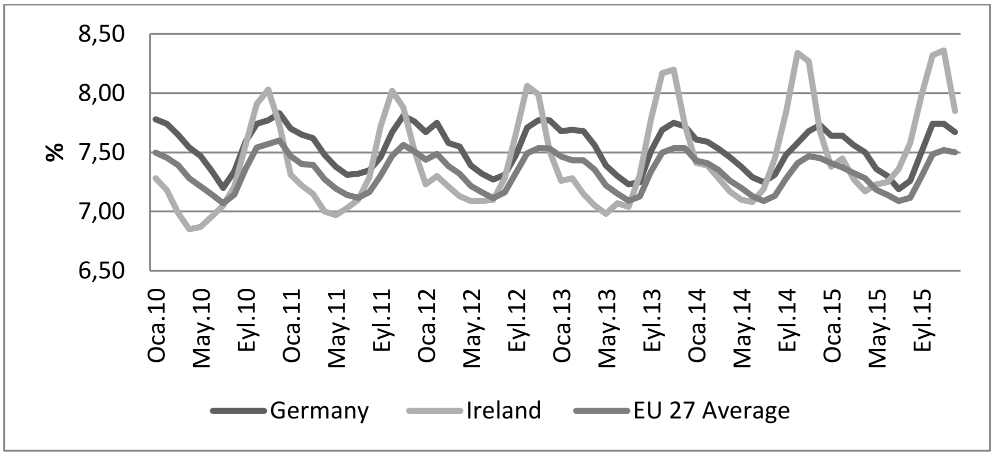
2. German, Irish and Average EU Farmgate Milk Prices

The EU dairy industry is the leading milk producing region in the World accounting for about 24 % of worldwide cow's milk output (International Dairy Federation, 2015).Within the EU milk production is hugely diverse (see Figure 1) with Germany being the largest producer accounting for approximately 21% of EU production in 2014 (IDF 2015).



Source: International Dairy Federation, 2015

Figure 1. Structure of Dairy Milk Production in the EU 2014.



Source: AHDB

Figure 2. Average Milk Solids (Fat and Protein) Contents in Germany, Ireland and EU-27 Average 2010-2015.

While the Irish production is far lower, at 4.3% of EU production, it offers an interesting contrast as its grass based feed system is somewhat unique within the EU. As a consequence Irish milk production is far more seasonal than production in all other member states as it relies on highly seasonal grass growth (Hurtado-Uria et al., 2014). For example the monthly peak to trough ratio in 2015 for Germany was 120 % compared to 760 % for Ireland. This seasonality is further emphasised when average monthly milk solids[5] content of raw milk is considered, Figure 2. While the German solids content is above the EU average it largely moves in tandem with the EU average and both display seasonal characteristics. The Irish milk solids content is more variable, trending upwards and seasonal. As milk solids content in part dictate farm gate prices this would suggest that farm gate price dynamics in Ireland might differ from those of Germany and the EU as a whole.

Table 1. Production of Processed Dairy Products in Germany and Ireland

	Germany (in 1000 tonnes)			
	Whole milk powder	**Skimmed milk powder**	**Butter**	**Cheese**
2006	71	197	384	1,995
2007	65	237	391	1,927
2008	70	227	419	1,941
2009	58	286	410	1,999
2010	60	259	404	2,083
2011		300	425	2,111
2012		314	441	2,161
2013		315	424	1,882
2014		357	441	1,893
	Ireland (in 1000 tonnes)			
	Whole milk powder	**Skimmed milk powder**	**Butter**	**Cheese**
2006	39	69	139	137
2007	34	83	141	140
2008	33	55	124	175
2009	25	75	120	163
2010	34	60	135	172
2011	38	67	146	180
2012	26	52	145	186
2013	35	49	152	183
2014	25	71	166	188

Source: Eurostat

Table 2. Intra and Extra EU Trade of Germany and Ireland in 20015 Source: EU Milk Market Observatory

	Germany 2015		**Ireland 2015**	
	Intra EU Net Export	**Extra EU Export**	**Intra EU Net Export**	**Extra EU Export**
Butter	-1,290	11,947	125,204	21,369
SMP	158,776	147,730	2,890	33,590
Cheese	355,371	93,831	93,916	39,779

A further driver of farmgate milk prices is the end use to which the milk is destined. Table 1 presents the dairy product portfolios in Germany and Ireland. In Germany cheese is the main product. Butter and SMP are produced to a lesser degree although production slightly shifted to these products in recent time while cheese production slightly declined. Ireland on the other hand historically produced mostly butter and SMP but recently has significantly increased its production of cheese. The Irish portfolio is again in part a result of the grass-based feed system which results in highly seasonal milk supply and the need to covert this raw milk into products suitable for storage like butter and SMP (Promar International, 2003). This in turn results in Ireland exporting most of its butter, SMP and cheese output, which can be seen from Table 2. In 2015 Germany was a net importer of butter from other EU member states and only exports a small portion of its butter to third countries. Ireland on the other hand is exporting nearly all of its butter production, most of it within the EU. Germany exports most of its SMP production with about equal portions within the EU and third countries. Ireland is also an exporter of SMP with most of its exports destined outside the EU. Finally both Germany and Ireland cheese export largely are to other member states within the EU with only a smaller portion outside the EU.

3. Methods and Data

A random walk (RW) model often performs well in forecasting studies (Green and Armstrong, 2015). Due to its simplicity, and often good performance, the random walk model is frequently used as a benchmark model against which more complex models are evaluated. The random walk model is described by:

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i \tag{1}$$

where $y_t$is the price at time t. $\varepsilon_i$is a disturbance term which is normally distributed with $\varepsilon_i \sim N(0, \sigma)$.

Autoregressive integrated moving average models (ARIMA) have also been widely applied to forecast agricultural time series (Hamulczuk, 2013). In this paper an ARIMA(1,1,1) model is applied. An ARIMA(1,1,1) model is given by:

$$(1 - \phi B)(1 - B)y_t = c + (1 - \theta B)\varepsilon_t \tag{2}$$

where $y_t$is the price at time t and ε_t is an error term. ϕ, θ and c are parameters to be estimated. B is called the backshift operator which shifts a variable one step back in time ($By_t = y_{t-1}$)[6].

This model can be generalized by including additional lags as well as seasonal lags. The so called seasonal autoregressive integrated moving average model (SARIMA) has been applied to milk prices in the studies of Lira (2013) as well as Hansen and Li (2016). A SARIMA process is described by Hamulczuk, (2013) and can be represented as follows:

$$\begin{aligned}\left(1 - \sum_{i=1}^{p} \phi_i B^i\right)\left(1 - \sum_{i=1}^{P} \Phi_i B^{iS}\right)(1 - B)^d(1 - B^S)^D y_t \\ = c + \left(1 + \sum_{i=1}^{q} \theta_i B^i\right)\left(1 + \sum_{i=1}^{Q} \Theta_i B^{iS}\right)\varepsilon_t\end{aligned} \tag{3}$$

again $y_t$is the price at time t, ε_t is an error term and B the backshift operator. The parameters of the model are ϕ_i, Φ_i, c,θ_i and Θ_i. The above process is integrated of order d with autoregressive order p and moving average order q as well as been seasonally integrated of order D with seasonal autoregressive order P and seasonal moving average order Q. This can be written as SARIMA(p,d,q)(P,D,Q). Here the X-12 ARIMA procedure as developed by the US Census Bureau is used to determine the order of the SARIMA process. The X-12 ARIMA procedure is fully automated and described in Findley et al. (1998).

Finally the structural time series approach introduced by Harvey (1989) is also used to forecast the milk prices. This model was employed by Bergmann et al. (2015)[7] to decompose EU milk prices into their trend, seasonal and cyclical components. It was also used by

Nicholson and Stephenson (2015) for the case of US milk prices. The model is described in Harvey (1989) and Durbin and Koopmann (2012) and can be represented as follows:

$$y_t = \mu_t + \gamma_t + \psi_t + \varepsilon_t \tag{4}$$

where $y_t$is the price at time t, μ_t is the trend, γ_t the seasonal component, $\psi_t$the cycle component and $\varepsilon_i \sim \mathrm{N}(0, \sigma_\varepsilon)$. The trend which represents the long term movement is modelled as a random walk with drift and given by:

$$\mu_{t+1} = \mu_t + \upsilon + \xi_t \tag{5}$$

with drift υ and $\xi_t \sim \mathrm{N}(0, \sigma_\xi)$. The seasonal component which captures cyclical patterns of up to one year is given by:

$$\gamma_{t+1} = -\sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t \tag{6}$$

where s is the number of periods per year ($s = 12$ for monthly data) and $\omega_t \sim \mathrm{N}(0, \sigma_\omega)$ is a normal distributed disturbance term. Due to the error term the magnitude of the seasonal component is allowed to change over time. The cyclical component is related to the seasonal component in so far as it also captures cyclical patterns. However it differs as it can capture cyclical variations longer than one year. It is given by:

$$\begin{pmatrix} \psi_{1,t} \\ \psi_{1,t}^* \end{pmatrix} = \rho \begin{pmatrix} cos\lambda & sin\lambda \\ -sin\lambda & cos\lambda \end{pmatrix} \begin{pmatrix} \psi_{1,t-1} \\ \psi_{1,t-1}^* \end{pmatrix} + \begin{pmatrix} \varpi_t \\ 0 \end{pmatrix} \tag{7}$$

where ρ is a dampening factor with $0 < \rho < 1$, λ the frequency and $\varpi_t \sim \mathrm{N}(0, \sigma_\varpi)$. The length of the cycle is$2\pi/\lambda$. As the trend, seasonal and cyclical component cannot be observed directly the Kalman filter is used to estimate the individual components.

The models described above have been chosen because they are either easy to implement (random walk model and ARIMA(1,1,1)) or because they are able to model trend, seasonal and cyclical components (X12-ARIMA and structural time series approach). According to Bergmann et al. (2015) as well as Nicholson and Stephenson (2015) these components have been shown to be present in EU and US milk prices.

Forecast accuracy can often be improved by combining the forecasts of individual models as stated by Timmermann (2006) and supported in an agricultural context by Colino et al. (2012). When combining individual forecasts simple equal weights or weights based on inverse past forecast errors often outperform more complex methods (see Timmermann, 2006 and the references therein). In this paper six different combination weighting methods are considered. The first is simply the average of the forecasts from the individual models. Here the weight of the forecast from each model is identical and thus:

$$\widehat{\omega}_{t+h,t,i} = \frac{1}{N} \tag{8}$$

The second method developed by Stock and Watson (2001) sets the weights inversely to the mean square error in previous periods (also called the hold out period). Thus the weight of the forecast from the i th model for forecast horizon h at time t is:

$$\widehat{\omega}_{t+h,t,i} = \frac{(1/MSE_{t+h,t,i})}{\sum_{j=1}^{N}(1/MSE_{t+h,t,j})} \tag{9}$$

where $MSE_{t+h,t,i} = 1/\upsilon \sum_{\tau=t-\upsilon}^{t} \varepsilon_{\tau,\tau-h,i}^{2}$ is the past mean squared forecast error. The third combination method sets the weights inversely proportional to the rank of the i th model forecast. This approach was proposed in Aiolfi and Timmermann (2006) and can be represented as:

$$\widehat{\omega}_{t+h,t,i} = \frac{(1/\mathcal{R}_{t+h,t,i})}{\sum_{j=1}^{N}(1/\mathcal{R}_{t+h,t,j})} \tag{10}$$

where $\mathcal{R}_{t+h,t,i}$ denotes the rank of the forecast from the i th model. Thus $\mathcal{R}_{t+h,t,i} = 1$ if the forecast of the i th model was the best in the previous period and $\mathcal{R}_{t+h,t,i} = 2$ for the second best and so on.

Table 3. Summary of the Forecast Models

Model	**Acronym**	**Short description**
Individual models		
Random walk model	RW	Simple random walk model used as benchmark model
ARIMA(1,1,1)	ARIMA(1,1,1)	ARIMA model with lag order p=1, d=1 and q=1
X12 ARIMA	X12ARIMA	X-12 ARIMA procedure used by the US census bureau
Structural time series approach	STSA	Structural time series approach based on Harvey (1989)
Combination models		
Combination Equal Weights	Combination Equal Weights	Combination based on equal weights
Combination Equal Weights w/o worst	Combination Equal Weights w/o worst	Combination based on equal weights neglecting the worst model
Combination Rank RMSE	Combination Rank RMSE	Combinations weights inversely based on the rank of the models RMSE in the three last periods
Combination Rank RMSE w/o worst	Combination Rank RMSE w/o worst	Combinations weights inversely based on the rank of the models RMSE in the three last periods neglecting the worst model
Combination RMSE	Combination RMSE	Combinations weights inversely based on the RMSE in the three last periods
Combination RMSE w/o worst	Combination RMSE w/o worst	Combinations weights inversely based on the RMSE in the three last periods neglecting the worst model

In addition to these three weighting combinations schemes three further combinations are also considered wherein for each case just mentioned the forecast of the worst model is ignored when building the combined forecast.

The hold out period is chosen to be three months which is a good compromise as it allows the weights to react to recent forecast performance while smoothing the weights at the same time[8]. All combinations along with the individual models and a short description of these are summarised in Table 3.

Estimation and out-of-sample forecasts are calculate by applying a rolling window framework with an estimation period of seven years (84 monthly observations) and forecasts with steps of one month, three months, 6 months and 12 months are derived.

4. Forecast Evaluation

In this paper both point forecasts and density forecasts are evaluated. Point forecasts are usually evaluated based on a loss function $L(y_{t+h}, y_{t+h,t})$ of the h step forecast $y_{t+h,t}$ at time t and the actual value y_{t+h} at time $t+h$ (Diebold, 2006). Measures of this nature considered here are the root mean square forecast error (RMSE), mean absolute forecast error (MAE), root mean square percentage error (RMSPE) and mean absolute percentage forecast error (MAPE). These measures are among the most popular measures (Diebold, 2006). In addition a directional forecast accuracy measure (DA) is applied. This measures the percentage of times an upward (downward) movement in the forecast price coincides with as upward (downward) movement in the actual price. The above measures are given by:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_{t+h} - y_{t+h,t})^2} \tag{11}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_{t+h} - y_{t+h,t}| \tag{12}$$

$$RMSPE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\frac{y_{t+h}-y_{t+h,t}}{y_{t+h}})^2} \tag{13}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{|y_{t+h}-y_{t+h,t}|}{|y_{t+h}|} \tag{14}$$

$$DA = \frac{1}{T}\sum_{t=1}^{T}\begin{cases}1\ , if\ sign(\Delta y_{t+h}) = sign(\Delta y_{t+h,t})\\ 0, if\ sign(\Delta y_{t+h}) \neq sign(\Delta y_{t+h,t})\end{cases} \tag{15}$$

where $y_{t+h,t}$ is the h step forecast at time t and y_{t+h} is the actual value at time $t+h$. Δy_{t+h} is the difference between y_{t+h} and y_t.

Next the forecast errors $e_{t+h} = y_{t+h} - y_{t+h,t}$ are tested for potential bias using a simple t-test of the hypothesis that the error have a zero mean using heteroskedasticity consistent standard errors to negate the possibility that the residuals are autocorrelated. In addition the Diebold and Mariano (1995) (DM) test as well as Pesaran and Timmermann (2009) (PT) test are applied. The DM test, tests whether the forecasts of a given model are significantly more accurate than the ones from a comparison model. It should be noted that the DM test was intended to compare forecasts rather than models, and is usually not valid if nested models[9] are compared. However as shown by Giacomini and White (2006) as well as Clark and McCracken (2011) the DM test may be applied to nested models under a rolling window framework. Directional accuracy is tested by the PT test. This test has as its null hypothesis that the forecast $y_{t+h,t}$does not help in predicting y_{t+h}. It is an extension of the well-known Pesaran and Timmermann (1992) test and allows for serial correlation in the forecast errors. Allowing for serial correlation is crucial as forecast errors of h step forecasts may have a moving average structure of order $h-1$ (Diebold, 2006).

The forecast densities are evaluated using the continuous ranked probability score (CRPS). This score is often applied in meteorology science and measures some element of distance between the forecast distribution and the observed value y_{t+h} (Hersbach, 2000). It is has been applied in an economic context in Arora et al. (2013) to evaluate GNP density forecasts and by Panagiotelis and Smith (2008) to evaluate electricity price density forecasts. It reduces to the MAE in the case of deterministic forecasts. This score is given by:

$$CRPS = \frac{1}{T}\sum_{t=1}^{T}\int_{-\infty}^{\infty}(\hat{F}_{t+h,t}^{f}(q) - F_{t+h}^{0}(q))^2 dq$$
$$F_{t+h}^{0} = \begin{cases} 0, for\ q < y_{t+h} \\ 1\ for\ q \geq y_{t+h} \end{cases} \tag{16}$$

where $\hat{F}_{t+h,t}^{f}$ is the forecast cumulative distribution function (cdf) and F_{t+h}^{0} is the observed cdf.

Table 4. Summary of Evaluation Tests

Test	**Acronym**	**Short description**
Point forecasts		
Bias	Bias test	Tests if forecast errors are biased
Diebold and Mariano (1995)	DM	Tests if the forecasts of a given model are significantly more accurate than the ones from a comparison model
Pesaran and Timmermann (2009)	PT	Tests if the forecasts actually do help in predicting the actual series
Density forecasts		
Knüppel (2015) test	Knüppel test	Test of correct calibration of the forecast density
Bai and Ng (2005)	Bai and Ng test	Test of correct calibration of the forecast density

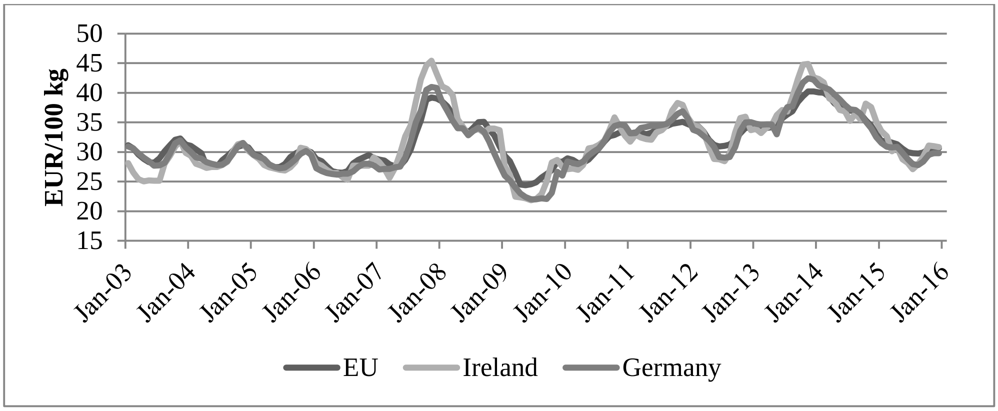
These forecast densities are then further evaluated by applying tests of correct forecast density calibration. These tests are a direct extension of the interval tests as proposed by Christoffersen (1998). The intuition behind the Christoffersen (1998)[10] test is that when evaluating the α % forecast interval and comparing this with the then observed value, the actual value should fall in the forecast interval about α % of the time. Consequently it should be outside the forecast interval in about $1 - \alpha$ % of the time. Forecast density evaluation tests now extend this notion by not only considering if the α % forecast interval is correctly calibrated but also that all intervals are correctly calibrated. These tests rely on the probability integral transform (PIT) of Rosenblatt (1952) which was first used by Diebold et al. (1998) in the context of forecast density evaluation. It is given by:

$$u_{t+h} = \int_{-\infty}^{y_{t+h}} \hat{f}_{t+h,t}(q)dq \tag{17}$$

withy_{t+h} being the actual value at time $t + h$ and $\hat{f}_{t+h,t}$ the forecast density function. As shown in Diebold et al. (1998) if $\hat{f}_{t+h,t}$ is the true density then u_{t+h} is uniformly distributed with $u_{t+h} \sim U(0,1)$. By applying the inverse standard normal cdf Φ^{-1} tou_t, a standard normal distributed $z_{t+h} = \Phi^{-1}(u_{t+h})$ can be constructed (see e. g. Knüppel, 2015). Tests can now be constructed by testing u_{t+h} for uniformity or z_{t+h} for normality. As h step forecast errors may be serially correlated appropriate tests need to take this into account. Two such tests, Knüppel (2015) and Bai and Ng (2005), and are therefore used. All tests used to evaluate the forecasts are summarized in Table 4.

5. Data

The monthly raw milk prices paid to milk producers in Germany and Ireland along with an average EU price are analysed and forecast. These monthly prices are published by the milk market observatory[11] and quoted as EUR/100kg. The average EU price is included as it should reflect broad market movements among all member states and is also a useful reference price for policy makers. This price is a weighted average price of prices in all member states. All prices are based on actual milk fat and protein content and exclude taxes and other costs[12].



Source: EU Milk Market Observatory

Figure 3. German, Irish and Average EU Farm Gate Milk Prices from January 2003 to December 2015.

Table 5. Summary Statistics of Farm Gate Prices and Monthly Returns

	EU	Germany	Ireland
Prices			
Mean	31.66	31.30	31.61
Standard Deviation	3.69	4.56	5.17
Skewness	0.45	0.40	0.62
Kurtosis	2.60	2.81	3.01
Coefficient of variation	12%	15%	16%
Monthly returns			
Mean	0.0060%	0.0215%	0.1698%
Standard Deviation	2.76%	3.60%	4.88%
Skewness	0.28	1.03	0.05
Kurtosis	4.25	5.13	4.16

Note: The Coefficient of variation is not applied to the monthly returns as the means are near zero.

From Figure 3 it can be seen that prior to mid-2007 all series exhibit a clear seasonal pattern which accounts for much of the variation in price. The prices peak at the start of 2008 before decreasing by more than one third by mid-2009. This trough marks the so called "milk crisis". After this prices started to recover and hit other peaks at the end of 2011and at the beginning of 2014 before dipping again in 2015. Comparing the three price series it appears that the Irish price seems to be the most seasonal which is to be expected given Irelands grass-based feed system. From Table 5 we see the Irish farm gate milk price is also the most variable as measured by the coefficient of variation. In addition the mean of the monthly returns for the Irish series is large compared to the others. From Figure 3the average EU and German prices look very similar which is not a surprise giving that Germany is the largest producer in the EU.

6. Results

Point and density forecasts generated by the time series methods are evaluated for the period January 2010 to December 2015 using a rolling window framework with an estimation period of seven years (84 monthly observations). Forecasts horizons of one, three, six and twelve months are considered.

6.1 Evaluation of Point Forecasts of Farm Gate Milk Prices

Results of the point forecasts for the average EU farm gate milk price are presented in Table 6 and Table 7. The Bias test in Table 7 suggests that the null hypothesis of zero mean for the forecast errors cannot be rejected for all models and all forecast horizons. From Table 6 it can be seen that the ARIMA (1,1,1) model and the X12 ARIMA have on average the lowest forecast errors for horizons up to three months as measured by the RMSE, MAE, RMSPE and MAPE. For these horizons they outperform the simple no-change random walk model by about 20 % to 25 %, which is confirmed by the significance of the DM[13] test in Table 7 for the forecast horizon of one month. It should be noted that while errors of less than 2% are reported across all models for the one month horizon these errors grow quickly and consistently over time. For example the RMSPE exceeds 4% at the 3 month horizon, 7.7% at the 6 month horizon and 13.3% at the one year horizon. The forecasts from the structural time series approach (STSA) do not perform significantly better than the RW benchmark for the shorter horizons as suggested by the DM test. For the six months horizon all four models perform equally well. For the twelve months horizon the STSA model does best, although the outperformance is not significant versus the RW benchmark and the RMSPE is in excess of 10%.

The finding of good short term performance for ARIMA type models and in the longer term for the STSA is consistent with Hansen and Li (2016). In that study a SARIMA model performs well for short horizons while the wavelet method is better at longer horizons. Those authors argue that the wavelet method is able to capture the cyclical nature of milk prices and thus better catches the business cycle which is more important in the long run. A similar argument can also be applied here to the STSA thus explaining why it outperforms the ARIMA type models for longer horizons.

From Table 6 and Table 7 it can be observed that the combined forecasts perform well and almost outperform all individual models for all forecast horizons with a noticeable decrease in the percentage errors[14].For example for the best performing combinations the RMSPE now drops to 1.28%, 3.35%, 5.58% and 9.51% as the horizons lengthen. In addition this outperformance is significant against the RW benchmark as confirmed by the DM test in Table 7 confirming (Timmermann, 2006 and Colino et al., 2012).

The directional accuracy measure (DA) in Table 6 in general confirms the result that the X12 ARIMA performs best for short term horizons (in excess of 80%) and that the STSA does well for a one year horizon (circa 72%). In addition the STSA performs well on shorter horizons correctly predicting the direction of market movements more than 70 % of the time. Surprisingly the ARIMA (1,1,1) performs poorly compared to the X12 ARIMA model for one and three months horizons based on this measure while performing well on the other scores. These results are in general confirmed by the PT tests in Table 7. This test suggests that the null can be rejected for the X12 ARIMA at the 1 % significance level for horizons up to three months and ARIMA (1,1,1) for a one month horizon. For longer horizons, significance lowers, and both models do not significantly help in predicting market movements for a twelve month horizon. The PT test for the STSA on the other hand implies that the STSA does predict the direction of market movements for all forecast horizons.

Table 6A. Forecasts Accuracy for EU Farm Gate Prices from January 2010 to December 2015

Forecast horizon	1 Month					3 Month				
Score measure	MAE	MAPE	RMSE	RMSPE	DA	MAE	MAPE	RMSE	RMSPE	DA
RW	0.49	1.46%	0.64	1.90%	N/A	1.35	3.99%	1.69	4.98%	N/A
ARIMA (1,1,1)	0.36	1.09%	0.46	1.37%	71%	1.17	3.49%	1.39	4.16%	64%
X12ARIMA	0.39	1.16%	0.49	1.49%	82%	1.11	3.33%	1.35	4.05%	81%
STSA	0.49	1.50%	0.59	1.82%	75%	1.38	4.16%	1.64	5.05%	74%
Combination Equal Weights	0.33	0.99%	0.43	1.28%	79%	0.89	2.64%	1.12	3.35%	80%
Combination Equal Weights w/o worst	0.33	1.00%	0.43	1.29%	82%	0.89	2.67%	1.13	3.39%	84%
Combination RMSE	0.33	1.00%	0.44	1.31%	81%	0.88	2.65%	1.10	3.32%	81%
Combination RMSE w/o worst	0.33	1.00%	0.44	1.31%	82%	0.89	2.66%	1.12	3.37%	79%
Combination Rank RMSE	0.34	1.01%	0.43	1.29%	82%	0.88	2.62%	1.10	3.28%	84%
Combination Rank RMSE w/o worst	0.34	1.02%	0.44	1.32%	81%	0.89	2.68%	1.13	3.41%	81%

Table 7B. Forecasts Accuracy for EU Farm Gate Prices from January 2010 to December 2015

Forecast horizon	6 month					12 month				
Score measure	MAE	MAPE	RMSE	RMSPE	DA	MAE	MAPE	RMSE	RMSPE	DA
RW	2.50	7.37%	2.92	8.59%	N/A	3.88	11.5%	4.53	13.60%	N/A
ARIMA (1,1,1)	2.33	6.84%	2.79	8.17%	67%	3.77	11.2%	4.53	13.62%	66%
X12ARIMA	2.29	6.78%	2.61	7.77%	66%	3.65	10.8%	4.47	13.38%	59%
STSA	2.43	7.25%	2.97	9.07%	72%	2.98	8.58%	3.65	10.39%	72%
Combination Equal Weights	1.78	5.21%	2.12	6.23%	75%	3.21	9.45%	3.94	11.69%	75%
Combination Equal Weights w/o worst	1.73	5.10%	2.07	6.14%	78%	2.88	8.46%	3.62	10.72%	82%
Combination RMSE	1.67	4.94%	1.97	5.83%	73%	2.65	7.76%	3.38	9.98%	84%
Combination RMSE w/o worst	1.69	5.00%	1.97	5.86%	73%	2.53	7.40%	3.23	9.51%	84%
Combination Rank RMSE	1.67	4.89%	1.99	5.85%	75%	2.79	8.19%	3.47	10.27%	84%
Combination Rank RMSE w/o worst	0.34	1.02%	0.44	1.32%	81%	0.89	2.68%	1.13	3.41%	81%

Note: RW: Random walk model; ARIMA(1,1,1): ARIMA model with order $p = d = q = 1$; X12ARIMA: X12ARIMA procedure; STSA: Structual time series approach; Combination Equal Weights: Combination based on equal weights; Combination Rank RMSE: Weights inversely based on the rank of the models RMSE in the three last periods; Combination RMSE: Weights inversely based on the RMSE in the three last periods; Q1, Q3: first respectively third quartile of forecast errors; w/o worst: neglecting the worst model; the DA score does not apply to the RW model, because it predicts prices will stay the same and thus does not predict direction.

Table 8A. Results of the Bias Test, Diebold and Mariano (1995) test along with the Pesaran and Timmermann (2009) Test Applied to the EU Average Farm Gate Price

Forecast horizon	1 month			3 month		
Score measure	Bias	DM	PT	Bias	DM	PT
RW	-0.33+++	N/A	N/A	-0.24+++	N/A	N/A
ARIMA(1,1,1)	0.02+++	-3.71***	4.07***	-0.04+++	-2.01**	2.26**
X12ARIMA	0.16+++	-2.47***	7.15***	0.35+++	-1.61*	7.71***
STSA	0.43+++	-0.75	4.74***	0.31+++	-0.15	3.67***
Combination Equal Weights	0.06+++	-4.47***	5.88***	0.10+++	-2.91***	5.42***
Combination Equal Weights w/o worst	0.4+++	-4.09***	7.25***	0.37+++	-2.93***	7.35***
Combination RMSE	0.19+++	-4.02***	7.28***	0.52+++	-2.91***	5.93***
Combination RMSE w/o worst	0.39+++	-3.84***	7.81***	0.59+++	-2.93***	5.07***
Combination Rank RMSE	0.30+++	-4.26***	7.07***	0.45+++	-3.00***	6.84***
Combination Rank RMSE w/o worst	0.47+++	-3.92***	6.68***	0.61+++	-2.90***	5.84***

Table 9B. Results of the Bias Test, Diebold and Mariano (1995) test along with the Pesaran and Timmermann (2009) Test Applied to the EU Average Farm Gate Price

Forecast horizon	6 month			12 month		
Score measure	Bias	DM	PT	Bias	DM	PT
RW	-0.19+++	N/A	N/A	-0.07+++	N/A	N/A
ARIMA(1,1,1)	0.05+++	-0.63	2.17**	0.34+++	0.03	1.51**
X12ARIMA	0.51+++	-0.96	1.85**	0.45+++	-0.26	0.65
STSA	0.18+++	0.08	2.75***	-0.53+++	-1.06	4.25***
Combination Equal Weights	0.14+++	-3.09***	3.49***	0.06+++	-3.1***	3.21***
Combination Equal Weights w/o worst	0.1+++	-3.12***	4.45***	0.05+++	-3.55***	3.37***
Combination RMSE	0.36+++	-3.66***	3.61***	-0.13+++	-3.05***	4.35***
Combination RMSE w/o worst	0.38+++	-3.62***	3.33***	-0.13+++	-3.08***	3.65***
Combination Rank RMSE	0.34+++	-3.41***	3.8***	0.01+++	-3.51***	4.17***
Combination Rank RMSE w/o worst	0.36+++	-3.32***	4.01***	0.00+++	-3.46***	3.54***

Note: RW: Random walk model; ARIMA(1,1,1): ARIMA model with order $p = d = q = 1$; X12ARIMA: X12ARIMA procedure; STSA: Structual time series approach; Combination Equal Weights: Combination based on equal weights; Combination Rank RMSE: Weights inversely based on the rank of the models RMSE in the three last periods; Combination RMSE: Weights inversely based on the RMSE in the three last periods; w/o worst: neglecting the worst model; Bias: Bias test of the forecast errors; DM: Diebold Mariano test; PT: Pesaran Timmermann test +, ++ and +++ indicate acceptance of the null at the 1 %, 5% and 10 %; *,** and *** indicate rejection of the null at the 1 %, 5% and 10 %; the DM test does not apply to the RW model, because the RW model is the benchmark model of the test; the PT test does not apply to the RW model, because it predicts prices will stay the same and thus does not predict direction.

From Table 6 it can again be seen that combining models improves DA performance. The directional performance of the combined forecasts is correct in excess of 70 % of the time across all models and horizons which is noteworthy. This is also confirmed by the PT test from Table 7 which shows that the combined forecasts significantly help in predicting market direction for all horizons at a 1 % confidence level. This suggests that farmers, for example, could improve decision making by using these forecasts. For example they could increase output or go long a futures contract[15] when forecasts indicate that prices will rise and reduce or go short in the opposite case.

Table 8 and Table 9 evaluate the point forecasts for the German farm gate milk price. Again all forecast errors show no sign of bias (Table 9). From Table 8 it can be seen that the MAE, MAPE, RMSE and RMSPE estimates for the German series are larger compared to the EU series. The RMSPE is close to 3% for the individual models for the one month horizon and in excess of 14% at the 12 month horizon. This is consistent with Table 5 which showed that the German farm gate price is more variable and thus more difficult to predict. Analysing the forecast performance of the individual models it appears that the STSA model produces the best forecasts although the DM test in Table 9 suggests that this outperformance is only significant for a one month horizon. An explanation consistent with Bergmann et al. (2015) might be that the seasonal and cyclical variations are stronger for the German series and thus the STSA is better able to capture these features. The suitability of the STSA is also confirmed by the directional accuracy measure in Table 8 which is about 70 % for all forecast horizons and the PT test in Table 9 which is significant at the 1 % confidence level for all horizons.

The forecasts from the combined models in Table 8 again confirm that the combined models outperform the individual models. The RMPSE is now as low as 2.74% for the one month horizon and 13.48% for the 12 month horizon. One exception is the equal weighted combination scheme which does not outperform the STSA model for a one year horizon. The outperformance of the combined models is further confirmed by the DM test in Table 9 which suggests that the outperformance of the combined models is significant against the RW benchmark. These combined models also perform very well on the directional accuracy measure with a correct prediction rate of direction of 70 % to 80 %.

Table 10A. Forecasts Accuracy or German Farm Gate Prices from January 2010 to December 2015

Forecast horizon	1 month					3 month				
Score measure	MAE	MAPE	RMSE	RMSPE	DA	MAE	MAPE	RMSE	RMSPE	DA
RW	0.80	2.37%	1.05	3.13%	N/A	2.08	6.16%	2.45	7.18%	N/A
ARIMA(1,1,1)	0.69	2.09%	0.96	2.93%	69%	1.97	5.84%	2.39	7.09%	71%
X12ARIMA	0.70	2.14%	1.00	3.10%	72%	1.93	5.76%	2.38	7.12%	71%
STSA	0.67	2.05%	0.93	2.89%	74%	1.67	5.07%	2.10	6.52%	79%
Combination Equal Weights	0.63	1.90%	0.91	2.77%	78%	1.66	4.93%	2.06	6.12%	79%
Combination Equal Weights w/o worst	0.62	1.88%	0.91	2.76%	75%	1.62	4.80%	2.02	5.97%	80%
Combination RMSE	0.62	1.89%	0.91	2.77%	79%	1.51	4.53%	1.94	5.77%	81%
Combination RMSE w/o worst	0.62	1.89%	0.91	2.78%	78%	1.51	4.51%	1.93	5.75%	81%
Combination Rank RMSE	0.61	1.86%	0.90	2.73%	79%	1.52	4.53%	1.91	5.68%	81%
Combination Rank RMSE w/o worst	0.61	1.86%	0.90	2.74%	76%	1.50	4.47%	1.89	5.63%	80%

Table 11B. Forecasts Accuracy or German Farm Gate Prices from January 2010 to December 2015

Forecast horizon	6 month					12 month				
Score measure	MAE	MAPE	RMSE	RMSPE	DA	MAE	MAPE	RMSE	RMSPE	DA
RW	3.42	10.1%	3.97	11.65%	N/A	5.09	15.17%	5.90	17.89%	N/A
ARIMA(1,1,1)	3.41	9.97%	4.10	11.80%	63%	5.22	15.57%	6.18	18.63%	51%
X12ARIMA	3.37	9.89%	4.01	11.65%	67%	4.92	14.70%	5.70	17.34%	61%
STSA	3.09	9.17%	3.79	11.56%	73%	4.39	12.67%	4.93	14.06%	69%
Combination Equal Weights	2.90	8.49%	3.45	9.99%	75%	4.53	13.44%	5.27	15.81%	70%
Combination Equal Weights w/o worst	2.78	8.10%	3.35	9.59%	76%	4.21	12.47%	4.95	14.80%	77%
Combination RMSE	2.56	7.43%	3.22	9.20%	79%	3.96	11.67%	4.70	13.94%	75%
Combination RMSE w/o worst	2.57	7.42%	3.20	9.11%	79%	3.84	11.29%	4.56	13.48%	75%
Combination Rank RMSE	2.64	7.68%	3.24	9.27%	75%	4.06	12.00%	4.80	14.25%	72%
Combination Rank RMSE w/o worst	2.61	7.57%	3.21	9.16%	78%	3.88	11.44%	4.62	13.70%	74%

Note: RW: Random walk model; ARIMA(1,1,1): ARIMA model with order $p = d = q = 1$; X12ARIMA: X12ARIMA procedure; STSA: Structual time series approach; Combination Equal Weights: Combination based on equal weights; Combination Rank RMSE: Weights inversely based on the rank of the models RMSE in the three last periods; Combination RMSE: Weights inversely based on the RMSE in the three last periods; Q1, Q3: first respectively third quartile of forecast errors; w/o worst: neglecting the worst model; the DA score does not apply to the RW model, because it predicts prices will stay the same and thus does not predict direction.

Table 12A. Results of the Bias Test, Diebold and Mariano (1995) Test along with the Pesaran and Timmermann (2009) Test Applied to German Farm Gate Price

Forecast horizon	1 month			3 month		
Score measure	Bias	DM	PT	Bias	DM	PT
RW	-0.35+++	N/A	N/A	-0.18+++	N/A	N/A
ARIMA(1,1,1)	-0.04+++	-1.16	3.22***	0.07+++	-0.33	3.57***
X12ARIMA	-0.22+++	-0.50	3.83***	-0.15+++	-0.46	4.06***
STSA	-0.13+++	-1.71**	4.07***	-0.09+++	-1.14	4.6***
Combination Equal Weights	-0.22+++	-2.48***	5.08***	-0.11+++	-2.77***	5.39***
Combination Equal Weights w/o worst	-0.41+++	-2.44***	4.57***	-0.31+++	-2.91***	5.94***
Combination RMSE	-0.6+++	-2.43***	5.19***	-0.67+++	-3.14***	7.36***
Combination RMSE w/o worst	-0.64+++	-2.31***	5.04***	-0.76+++	-3.09***	6.72***
Combination Rank RMSE	-0.5+++	-2.82***	5.28***	-0.38+++	-3.2***	7.55***
Combination Rank RMSE w/o worst	-0.62+++	-2.74***	4.61***	-0.52+++	-3.18***	6.4***

Table 13B. Results of the Bias Test, Diebold and Mariano (1995) Test along with the Pesaran and Timmermann (2009) Test Applied to German Farm Gate Price

Forecast horizon	6 month			12 month		
Score measure	Bias	DM	PT	Bias	DM	PT
RW	-0.07+++	N/A	N/A	0.06+++	N/A	N/A
ARIMA(1,1,1)	0.26+++	0.50	1.78**	0.47+++	1.65	-0.03
X12ARIMA	0.02+++	0.25	2.53***	0.2+++	-2.49***	0.89
STSA	-0.22+++	-0.26	3.3***	-0.95+++	-1.11	4.6***
Combination Equal Weights	-0.01+++	-2.4***	3.37***	-0.02+++	-4.46***	3***
Combination Equal Weights w/o worst	-0.12+++	-2.32***	4.35***	0.00+++	-3.35***	3.62***
Combination RMSE	-0.15+++	-3.18***	4.49***	-0.15+++	-3.48***	4.18***
Combination RMSE w/o worst	-0.16+++	-2.92***	5.05***	-0.14+++	-3.16***	3.59***
Combination Rank RMSE	-0.11+++	-2.9***	3.55***	-0.07+++	-3.85***	3.74***
Combination Rank RMSE w/o worst	-0.17+++	-2.75***	5.03***	-0.07+++	-3.35***	3.41***

Note: RW: Random walk model; ARIMA(1,1,1): ARIMA model with order $p = d = q = 1$; X12ARIMA: X12ARIMA procedure; STSA: Structual time series approach; Combination Equal Weights: Combination based on equal weights; Combination Rank RMSE: Weights inversely based on the rank of the models RMSE in the three last periods; Combination RMSE: Weights inversely based on the RMSE in the three last periods; w/o worst: neglecting the worst model; Bias: Bias test of the forecast errors; DM: Diebold Mariano test; PT: Pesaran Timmermann test +, ++ and +++ indicate acceptance of the null at the 1 %, 5% and 10 %; *,** and *** indicate rejection of the null at the 1 %, 5% and 10 %; the DM test does not apply to the RW model, because the RW model is the benchmark model of the test; the PT test does not apply to the RW model, because it predicts prices will stay the same and thus does not predict direction .

Results for the Irish farm gate price are reported in Table 10 and Table 11 As with both other series the forecast errors of all models show no sign of bias (Table 11). Comparing the forecast error scores in Table 10 to those of the German series, it can be seen that for one to three months horizons the errors are slightly higher for the Irish series as measured by the MAPE and RMSPE for the better performing models. For six and twelve months horizons the MAPE and RMSPE are approximately of the same magnitude as the German series. From Table 10 it can be observed that STSA outperforms all other individual models similar to the results for the German series. In contrast to the German series these outperformances are significant when considered against the RW benchmark for all horizons at the 5 % confidence level as shown by the DM test in Table 11. Also in contrast to the EU and German series the ARIMA does not outperform the RW benchmark for short horizons. This suggests the ARIMA does equally well as the RW for the one month horizon but worse for a three month horizon. The X12 ARIMA model on the other hand performs well for short horizons. This result is somewhat expected given that the Irish price can be considered the most seasonal. The good performance of the STSA is further confirmed by the directional accuracy measure which is above 70 % for all forecast horizons. This is significant at the 1 % level as suggested by the PT test in Table 11.

Table 14A. Forecasts Accuracy for Irish Farm Gate Prices from January 2010 to December 2015

Forecast horizon	1 month					3 month				
Score measure	MAE	MAPE	RMSE	RMSPE	DA	MAE	MAPE	RMSE	RMSPE	DA
RW	1.19	3.49%	1.49	4.34%	N/A	2.81	8.20%	3.31	9.60%	N/A
ARIMA(1,1,1)	1.15	3.37%	1.41	4.10%	64%	2.85	8.30%	3.38	9.80%	60%
X12ARIMA	0.98	2.91%	1.27	3.78%	76%	2.08	6.22%	2.70	8.12%	74%
STSA	0.95	2.85%	1.17	3.53%	79%	1.82	5.49%	2.30	7.07%	84%
Combination Equal Weights	0.92	2.70%	1.18	3.45%	81%	2.07	6.08%	2.47	7.28%	83%
Combination Equal Weights w/o worst	1.01	2.98%	1.23	3.61%	74%	2.00	5.91%	2.42	7.22%	80%
Combination RMSE	0.92	2.73%	1.16	3.42%	79%	1.77	5.26%	2.22	6.65%	83%
Combination RMSE w/o worst	0.98	2.90%	1.20	3.53%	78%	1.78	5.32%	2.25	6.77%	80%
Combination Rank RMSE	0.94	2.79%	1.18	3.47%	79%	1.84	5.44%	2.24	6.65%	83%
Combination Rank RMSE w/o worst	0.99	2.93%	1.21	3.58%	78%	1.82	5.41%	2.24	6.71%	80%

Table 15B. Forecasts Accuracy for Irish Farm Gate Prices from January 2010 to December 2015

Forecast horizon	6 month					12 month				
Score measure	MAE	MAPE	RMSE	RMSPE	DA	MAE	MAPE	RMSE	RMSPE	DA
RW	4.14	12.14%	4.92	14.49%	N/A	5.08	14.85%	5.83	17.11%	N/A
ARIMA(1,1,1)	4.30	12.60%	5.19	15.25%	54%	5.28	15.43%	6.13	17.90%	51%
X12ARIMA	3.44	10.08%	4.17	12.22%	61%	5.18	15.19%	6.09	18.00%	43%
STSA	3.05	8.99%	3.78	11.22%	82%	4.23	11.91%	5.06	13.84%	75%
Combination Equal Weights	3.21	9.34%	3.71	10.79%	75%	4.41	12.74%	5.28	15.13%	70%
Combination Equal Weights w/o worst	3.09	9.07%	3.57	10.47%	73%	4.40	12.69%	5.13	14.61%	69%
Combination RMSE	2.49	7.20%	3.03	8.66%	76%	3.91	11.08%	4.66	12.76%	72%
Combination RMSE w/o worst	2.50	7.25%	3.04	8.76%	78%	3.93	11.14%	4.64	12.73%	69%
Combination Rank RMSE	2.77	8.04%	3.25	9.37%	72%	4.14	11.89%	4.86	13.71%	74%
Combination Rank RMSE w/o worst	2.68	7.80%	3.20	9.25%	73%	4.10	11.77%	4.78	13.41%	70%

Note: RW: Random walk model; ARIMA(1,1,1): ARIMA model with order $p = d = q = 1$; X12ARIMA: X12ARIMA procedure; STSA: Structual time series approach; Combination Equal Weights: Combination based on equal weights; Combination Rank RMSE: Weights inversely based on the rank of the models RMSE in the three last periods; Combination RMSE: Weights inversely based on the RMSE in the three last periods; Q1, Q3: first respectively third quartile of forecast errors; w/o worst: neglecting the worst model; the DA score does not apply to the RW model, because it predicts prices will stay the same and thus does not predict direction.

Table 16A. Results of the Bias test, Diebold and Mariano (1995) Test along with the Pesaran and Timmermann (2009) Test Applied to the Irish Farm Gate Price

Forecast horizon	1 month			3 month		
Score measure	Bias	DM	PT	Bias	DM	PT
RW	-0.23+++	N/A	N/A	-0.25+++	N/A	N/A
ARIMA(1,1,1)	0.01+++	-1.15	2.25**	-0.05+++	0.49	1.70**
X12ARIMA	-0.88+++	-1.66**	4.87***	-0.7+++	-1.58*	3.76***
STSA	0.04+++	-2.65***	5.84***	0.22+++	-2.65***	6.12***
Combination Equal Weights	-0.29+++	-3.92***	6.16***	-0.24+++	-3.73***	6.4***
Combination Equal Weights w/o worst	-0.57+++	-2.93**	4.48***	-0.77+++	-3.21***	5.8***
Combination RMSE	-0.65+++	-3.3***	5.61***	-0.86+++	-3.59***	6.88***
Combination RMSE w/o worst	-0.73+++	-2.84**	5.49***	-1.12+++	-3.27***	5.72***
Combination Rank RMSE	-0.65+++	-3.38***	5.64***	-0.65+++	-3.85***	6.88***
Combination Rank RMSE w/o worst	-0.81+++	-2.83***	5.49***	-0.96+++	-3.47***	5.91***

Table 17B. Results of the Bias test, Diebold and Mariano (1995) Test along with the Pesaran and Timmermann (2009) Test Applied to the Irish Farm Gate Price

Forecast horizon	6 month			12 month		
Score measure	Bias	DM	PT	Bias	DM	PT
RW	-0.13+++	N/A	N/A	0.03+++	N/A	N/A
ARIMA(1,1,1)	0.13+++	2.15	0.60	0.42+++	2.27	0.07
X12ARIMA	-0.26+++	-1.28	1.35*	0.24+++	1.14	-0.97
STSA	0.16+++	-2.02**	5.67***	-0.29+++	-0.82	3.80***
Combination Equal Weights	-0.03+++	-4.57***	3.56***	0.12+++	-1.76**	2.57***
Combination Equal Weights w/o worst	-0.13+++	-4.16***	4.08***	0.04+++	-1.66**	2.52***
Combination RMSE	-0.41+++	-4.89***	4.25***	-0.23+++	-1.75**	2.78***
Combination RMSE w/o worst	-0.33+++	-4.64***	4.74***	-0.31+++	-1.61**	2.52***
Combination Rank RMSE	-0.1+++	-4.91***	3.2***	-0.04+++	-1.9**	3.22***
Combination Rank RMSE w/o worst	-0.16+++	-4.56***	3.12***	-0.11+++	-1.76**	2.96***

Note: RW: Random walk model; ARIMA(1,1,1): ARIMA model with order $p = d = q = 1$; X12ARIMA: X12ARIMA procedure; STSA: Structual time series approach; Combination Equal Weights: Combination based on equal weights; Combination Rank RMSE: Weights inversely based on the rank of the models RMSE in the three last periods; Combination RMSE: Weights inversely based on the RMSE in the three last periods; w/o worst: neglecting the worst model; Bias: Bias test of the forecast errors; DM: Diebold Mariano test; PT: Pesaran Timmermann test +, ++ and +++ indicate acceptance of the null at the 1 %, 5% and 10 %; *,** and *** indicate rejection of the null at the 1 %, 5% and 10 %; the DM test does not apply to the RW model, because the RW model is the benchmark model of the test; the PT test does not apply to the RW model, because it predicts prices will stay the same and thus does not predict direction.

The results for the combined forecasts for the Irish series are similar to the results of both of the other series. This means combining forecasts of the individual models can improve forecast performance as can be seen from Table 10. This is supported by the DM test which is significant at least at the 5 % confidence level for all combined forecasts and horizons indicating that the combined forecasts significantly outperform the RW benchmark. In addition the combined forecasts also do very well on the directional accuracy measure (69% or better in all cases) and the corresponding PT test which is significant at the 1 % level for all combination methods and forecast horizons.

Comparing the scores across the three series one can see that the forecast errors of the point forecasts for the EU average farm gate price is much smaller compared to the other two farm gate prices. For example on a twelve month horizon the RMSPE measure of the best model is slightly less than 10 %, while it is about 13 % to 14 % for the German and Irish series respectively.

6.2 Evaluation of Density Forecasts of Farm Gate Milk Prices

Density forecasts provide additional information with regards to the uncertainty of point forecasts and can thus further improve decision making. Selected density forecasts for the EU average price for a one, three, six and twelve month horizon are shown in Figure 4[16]. In particular the plots show the 5 % and 95 % quantiles (light grey area) as well as the 25 % and 75 % quantiles (dark grey area) and are commonly referred to as fan charts. In addition the point forecast (solid line) and the actual price (dashed line) are also shown. From these charts it can be seen that the forecast distribution gets wider with increasing forecast horizons. This is expected as the uncertainty of forecasts generally increases the further into the future the forecasts go.

The density forecasts from the ARIMA and X12ARIMA model, as well as from the combinations models, produce a jagged density forecasts. For the ARIMA and X12ARIMA model this can be explained by the fact that the density forecasts depend heavily on the lagged series and errors which change at each time step. For the density forecasts of the combination models the ragged shape of the density forecasts may be explained by the fact the weights of the individual models may change in each period.

The STSA model on the other hand appears to overestimate the seasonal effect at the start of the forecast period for the three and six month horizons. This can be seen by the forecasted seasonal peaks up to 2012/2013. The actual prices do not display similar peaks. A further point of note is that the density forecasts from the ARIMA and X12ARIMA model are very wide for six and twelve month forecast horizons. Given these wide bands it should be noted that the forecasts do not account for a potential minimum price floor as implied by EU intervention purchasing in butter and SMP markets. This potential minimum price floor is not explicitly modelled in the forecasts as there may be occasions when EU butter and SMP prices may fall below its intervention prices. This may for example be because these commodities do not fulfil the required specification eligible for intervention[17] and because intervention scheme volumes[18] are limited.

Sample density forecasts for the German farm gate milk price are shown in Figure 5 while the density forecasts of the Irish milk price are presented in Figure 6. In general the same general conclusions from the graphical analysis of the EU density can be drawn for the density forecasts of these series. An additional observation is that the width of the density forecasts for the German and Irish prices from the X12ARIMA models are even wider compared to the density forecast from the same model for the EU price at the twelve month horizon implying that these forecasts may be of very limited use.

To decide which models perform best the continuous ranked probability score (CRPS) is applied to the density forecasts and presented in Table 12. These results generally confirm the results of the point forecast evaluation. The ARIMA and X12ARIMA perform well for shorter forecast horizons while the STSA performs better for longer forecast horizons. Again model combinations outperform the individual models. To further evaluate the density forecasts the

Bai and Ng (2005) test as well as the Knüppel (2015) test of correct calibration are applied in Table 13. From this table it can be seen that almost all models produce plausible density forecasts for all horizons and for all three farm gate prices at the 5 % confidence level or better. Exceptions were the X12ARIMA density forecast for the EU average price at a one month horizon and the X12ARIMA density forecast for the German series at a six month horizon where the null of correct calibration cannot be accepted at the 5 % level. For the Irish series the null of correct calibration cannot be accepted for the ARIMA and X12 ARIMA models at a three month horizon, the combination model with equal weights at a one and three month horizons as well as the combination model with weights inverse to the RMSE (Combination RMSE) at a one month horizon. The results of the test therefore somehow differ from the graphical analysis given that one may have concluded that the X12ARIMA produces density forecasts which are too wide to be of practical use.

Table 18. Continuous Ranked Probability Score (CRPS) of EU, German and Irish Farm Gate Price One Month, Three Month, Six Month and Twelve Month Ahead Forecasts

	Continuous ranked probability score											
	EU				Germany				Ireland			
Forecast horizon	1 month	3 month	6 month	12 month	1 month	3 month	6 month	12 month	1 month	3 month	6 month	12 month
Individual models												
RW	0.38	0.97	1.74	2.74	0.59	1.43	2.37	3.56	0.85	1.93	2.88	3.38
ARIMA(1,1,1)	0.31	0.83	1.61	2.60	0.51	1.38	2.42	3.75	0.83	2.11	3.30	3.79
X12ARIMA	0.29	0.80	1.54	2.58	0.54	1.38	2.32	3.34	0.73	1.56	2.41	3.74
STSA	0.34	0.97	1.77	2.16	0.51	1.18	2.18	2.95	0.67	1.32	2.14	2.90
Combination models including all four individual models												
Combination Equal Weights	0.24	0.63	1.25	2.35	0.47	1.21	2.10	3.32	0.68	1.47	2.31	3.24
Combination Equal Weights w/o worst	0.25	0.63	1.20	2.08	0.47	1.15	1.98	2.96	0.72	1.41	2.15	3.12
Combination RMSE	0.25	0.64	1.14	1.97	0.48	1.10	1.89	2.86	0.68	1.28	1.79	2.76
Combination RMSE w/o worst	0.26	0.65	1.14	1.86	0.48	1.09	1.85	2.71	0.70	1.29	1.79	2.72
Combination Rank RMSE	0.25	0.62	1.15	2.04	0.47	1.09	1.92	2.91	0.68	1.30	1.93	2.96
Combination Rank RMSE w/o worst	0.26	0.64	1.16	1.91	0.47	1.07	1.87	2.74	0.70	1.29	1.87	2.88

Note: RW: Random walk model; ARIMA(1,1,1): ARIMA model with order $p = d = q = 1$; X12ARIMA: X12ARIMA procedure; STSA: Structual time series approach; Combination Equal Weights: Combination based on equal weights; Combination Rank RMSE: Weights inversely based on the rank of the models RMSE in the three last periods; Combination RMSE: Weights inversely based on the RMSE in the three last periods; w/o worst: neglecting the worst model.

Table 19A. Bai And Ng (2005) Test of Correct Calibration of EU, German and Irish Farm Gate Price Density Forecasts

	Bai and Ng test											
	EU				Germany				Ireland			
Forecast horizon	1 month	3 month	6 month	12 month	1 month	3 month	6 month	12 month	1 month	3 month	6 month	12 month
Individual models												
RW	0.99 +++	1.69 +++	2.32 +++	1.88 +++	3.51 +++	2.30 +++	2.28 +++	1.69 +++	2.95 +++	3.40 +++	3.26 +++	1.72 +++
ARIMA (1,1,1)	1.32 +++	2.20 +++	1.77 +++	1.49 +++	2.94 +++	2.95 +++	2.67 +++	1.47 +++	3.40 +++	6.25 +	3.76 +++	2.07 +++
X12 ARIMA	2.12 +++	4.00 +++	3.93 +++	1.27 +++	2.80 +++	4.51 +++	2.91 +++	1.55 +++	1.64 +++	2.87 +++	1.91 +++	1.72 +++
STSA	1.96 +++	1.89 +++	1.77 +++	0.86 +++	3.06 +++	0.68 +++	1.07 +++	1.38 +++	3.36 +++	2.38 +++	2.43 +++	2.96 +++
Combination models including all four individual models												
Combination Equal Weights	0.60 +++	1.33 +++	1.68 +++	0.92 +++	4.57 ++	2.42 +++	1.92 +++	1.47 +++	0.55 +++	2.93 +++	3.80 +++	1.04 +++
Combination Equal Weights w/o worst	0.18 +++	0.22 +++	1.63 +++	2.72 +++	5.08 ++	2.53 +++	2.78 +++	1.21 +++	1.90 +++	1.84 +++	5.23 ++	1.30 +++
Combination RMSE	1.65 +++	0.32 +++	1.68 +++	2.81 +++	5.32 ++	2.54 +++	2.95 +++	0.94 +++	1.73 +++	0.59 +++	2.41 +++	1.59 +++
Combination RMSE w/o worst	0.79 +++	1.09 +++	2.22 +++	3.63 +++	5.29 ++	2.07 +++	4.32 +++	0.81 +++	2.56 +++	0.41 +++	2.27 +++	1.75 +++
Combination Rank RMSE	0.99 +++	0.39 +++	1.57 +++	2.15 +++	5.30 ++	2.59 +++	2.33 +++	1.23 +++	0.22 +++	0.85 +++	2.97 +++	1.44 +++
Combination Rank RMSE w/o worst	0.78 +++	0.85 +++	2.26 +++	3.21 +++	5.21 ++	2.28 +++	3.58 +++	1.04 +++	0.92 +++	0.58 +++	2.99 +++	1.77 +++

Table 13B. Knüppel (2015) Test of Correct Calibration of EU, German and Irish Farm Gate Price Density Forecasts

	Knüppel test											
	EU				Germany				Ireland			
Forecast horizon	1 month	3 month	6 month	12 month	1 month	3 month	6 month	12 month	1 month	3 month	6 month	12 month
Individual models												
RW	7.59 +++	1.04 +++	3.47 +++	3.50 +++	5.25 +++	3.45 +++	3.15 +++	3.89 +++	3.46 +++	3.66 +++	4.12 +++	2.89 +++
ARIMA (1,1,1)	4.74 +++	7.87 ++	4.94 +++	1.81 +++	6.81 +++	5.10 +++	4.28 +++	4.59 +++	8.43 ++	13.4	8.46 ++	4.14 +++
X12 ARIMA	10.7 +	9.31 ++	7.11 +++	3.55 +++	9.33 ++	8.89 ++	9.53 +	5.02 +++	8.23 ++	12.4 +	8.47 ++	5.53 +++
STSA	2.06 +++	6.03 +++	4.83 +++	1.46 +++	5.02 +++	0.23 +++	4.22 +++	4.00 +++	2.99 +++	2.73 +++	0.10 +++	2.89 +++
Combination models including all four individual models												
Combination Equal Weights	3.30 +++	3.39 +++	2.58 +++	1.96 +++	2.70 +++	6.61 +++	4.14 +++	4.70 +++	12.6 +	9.79 +	6.24 +++	1.81 +++
Combination Equal Weights w/o worst	8.08 ++	1.40 +++	2.25 +++	2.98 +++	4.62 +++	4.98 +++	3.96 +++	3.54 +++	8.84 ++	8.88 ++	7.79 ++	2.23 +++
Combination RMSE	6.94 +++	0.54 +++	1.89 +++	5.04 +++	2.68 +++	4.06 +++	3.91 +++	4.12 +++	11.9 +	3.4 +++	2.87 +++	2.93 +++
Combination RMSE w/o worst	8.18 ++	1.20 +++	2.78 +++	4.58 +++	4.19 +++	2.77 +++	3.52 +++	3.68 +++	6.70 +++	3.67 +++	3.32 +++	3.30 +++
Combination Rank RMSE	6.86 +++	1.26 +++	1.81 +++	3.75 +++	3.49 +++	4.76 +++	3.83 +++	5.75 +++	7.70 +++	5.62 +++	5.94 +++	2.73 +++
Combination Rank RMSE w/o worst	9.21 ++	1.03 +++	3.12 +++	3.72 +++	5.35 +++	3.18 +++	3.18 +++	4.68 +++	5.26 +++	3.08 +++	3.07 +++	2.86 +++

Note: RW: Random walk model; ARIMA(1,1,1): ARIMA model with order $p = d = q = 1$; X12ARIMA: X12ARIMA procedure; STSA: Structual time series approach; Combination Equal Weights: Combination based on equal weights; Combination Rank RMSE: Weights inversely based on the rank of the models RMSE in the three last periods; Combination RMSE: Weights inversely based on the RMSE in the three last periods; w/o worst: neglecting the worst model; Bai and Ng test: Bai and Ng (2005) test; Knüppel test: Knüppel (2015) test; +, ++ and +++ indicate acceptance of the null at the 1 %, 5% and 10 %;

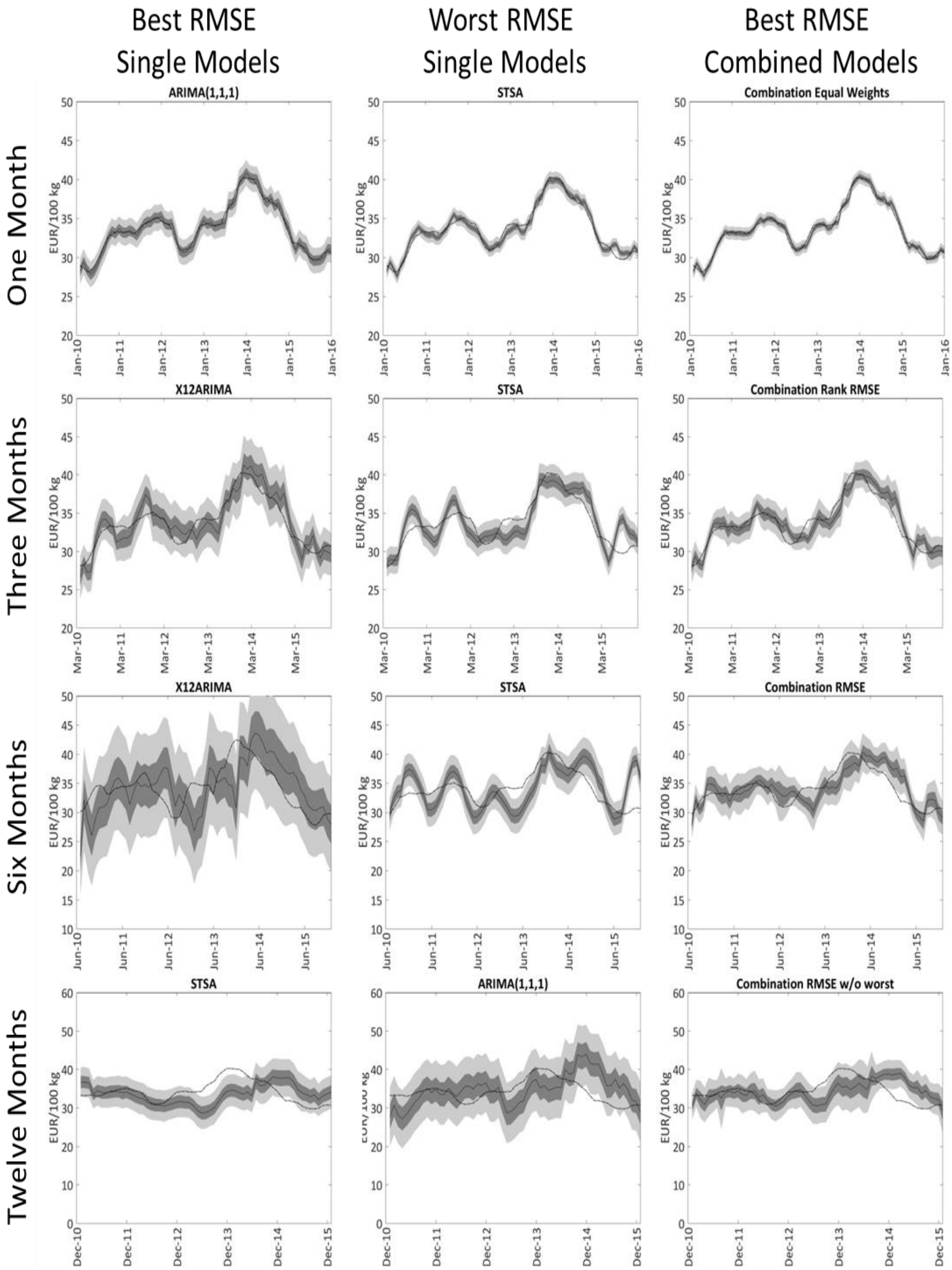


Figure 4. Selected Density Forecasts of EU Average Farm Gate Prices (Grey Shaded Area) Compared to Actual Prices (Dashed Line).

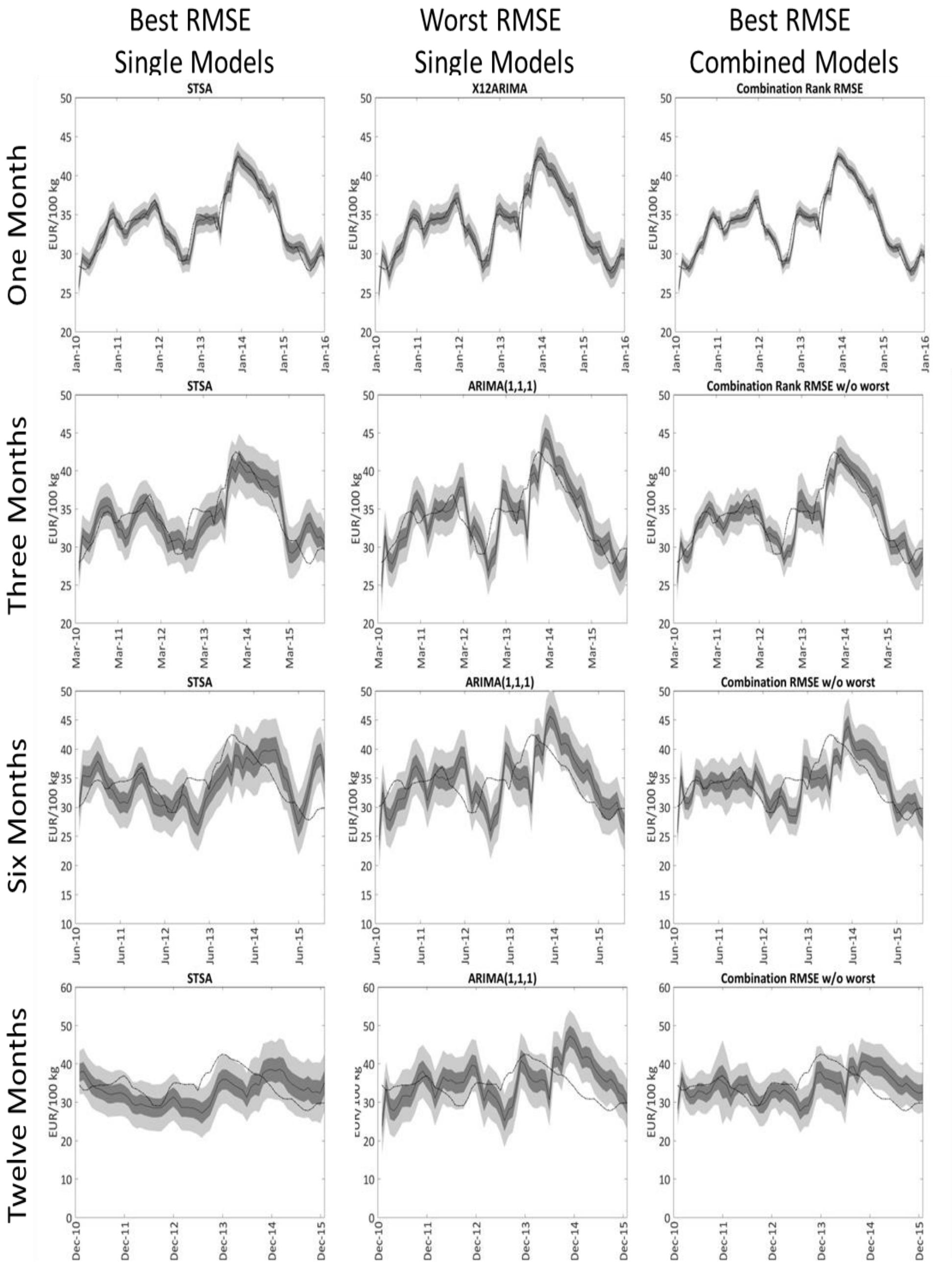


Figure 5: Selected Density Forecasts of German Farm Gate Prices (Grey Shaded Area) Compared to Actual Prices (Dashed Line).

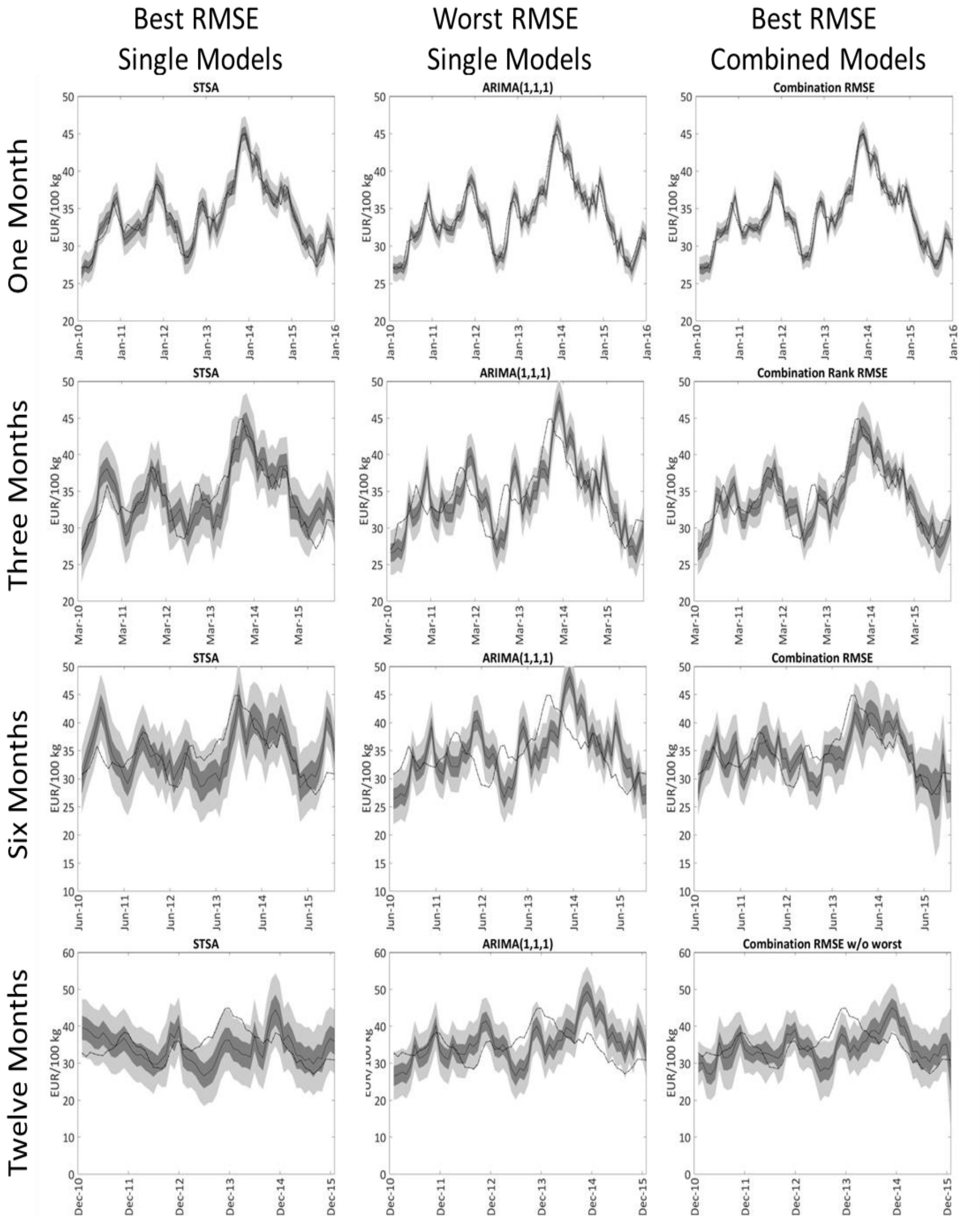


Figure 6: Selected Density Forecasts of Irish Average Farm Gate Prices (Grey Shaded Area) Compared to Actual Prices (Dashed Line).

7. Conclusion

The different levels of milk solids and their seasonal production as well as the different product portfolios and end markets result in different price dynamics for farm gate milk prices in Germany and Ireland. In addition recent policy changes in the EU have made forecasting of dairy prices more difficult as the volatility of EU dairy prices has sharply increased. However in this volatile environment accurate forecasts of dairy prices are of great importance as they aid cash flow management and improve decision making for production planning and investment. The analysis in this paper addresses these issues as it provides EU average, German and Irish farm gate milk price forecasts obtained from various time series models and evaluates the reliability of these forecasts. Both point estimates and density forecasts have been constructed and evaluated. Forecast accuracy depends on the forecast horizon thus overall no single best model could be identified. The main findings are summarized as follows:

- ARIMA and SARIMA models perform well on short forecast horizons (1 to 3 month).
- The structural time series approach by Harvey (1989) performs well on longer forecast horizons (12 month).
- Combining individual forecasts of different models significantly improves the forecast performance for all forecast horizons.
- These models perform well in terms of forecast direction with accuracy in excess of 75% in many cases.
- As the forecast horizon increases the forecast interval increases substantially

Decisions by farmers, like lowering or increasing production, should be based on price expectations in the future rather than current prices. This is especially true given the abolishment of the milk quota in 2015. In such an environment farmers now need to react more closely to market signals than previously. Thus forecasts of prices in the future can provide a comparative advantage as they can form the basis for improved decision making. This is especially true given the currently large volatility of dairy prices in the EU. The models identified in this paper perform well in forecasting farm gate milk prices especially with regards to the directional accuracy. As shown in Leitch and Tanner (1991) the directional accuracy can be directly linked to profits. With these forecasts farmers have a solid foundation to base their decision on.

While other studies which analyse milk price forecast performance are based on point forecasts (Hansen and Li, 2016; Lira, 2013; Glauco et al. 2015) this study also considers density forecasts. Density forecasts give important information about the uncertainty of prices in the future and thus form an important risk management tool. For cash flow management the point forecasts could be used to derive expected scenarios. The density price forecasts on the other hand could be a used for cash flow simulations and thus complementing the point forecasts. Overall the budgeting process should be improved. For example a farmer/company could identify a worst case price scenario from the forecast density and then design counter measures so that she/it remains solvent in such a situation[19]. Possible counter measures could then include the use of dairy futures or over the counter contracts to hedge potential exposure. Furthermore this may allow farmers to expand in a countercyclical nature as they may be in a better position than their peers when land prices drop in tandem with milk prices as outlined in Nicholson and Stephenson (2015).

Density forecasts can also assist decisions made by policymakers. Recent policies for example introduced some measures to deal with a market crisis situation. For example in March 2016 the EU Commission enabled producer organisations, inter-branch organisations and cooperatives in the dairy sector to establish voluntary agreements on their production and supply[20]. In addition the EU Commission is able to raise intervention ceilings for butter and

SMP or support promotion of agricultural products. With density forecasts of milk prices at hand policymakers can quantify the probability associated with these adverse market situations and decide whether to intervene or not. Likewise these interval forecasts could be used to quantify the potential costs of market interventions.

The forecasts presented in this paper could also be the basis for forward contracting as forecasts can help by making the price setting process more transparent. However losses might still be substantial for one party if the fixed price is set to a forecast price.

References

Aiolfi, M. and A. Timmermann. 2006. "Persistence in Forecasting Performance and Conditional Combination Strategies" *Journal of Econometrics* 135:31-53.

Allen, P. G. 1994. "Economic forecasting in agriculture" *International Journal of Forecasting* 10(1):81-135.

Arora, S. M., M. A. Little and P. E. McSharry. 2013. "Nonlinear and Nonparametric Modelling Approaches for Forecasting the US GNP" *Studies in Nonlinear Dynamics and Control* 17(4):395-420.

Bai, J. and S. Ng. 2005. "Tests of Skewness, Kurtosis, and Normality for Time Series Data" *Journal of Business and Economic Statistics* 23(1):49-60.

Bergmann, D., D. O'Connor and A. Thümmel. 2015. "Seasonal and cyclical behaviour of farm gate milk prices" *British Food Journal* 112(12):2899 - 2913.

Bressler, D. A. and J. L. King. 1989. "The Forecast and Policy Analysis." *American Journal of Agricultural Economics* 71(1989):503-506.

Bottum, J. C. 1966. "Changing Functions of Outlook in the U.S." *Journal of Farm Economics* 48:1154–1159.

Christoffersen, P. 1998. "Evaluating Interval Forecasts" *International Economic Review* 39:841-862.

Clark, T. E. and M. W. McCracken. 2011. "Testing for Unconditional Predictive Ability" in Clements, M. P. and D. F. Hendry (ed.), *the Oxford Handbook of Economic Forecasting.*

Colino, E. V., S. H. Irwin, P. Garcia and X. Etienne. 2012. "Composite and Outlook Forecast Accuracy" *Journal of Agricultural and Resource Economics* 37(2):228–246

Diebold, F. X. 2006. *Elements of Forecasting*, South Western Cengage Learning 4th edition.

Diebold, F. X. and R. S. Mariano. 1995. "Comparing Predictive Accuracy" *Journal of Business & Economic Statistics* 13(3):253-263.

Diebold, F.X., T.A. Gunther, and A.S. Tay. 1998. "Evaluating Density Forecasts with Applications to Financial Risk Management." *International Economic Review* 39:863–883.

Durbin, J. and S. J. Koopman. 2012. *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.

FAO and OECD. 2015. "Agricultural Outlook 2015-2024".

Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto and B.-C. Chen. 1998. "New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program" *Journal of Business and Economic Statistics* 16:127-152.

Giacomini, R. and H. White. 2006. "Tests of Conditional Predictive Ability," *Econometrica* 74:1545-1578.

Glauco R. G. R. Carvalho, D. Bessler, T. Hemme and E. Schröer-Merker. 2015. "Understanding International Milk Price Relationships". In *Selected Paper prepared for presentation at the Southern Agricultural Economics Association's 2015 Annual Meeting*, January 31 - February 3, 2015. Atlanta, Georgia.

Gloy, B. A., E. L. LaDue and K. Youngblood. 2002. "Financial Management Practices of New York Dairy Farms" Cornell Program on Agricultural and Small Business Finance R. B. 2002-09

Green, K. C and J. S. Armstrong. 2015. "Simple versus complex forecasting: The evidence" *Journal of Business Research* 68(8):1678–1685.

Hamulczuk, M., S. Grudkowska, S. Gędek and C. Klimkowski. 2013. "Essential Econometric Methods of Forecasting Agricultural Commodity Prices"

Hansen, B. G. and Y. Li. 2016. "An Analysis of Past World Market Prices of Feed and Milk and Predictions for the Future" *Agribusiness*doi:10.1002/agr.21474.

Harvey, A. C. 1989. *Forecasting, structural time series models and the Kalman flter* Cambridge University Press, Cambridge.

Hersbach, H. 2000. "Decomposition of the continuous ranked probability score for ensemble prediction systems" *Weather and Forecasting* 15:559–570.

Hurtado-Uria, C., D. Hennessy, L. Shalloo, D. O'Connor and L. Delaby. 2014. "Relationships between meteorological data and grass growth over time in the south of Ireland" *Irish Geography* 46(3):175-201.

International Dairy Federation. 2014. "The World Dairy Situation" *Bulletin of the International Dairy Federation* 476/2014 Brussels

Isengildina, O., S. H. Irwin, and D. L. Good. 2004. "Evaluation of USDA Interval Forecasts of Corn and Soybean Prices" *American Journal of Agricultural Economics* 86:990–1004.

Knüppel, M. 2015. "Evaluating the calibration of multi-step-ahead density forecasts using raw moments" *Journal of Business & Economic Statistics* 33(2).

Leitch, G. and J. E. Tanner. 1991. "Economic Forecast Evaluation: Profits versus the Conventional Error Measures" *American Economic Review* 81(3):580-90.

Lira, J. 2013. "A Comparison of the usefullness of Winters' and SARIMA models in forecasting of procurement prices of milk in Poland" *Quantitative Methods in Economics* 14(1):325-333.

MacDonald, J. M., J. Cessna, and R. Mosheim. 2016. "Changing Structure, Financial Risks, and Government Policy for the U.S. Dairy Industry" U.S. Department of Agriculture, Economic Research Service, March 2016.

Mosheim, R. 2012. "A Quarterly Econometric Model for Short-Term Forecasting of the U.S. Dairy Industry" U.S. Department of Agriculture. Economic Research Service. Technical Bulletin No. 1932. Jan.

Nicholson, C.F. and M. W. Stephenson. 2015. "Milk Price Cycles in the U.S. Dairy Supply Chain and Their Management Implications" *Agribusiness* 31:507-520.

Panagiotelis, A. and M. Smith. 2008. "Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions" *International Journal of Forecasting* 24(4):710-727.

Pesaran, M. H. and A. Timmermann. 1992 "A simple nonparametric test of predictive performance" *Journal of Business and Economic Statistics* 10:461-465.

Pesaran, M. H. and A. Timmermann. 2009. "Testing Dependence among Serially Correlated Multicategory Variables" *Journal of the American Statistical Association* 104(485):325-337.

Promar International. 2003. Strategic Development Plan for the Irish Dairy Processing Sector.

Rosenblatt, M. 1952. "Remarks on a Multivariate Transformation" *Annals of Mathematical Statistics* 23(3):470-472.

Sanders, D. R. and M. R. Manfredo. 2003. "USDA livestock price forecasts: A comprehensive evaluation" *Journal of Agricultural and Resource Economics* 28(2):316-334.

Sanders, D. R. and M. R. Manfredo. 2005. “Forecast Encompassing as the Necessary Condition for Futures Market Efficiency: Fluid Milk Futures” *American Journal of Agricultural Economics* 87(2005):610-620.
Stock, J. H. and M. W. Watson. 2004. “Combination Forecasts of Output Growth in a Seven-Country Data Set” *Journal of Forecasting* 23:405-430.
Timm, T. R. 1966. “Proposals for Improvement of the Agricultural Outlook Program of the United States” *Journal of Farm Economics* 48:1179.
Timmermann, A. 2006. “Forecast Combinations”, in G. Elliott et al. (ed.) *Handbook of Economic Forecasting.*
Trujillo-Barrera, A., P. Garcia, and M. Mallory. 2012b. “Density Forecasts of Lean Hog Futures Prices” *Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. St. Louis, MO. [http://www.farmdoc.illinois.edu/nccc134].

[1] It should be noted that in the short run farmers can vary production marginally by managing feeding but a significant change is only possible by making long run changes to herd size and genetics.
[2] Such a scenario could for example be the 95 % percentile of the forecast density.
[3] The latest OECD/FAO projection of dairy prices can be found in FAO and OECD (2015). These are annual point estimates up to 2024.
[4] Rabobank publishes a quarterly dairy outlook report with dairy forecasts.
[5] While milk contains solids other than fat and protein such as lactose and minerals these latter solids are not published and the former solids are the ones most commonly used when pricing milk at farm level
[6] ARIMA models with GARCH (generalized autoregressive conditional heteroscedasticity) residuals have also been tested but found to not improve forecasting accuracy.
[7] In this study the same price series as Bergmann et al. (2015) are considered albeit over a different time horizon.
[8] Choosing a hold out period which is too short might result in the best model changing every period while choosing too long a hold out period might result in recent forecast performance not transmitting fast enough to the weights.
[9] E.g. the random walk model is nested in the structural time series approach.
[10] In this context the unconditional coverage test is intended.
[11]http://ec.europa.eu/agriculture/milk-market-observatory/index_en.htm (accesed1st September 2016)
[12] Commission Regulation (EU) No 479/2010
[13] The DM test has the null hypothesis that the performance of two forecast series is equal. Thus a rejection implies that one model is significantly outperforming the other.
[14] An exception is the equal weighted combination scheme which does not outperform the STSA model for a one year horizon.
[15] It should be noted that hedging opportunities are limited at present for EU dairy farmers. The EEX exchange which offers Butter, SMP and Whey futures is still at an embryonic stage and cross hedging milk using these contracts is novel and limited. In addition the basis (the futures price minus spot price) must be considered when investing in futures. For example if the basis is large loses can be realized even when spot prices rise.
[16] As 40 models were developed for each series (including combinations models) it was decided to limit this figure to a sample of the best and worst single models based on the RMSE along with an example of a good combinations model for each time horizon. All fan charts are available on request from the authors. It should be noted that the RMSE difference between the best and worst models for especially the short horizons might be small, which means the model with the worst RMSE may not necessarily be a poor model.
[17] For example the commodities may be too old or do not meet quality standards.
[18] Limits in January 2016 were 50,000 tonnes for butter and 109,000 tonnes for SMP and have doubled in April 2016. In May 2016 the SMP limits were further raised to 350,000 tonnes.
[19] This is sometimes also called cash flow at risk (https://www.risknet.de/wissen/rm-methoden/cash-flow-at-risk/accessed 1st September 2016).
[20] http://europa.eu/rapid/press-release_IP- en.htm(accessed 1st September 2016)