Analysis of global food prices with new Bayesian model combination schemes

Krzysztof Drachal[1], Justyna Góral[2], Włodzimierz Rembisz[3]

Abstract

The purpose of this paper is to indicate key determinants of food prices and methods of their forecasting. The analysed data consist of 5 food price indices provided by FAO: food, meat, dairy indices (data from 1990 to 2016). Model combination scheme is one of the possible econometric tool to deal with model uncertainty. For various commodities prices some of these methods were found to produce more accurate price forecasts than the previously studied (applied) methods. The novel methods applied herein are characterized by the use of time-varying parameters approach and dealing with model uncertainty.

Keywords: food prices, model averaging, Bayesian model combination schemes

JEL Classification: E37; E31; C53

1. Introduction

Farmers' production decisions are based partly on their expectations about future prices. Producers usually respond with a shift in time by one period (t+1). Its objective is to determine a proper price path for this good so that the supply and demand are fully implemented in each year (i.e. market clearing price). One way to describe the systematic behaviour of prices through time is by using linear statistical models. For the rational expectations model the information set determining the expected price is hypothesized to be the current information on all the variables thought to determine prices in the forthcoming period (Tomek and Kaiser, 2014). The expected price is equivalent to a forecast from a structural econometric model. Basing on FAO (2019) food price indices, it can be observed that there exists a rising trend in global food prices since 2000 (Kaaresvirta and Mehrotra, 2009).

One of the solutions to this problem is to initially consider several variables, and construct multiple models (for example, linear regressions ones). Then, each of the model can be evaluated on the basis of some criterion (Akaike Information Criterion, Mean Squared Error, etc.). This leads to ascribing to each of the model certain numerical value, which can be used to weighted

[1] Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland, kdrachal@wne.uw.edu.pl.
[2] Collegium of Socio-Economics, Warsaw School of Economics, Warsaw, Poland, justyna.goral@sgh.waw.pl.
[3] Mathematics Application in Agricultural Economics Department, Institute of Agricultural and Food Economics – National Research Institute, Warsaw, Poland, wlodzimierz.rembisz@ierigz.waw.pl.

averaging across all the models, or to selecting only one model (Burnham and Anderson, 2002). Some of the recently proposed methods have interesting characteristics. First of all, they are Bayesian ones. Therefore, they are very in-line with the real market approach and recursive updating. Consequently, it is interesting to analyse some new methods in this context. Moreover, the interest in the Bayesian methods for food prices forecasting is increasing amongst the researchers recently (Mazur, 2018).

2. Potential food price drivers

Rezitis and Sassi (2013) noticed the important role of supply and demand as factors driving food prices. The results of Kalkuhl's team (2014) pointed to the increasing linkages among food, energy, and financial markets (the financialisation of food and farming), which explained much of the observed food price spikes and volatility. The analysis had indicated that exogenous shocks as well as the linkages between food, energy, and financial markets played a significant role in explaining food price volatility and spikes. The results indicate that food prices, energy prices, and the wage rate were integrated of order one, whether a deterministic trend was included or not in the test equation.

Food price volatility may increase because of stronger linkages between agricultural and energy markets. In addition to demand and supply shocks, speculation is an important factor in explaining and triggering extreme price spikes. C. Erdem *et al.* (2013) investigated volatility spillover between oil and selected agricultural commodity markets (wheat, corn, soybean, and sugar) that are key agricultural products for biofuels and for food in the world. The variance causality test showed that while there was no risk transmission between oil and agricultural commodity markets in the pre-crisis period, oil market volatility spilled on the agricultural markets in the post-crisis period. C. Erdem *et al.* showed that the dynamics of volatility transmission changes significantly following the food price crisis. After the crisis, risk transmission emerged as another dimension of the dynamic interrelationships between energy and agricultural markets. Q.E. Bouri *et al.* (2018) employed a relatively newer modelling technique – a time-varying copula with a switching dependence - to characterise the conditional dependence between energy and agricultural commodity markets in a more realistic way. However, Nazlioglu and Soytas (2012) found that crude oil price and exchange rates significantly impact food prices. As a result, the information from financial market can be the early signal for some future changes on agricultural markets. There are also studies focusing on the role of the global economic conditions in driving food prices (Alam and Gilbert, 2017).

3. Data

This analysis covers the period between 1990 and 2016. It is based on monthly data. Prices of food are taken from FAO (2019). In Table 1 the explanatory variables are listed. Denton-Cholette method is used, as it nicely deals with trends, stationarity and cointegration issues (Dagum, Cholette, 2006). FAO indices are transformed into returns for the purpose of modelling. This allows to set the initial variance for the Bayesian models to 0.01 (Raftery *et al.*, 2010).

Table 1. Explanatory variables

Symbol	Description	Source
BIO	traditional biofuels production	Ritchie and Roser (2019)
GDP	average of annual growths of GDP per capita from BRIC countries	The World Bank (2019)
MS	Moody's seasoned Baa corporate bond minus federal funds rate; market stress indicator	FRED (2019)
KEI	Kilian index of global economic activity	Kilian (2019)
TB3MS	3-month treasury bill: secondary market rate for U.S.	FRED (2019)
SP500	S&P 500 index	Stooq (2019)
R	long-term government bond yields: 10-year: main (including benchmark) for U.S.	FRED (2019)
AUD	Australian dollar to U.S. dollar exchange rate	Stooq (2019)
CAD	Canadian dollar to U.S. dollar exchange rate	FRED (2019)
OIL	average of Brent, Dubai and WTI crude oil spot prices	The World Bank (2019)
E	energy prices	The World Bank (2019)
F	fertilizer prices	The World Bank (2019)

4. Methodology

There are a few types of short-term forecasting methods to predict the food prices. These methods are usually used to forecast some specific kinds of agricultural commodities. The latest researches show that the mixed model (price warning model based on neural networks) has obtained satisfactory forecasting results. The mixed model makes an improvement both on the forecasting accuracy and efficiency compared with any other single models.

The computations are done in R (R Core Team, 2018) with the help of "fDMA" and "forecast" R packages (Hyndman and Khandakar, 2008). The details of Dynamic Model Averaging (DMA) are described in the paper by Raftery *et al.* (2010). Roughly speaking with $n = 12$ potential explanatory variables, there can be constructed $K = 2^{12} = 4096$ multiple linear regression models (with the constant term included). For each of these regression models, the regression coefficients are recursively updated by the Kalman filter. Whereas the updating of variance matrices is done with the forgetting procedure, which requires setting a forgetting factor λ (Raftery *et al.*, 2010). Smaller values of λ correspond to more abrupt changes of the regression coefficients, whether λ = 1 assumes they are fixed. For the purpose of model averaging, a set of two time-varying weights is recursively computed, as the following equations present:

$$\pi_{t|t-1,k} = \left(\pi^{\alpha}_{t-1|t-1,k} + c\right) \Big/ \left(\sum_{i=1}^{K} \pi^{\alpha}_{t-1|t-1,i} + c\right) \quad (1)$$

$$\pi_{t|t,k} = \left[\pi_{t|t-1,k} f_k \left(y_t|Y^{t-1}\right)\right] \Big/ \left[\sum_{i=1}^{K} \pi_{t|t-1,i} f_i \left(y_t|Y^{t-1}\right)\right] \quad (2)$$

where:

$f_k(y_t|Y^{t-1})$ is the predictive density of the k-th model at y_t, under the assumption that the data up to time t is known, and k = 1, …, K.

The second forgetting factor α must also be specified. Herein, the following combinations are tested α = {1, 0.99, 0.98, 0.95, 0.90} = λ. To initialize the computations $\pi_{0|0,k}$ are set to 1 / K, and c = 0.001 / K is taken, following Raftery *et al.* (2010).

The weights $\pi_{t|t-1,k}$ are called posterior predictive probabilities. Their sums for the models, which contain a given explanatory variable are called relative variable importance. They can be used (together with averaged regression coefficients) to analyse the predictive power of a given variable (Burnham and Anderson, 2002).

The DMA forecast is computed as the weighted average from the forecasts given by the component models:

$$y_t^{DMA} = \sum_{k=1}^{K} \hat{\pi}_{t|t-1,k} \hat{y}_t^{(k)} , \quad (3)$$

where:

$\hat{y}_t^{(k)}$ is the prediction from the k-th multilinear regression model.

If α = 1 = λ, then, as noticed by Raftery *et al.* (2010), DMA is a computationally efficient way towards Bayesian Model Averaging (BMA). If Eq. (3) is modified in such a way that only the forecast from the model with the highest $\pi_{t|t-1,k}$ is taken, then such a scheme is called Dynamic Model Selection (DMS). Similarly, if α = 1 = λ is set, the scheme is called Bayesian Model Selection (BMS).

However, Barbieri and Berger (2004) observed that the selection of the model with the highest posterior probability is not always an optimal choice. They proposed to select the model, which contains exactly those explanatory variables, for which the relative variable importance is greater than or equal to 0.5. Such a scheme is called Median Probability Model (MED). Similarly, for $\alpha = 1 = \lambda$ it is called Bayesian Median Probability Model (BMED).

As the benchmark models the automatic version of ARIMA by Hyndman and Khandakar (2008), and the naïve method (NAÏVE) are taken. Besides, a time-varying parameters regression (TVP) with all explanatory variables, as given in Table 1, is considered. This is simply the DMA scheme reduced to only one model, $K = 1$. For TVP, $\lambda = 0.99$ is taken, as such a choice is the standard assumption (Raftery *et al.,* 2010). Also, the historical average (HA) is computed. By HA it is understood that the forecast for time t is the mean value of all observations up to time *t-1*.

The in-sample consists of first 80 observations. The further observations constitute the out-of-sample set, which is used to make conclusions and computation of forecast accuracy measures. In particular, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Scaled Error (MASE) are computed. These measures are computed for the raw values of analysed indices. (The models are forecasting the returns, but simple algebraic transformations can convert these forecasts into the forecasts of price indices.) MASE is an interesting forecast accuracy measure proposed by Hyndman and Koehler (2008). It is scale invariant, penalizes positive and negative errors equally, and penalizes large and small differences between real and forecasted values equally. Absolute Scaled Error loss function is also compatible with the Diebold and Mariano (1995) test procedures (Frases, 2016). If not stated otherwise, 5% significance level was taken.

5. Results

Table 2 presents forecast accuracy measures. It can be seen that the model minimising MASE for food price index is MED with $\alpha = 0.99 = \lambda$. For meat price index – DMA with $\alpha = 0.99 = \lambda$. For the other indices, it is ARIMA model. As a result, the considered model combination schemes do not produce more accurate forecast for most of the analysed price indices. However, in one case it is the model selection scheme proposed by Barbieri and Berger (2004), which is not based on selecting the model with the highest posterior probability, which minimises MASE. In one case, it is the DMA scheme with the commonly advised forgetting parameters combination (Raftery *et al.,* 2010). Table 3 presents the results from the Diebold-Mariano test. It can be seen that in the case of food price index, the selected MED model produces statistically significantly more accurate forecasts than many other forecasts from other Bayesian model combination schemes. These forecasts are also more accurate than the ones from the naïve method. However, it cannot be concluded that they are more accurate than the ones from ARIMA method. In the case of the meat price index, the conclusion about the selected DMA model is the same.

Table 2. Forecast accuracy measures

Items	Food Price Index			Meat Price Index			Dairy Price Index		
	RMSE	MAE	MASE	RMSE	MAE	MASE	RMSE	MAE	MASE
DMA ($\alpha = 0.99$)	3.973	2.597	0.986	3.516	2.704	0.943	7.775	4.927	1.115
DMA ($\alpha = 0.98$)	3.900	2.596	0.985	3.519	2.710	0.945	7.747	4.889	1.107
DMA ($\alpha = 0.95$)	3.977	2.777	1.054	3.657	2.793	0.974	7.640	4.853	1.099
DMA ($\alpha = 0.90$)	4.249	2.932	1.113	3.907	2.959	1.032	8.185	5.130	1.161
BMA	4.016	2.588	0.983	3.556	2.727	0.951	7.794	4.907	1.111
DMS ($\alpha = 0.99$)	4.364	2.634	1.000	3.571	2.756	0.962	7.864	5.018	1.136
DMS ($\alpha = 0.98$)	3.955	2.577	0.978	3.641	2.816	0.982	8.012	5.011	1.134
DMS ($\alpha = 0.95$)	4.144	2.852	1.083	3.941	3.005	1.048	8.154	5.212	1.180
DMS ($\alpha = 0.90$)	4.782	3.152	1.197	4.249	3.214	1.121	8.635	5.419	1.227
BMS	4.246	2.640	1.002	3.588	2.742	0.956	8.159	5.201	1.177
MED ($\alpha = 0.99$)	3.894	2.556	0.970	3.615	2.791	0.973	7.897	5.103	1.155
MED ($\alpha = 0.98$)	4.001	2.652	1.007	3.610	2.773	0.967	7.985	5.133	1.162
MED ($\alpha = 0.95$)	4.076	2.796	1.062	3.791	2.928	1.021	7.918	5.111	1.157
MED ($\alpha = 0.90$)	4.629	3.094	1.175	4.171	3.134	1.093	9.057	5.547	1.256
BMED	4.202	2.630	0.998	3.616	2.770	0.966	8.088	5.141	1.164
TVP ($\lambda = 0.99$)	4.082	2.676	1.016	3.601	2.780	0.969	8.056	5.219	1.182
HA	50.166	37.948	14.406	34.709	28.460	9.922	65.026	46.688	10.569
ARIMA	3.699	2.623	0.996	3.736	2.840	0.990	7.053	4.225	0.956
NAÏVE	4.295	2.902	1.102	3.856	2.893	1.009	8.028	4.988	1.129

Source: own calculations.

On the other hand, in the case of the diary price index, all models can be assumed to produce statistically significantly less accurate forecasts than the ARIMA method. It can be seen that the role of the growth of GDP per capita in BRIC countries has been increasing since 2010. The role of market stress was important around 2008. The role of global economic activity was the highest around 2000, and since then it has been decreasing. The role of short-term interest rate has always been important, however in 1990s its role was mostly positive, whereas in 2010s it is negative. The role of long-term interest rate has been most of the time positive. Similarly, stock prices index has been an important explanatory variable all the time. However, around 2000 its role was negative, but recently it is positive. In the case of exchange rates the role of AUD currency is the highest. Its importance has also increased recently. The role of oil price has an increasing trend. Its role increased around 2000, later it increased even more in 2010 and keeps to be an important factor. Energy prices in general played an important role only in the end of 1990s and in 2000s. Fertilizer prices were most important around 2010[4].

Table 3. The Diebold-Mariano test results

Items	Food Price Index		Meat Price Index		Dairy Price Index	
	DM test statistic	DM test p-value	DM test statistic	DM test p-value	DM test statistic	DM test p-value
DMA ($\alpha = 0.99 = \lambda$)	0.679	0.249			2.806	0.003
DMA ($\alpha = 0.98 = \lambda$)	0.778	0.218	0.225	0.411	2.550	0.005
DMA ($\alpha = 0.95 = \lambda$)	2.176	0.015	1.206	0.114	2.388	0.008
DMA ($\alpha = 0.90 = \lambda$)	2.449	0.007	2.238	0.013	3.634	0.000
BMA	0.464	0.321	0.690	0.245	2.709	0.003
DMS ($\alpha = 0.99 = \lambda$)	0.703	0.241	1.629	0.052	2.946	0.002
DMS ($\alpha = 0.98 = \lambda$)	0.328	0.372	2.276	0.011	2.887	0.002
DMS ($\alpha = 0.95 = \lambda$)	2.476	0.007	2.961	0.002	3.475	0.000
DMS ($\alpha = 0.90 = \lambda$)	2.909	0.002	3.513	0.000	4.346	0.000
BMS	0.813	0.208	0.784	0.217	3.582	0.000
MED ($\alpha = 0.99 = \lambda$)			2.632	0.004	3.260	0.001
MED ($\alpha = 0.98 = \lambda$)	1.775	0.038	1.389	0.082	3.295	0.001
MED ($\alpha = 0.95 = \lambda$)	1.959	0.025	2.224	0.013	3.192	0.001
MED ($\alpha = 0.90 = \lambda$)	2.905	0.002	3.051	0.001	4.299	0.000
BMED	0.754	0.225	1.243	0.107	3.488	0.000
TVP ($\lambda = 0.99$)	1.396	0.081	1.691	0.045	3.766	0.000
HA	17.330	0.000	20.130	0.000	15.260	0.000
ARIMA	0.631	0.264	1.388	0.083		
NAÏVE	2.291	0.011	1.947	0.026	2.952	0.002

6. Conclusions

Agricultural commodity prices are substantially more volatile than are the prices of most nonfarm goods and services. Besides the traditional causes for price fluctuations, agricultural commodities are increasingly connected to energy and financial markets, with potentially destabilizing impacts on prices. Trend in price depends on the long-run supply and demand conditions. Food prices exhibited a large volatility in recent years. In particular, there have been

[4] The conclusions from the Giacomini-White test are the same as those from the Diebold-Mariano one.

huge price shocks in the years 2007/2008, 2010/2011, and 2012/2013. Some of the reasons for rising food prices are bad weather, increasing factor prices, in particular oil prices, or increasing usage of bio-fuels. Price differentiation may also result from different price elasticity of demand. The nature of demand for agricultural products is another factor determining the prices of agricultural products. The constantly increasing complexity and diversity of products as well as changing consumer preferences make demand analysis more and more difficult. Speculative demand also needs to be taken into account.

The simplest way to measure price volatility is the coefficient of variation (CV). This is the standard deviation of prices over a particular time interval divided by the mean price over the same interval. As an alternative is used the standard deviation of changes in the logarithm of prices. It is less affected by strong trends over time. Different prediction methods have been presented in recent years for price forecasting. Non-stationary time series models like Generalised Auto-Regressive Conditional Heteroskedastic, stationary time series models such as Auto-Regressive (AR), dynamic regression and transfer function, and Auto-Regressive Integrated Moving Average (ARIMA) have been proposed for this purpose. Other researchers proposed neural networks (NNs) and fuzzy neural networks (FNNs) for price forecasting. The selected Bayesian model combination schemes produce the most accurate forecasts only for food and meat price index, amongst all the considered models. In the case of other indices, the most accurate forecasts are produced by the ARIMA method. In the case of meat price index, according to Dynamic Model Averaging (DMA) scheme, the growth of GDP per capita in BRIC countries, market stress, global economic activity, interest rates, stock prices, exchange rates, oil price and fertilizer prices are found to be the most important price driving factors amongst the considered explanatory variables. Moreover, time-varying patterns in these importance are found.

Acknowledgements

(K. Drachal) Research funded by the Polish National Science Centre grant under the contract number DEC-2015/19/N/HS4/00205.

References

Alam, M.R., Gilbert, S. (2017). Monetary policy shocks and the dynamics of agricultural commodity prices: evidence from structural and factor-augmented VAR analyses. *Agricultural Economics*, 48, 15-27.

Barbieri, M.M., Berger, J.O. (2004). Optimal predictive model selection. *The Annals of Statistics*, 32, 870-897.

Bouri, Q.E., Roubaud, D., Jawad, S., Shahzad, H. (2018). Risk spillover between energy and agricultural commodity markets: a dependence-switching CoVaR-copula model. *Energy Economics*, 75, 14-27.

Burnham, K.P., Anderson, D.R. (2002). Model Selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York, 49-117.

Dagum, E.B., Cholette, P.A. (2006). Benchmarking, Temporal Distribution, and Reconciliation Methods for TimeSeries. Lecture Notes in Statistics. Springer-Verlag, New York, 80-82.

Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics, 13, 253-263.

Drachal, K. (2015). Review of GARCH model applicability in view of some recent research. *Journal of Academic Research in Economics*, 7/2, 191-200.

Erdem, C., Nazlioglua, S, Soytas, U. (2013). Volatility spillover between oil and agricultural commodity markets. *Energy Economics,* 36, 658-665.

FAO (2019). World Food Situation (http://www.fao.org/worldfoodsituation/foodprices-index/en). Food and Agriculture Organization of the United States, Rome.

Franses, P.H. (2016). A note on the mean absolute scaled error. *International Journal of Fore-casting*, 32, 20-22.

FRED (2019). Economic Research (https://fred.stlouisfed.org). Federal Reserve Bank of St. Louis, St. Louis, MO.

Giacomini, R., White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74, 1545-1578, https://stooq.pl (15.11.2019).

Hyndman R.J., Khandakar Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26, 1-22.

Irz X., Niemi J., Xing L. (2013). Determinants of food price inflation in Finland – the role of energy. *Energy Policy*, Elsevier 63C, 656-663.

Kaaresvirta, J., Mehrotra, A. (2009), Business surveys and inflation forecasting in China, Economic Change and Restructuring, 42, 263-271.

Kalkuhl M., Tadesse G., Algieri, B. & von Braun, J. (2014). Drivers and triggers of inter-national food price spikesand volatility. *Food Policy*, 47, 117-128.

Kilian L., Zhou X. (2019). Oil Prices, Exchange Rates and Interest Rates, CEPR Discussion Paper No. DP13478, January 2019.

Mazur, B. (2018). Cyclical fluctuations of global food prices: a predictive analysis. Conference Proceedings, Foundation of the Cracow University of Economics, Cracow, 286-295.

Nazlioglu, S., Soytas, U. (2012). Oil price, agricultural commodity prices, and the dollar: a panel cointegration and causality analysis, *Energy Economics*, 34, 1098-1104.

Popp J., Szenderak J., Harangi-Rakos, M. (2019). Price and volatility spillovers of the producer price of milk bet-ween some EU member states. *German Journal*, 68/2, 61-76.

Raftery, A.E., Karny, M., Ettler, P. (2010). Online prediction under model uncertainty via Dynamic Model Aver-aging: application to a cold rolling mill. *Technometrics*, 52, 52-66.

Rezitis, A.N., Sassi, M. (2013). Commodity food prices: review and empirics. *Economics Research International* 2013, ID 694507.

Ritchie, H., Roser, M. (2019). Renewable Energy. In: Our World in Data, Oxford.

Tomek, W.G., Kaiser, H.M. (2014). Agricultural product prices (5th edition), Cornell University Press, USA.