On the dynamics of the correlation between macroeconomic variables based on petroleum raw materials

Jacek Stelmach[1]

Abstract

The ability to bring accurate forecasts for commodity prices – which are essential for the enterprise, may be an important competitive advantage. Unfortunately, created regression models may be outdated, which is the result of changes in the macroeconomic environment. This is, among others, the effect of changes of impact many unknown factors on dependent variables. It is important, therefore, to identify these changes and estimate their direction. The article presents the historical dynamics of the correlation between selected variables representing oil raw materials and semiproducts in the years 1995-2015, indicating the possibility of the forecast of changes for such correlations.

Keywords: *regression models, commodity prices' forecast, correlation*

JEL Classification: C530

1. Introduction

The ability to build regression models to create the forecasts of variables important for business activity can lead to significant competitive advantage. It may be for example the price of strategic raw materials, demand for ready products, variable used in price formulas, et cetera. The accuracy of such predictions obviously depends on the quality of the model, with the impact of the proper selection of independent variables and time range of observations taking into consideration. Higher number of observations usually add more information. Today however, increased instability in the world economy is observed – a result of political and social turmoils, speculations or military actions (Cariolle and Goujon, 2015; Stock and Watson, 2010). It increases the risk of use of such observations that describe no longer valid dependencies and best known tests do not allow to reject the hypothesis about stationarity. Such risk can be, of course limited – with Weighted OLS method in which certain observations have greater impact. Another solution can be aggregated regression models in which base models are created from the set of observations coming from different time ranges (Gatnar, 2008; Powell and Stoker, 2005). Still, a crucial issue remains the choice of the range of observations that can vary depending on a particular variable. Presented investigations, based on data that represent the prices of oil raw material, its semiproducts and

[1] Polwax S.A., Łukasiewicza 2, 43-502 Czechowice-Dziedzice, Poland, e-mail: jacek.stelmach@polwax.pl.

other energy commodities, show another approach to the problem: taking into account the dynamics of correlation between variables. This study is particularly important for the author since they reach directly to the present.

2. Description of investigations

2.1. Data review

Professional interest was the reason for the author to focus on variables related to the refining industry. All the quotations were obtained from very popular website: www.indexmundi.com. First – the main raw material – crude oil was taken into consideration:

- *Brent*-light blend 38 API, fob U.K., (US Dollars per Barrel) – the most important in EU market.

Then the following raw materials and intermediates were taken into account:

1. *Gasoline* – New York Harbor Conventional Gasoline Regular Spot Price FOB, (US Dollars per Gallon).
2. *Heating_oil* – New York Harbor No. 2 Heating Oil Spot Price FOB, (US Dollars per Gallon).
3. *Natural_gas* – Natural Gas spot price at the Henry Hub terminal in Louisiana, (US Dollars per Million Metric British Thermal Unit).
4. *Propane* – Mont Belvieu, TX Propane Spot Price FOB, (US Dollars per Gallon).
5. *Coal_AU* – Australian thermal coal, 12,000- btu/pound, less than 1% sulfur, 14% ash, FOB Newcastle/Port Kembla, (US Dollars per Metric Ton).
6. *DAP* – DAP (diammonium phosphate) fertilizer, standard size, bulk, spot, f.o.b. US Gulf, (US Dollars per Metric Ton).
7. *Urea* – Urea fertilizer (Black Sea), bulk, spot, f.o.b. Eastern Europe, (US Dollars per Metric Ton).

Chosen wide time range: from November 1995 till October 2015 (20 years) allows comparing very interesting changes in dependencies between global variables in historically important periods:

1. dot-com bubble and bessa (1995-2001),
2. hossa connected with housing bubble (2001-2007),
3. global crisis and recession (2008-2012),
4. fluctuations and growth (2012-2015).

Variables values were indexed to the first observation for easier further analysis.

2.2. Overview of trend and dispersion

First, moving average and moving variance were calculated for two intervals cases: 12 months and 48 months. The latter case (48 months) excludes however first four years (1995-1999) from further analysis. Fig. 1 presents charts of these measures.

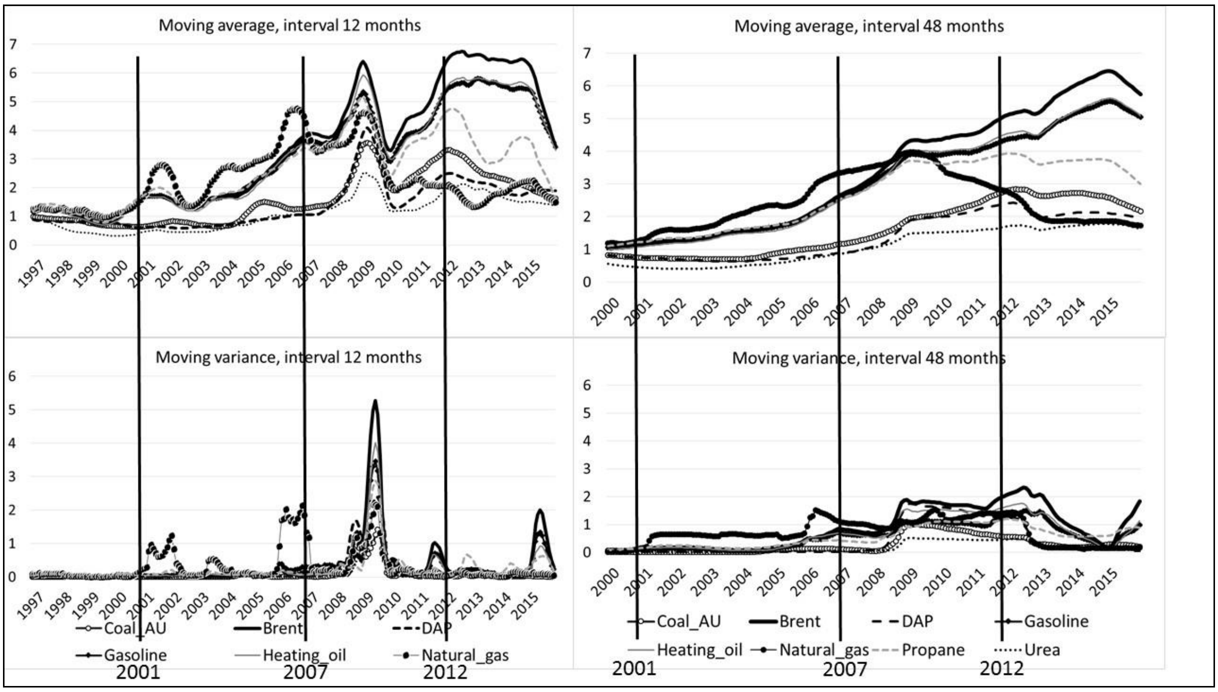

Fig. 1. Moving average and moving variance (interval 12 and 48 months) of analyzed variables.

Theses charts show the variety of trends of these variables, except years 2007-2010 (first part of the crisis). The same period is characterized by exceptionally high variability which is enhanced by the investors. Their perceptions of risk is less volatile than return expectations (Hoffmann et al., 2013). Only in the period before 2001, the trends were partially similar, with small dispersion. Regression models based on observations collected in such long period are not able to describe current dependencies, although their goodness-of-fit may be high. Furthermore, in a period of large variability (economic instability) any regression model could be of poor quality.

Such statement can lead to a sad conclusion that today higher economic variability makes it difficult to calculate sufficiently accurate forecasts (Caballero, 2010; Borio, 2014). Of course, it is possible to build models based only on a very small number of the latest observations but it may limit their quality and ability to generalize the dependencies between the dependent variable and predictors. Risk of improper regression models increases due to unidentified spurious regression – either proper tests are not carried out or the results of such tests reject the hypothesis about unit root.

2.3. Investigation of correlations in longer period

Fortunately, the rules that control economic changes do not change rapidly. There is a delay, an effect of inertia of industrial and economic processes, long-term agreements, conservative behaviour of society, willingness to keep ‘status quo’. It means that the changes of dependencies can be rather smooth and predictable in short period. This paper presents the investigations carried out to verify the above statement. There were calculated the moving correlations between such variables that are linked by supply chain, production processes or market competition. Due to the finite (but greater than zero) time of the production process and economic effects, also the correlation with lagged variables was analyzed. There is, of course, a risk of spurious correlation – indication of the relationship between two variables where one does not exist. However, analysing forecasting ability, the measure of accuracy of forecast seems to be considered first of all. Moreover, the same risk exists building regression models with no sufficient knowledge about non-stationarity.

This paper presents particularly a study with *DAP* variable. This variable describes the quotation of diammonium phosphate fertilizer and is linked directly with: *Natural_gas* – important raw material and a source of energy, *Urea* – another type of fertilizer that can be treated as complementary fertilizer and indirectly with *Brent, Gasoline, Propane, Heating_oil* and *Coal_AU* as either competitive energy commodities or competing products from processing of crude oil and gas. The investigations were carried out for different intervals: 6 months, 12 months, 24 months and 48 months with lag of variable +/- 3 months. Too short interval allows indication of faster coming changes but is more sensitive to temporary fluctuations, too long may be too conservative and lag the response to changes in trends. Finally, as a compromise, 12 months interval was chosen. For all the intervals Augmented Dickey–Fuller and Phillips–Perron tests were carried out. Calculated *p-value* allows to reject null hypothesis about unit root. Moreover, KPSS test fails to reject null hypothesis about stationarity. Fig. 2-8 show the correlation between above described variables, presenting moving correlation with lags between variables as a “ribbon”. The chart on the left-hand side presents the dependencies for whole period (1997-2015), the chart on the right-hand side – last 5 years, as the period most important for future forecasts.

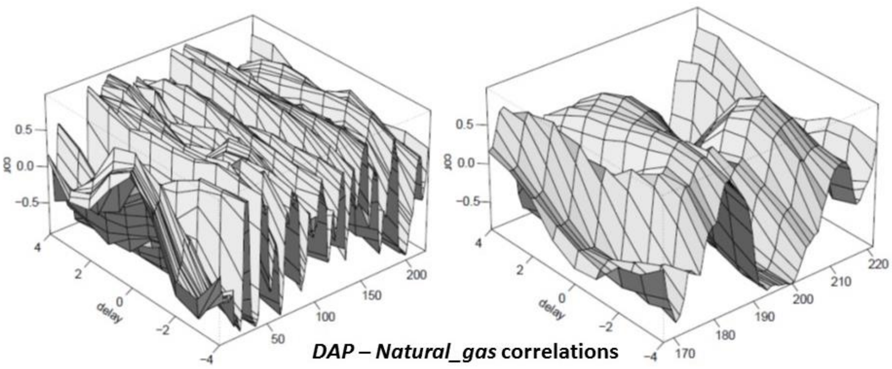

Fig. 2. Moving correlations *DAP – Natural_gas*.

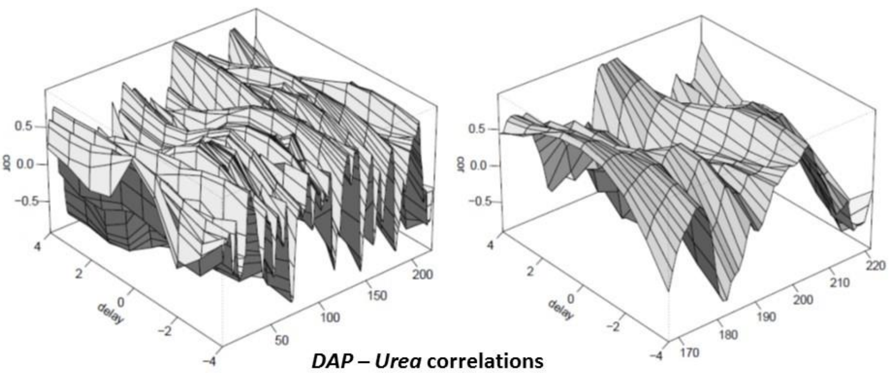

Fig. 3. Moving correlations *DAP – Urea*.

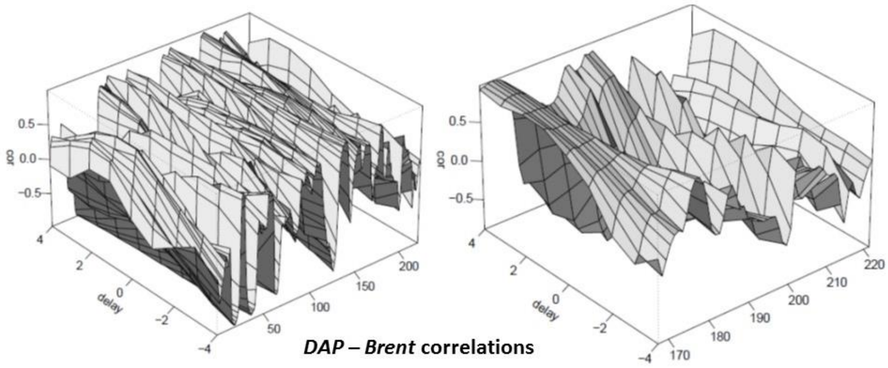

Fig. 4. Moving correlations *DAP – Brent*.

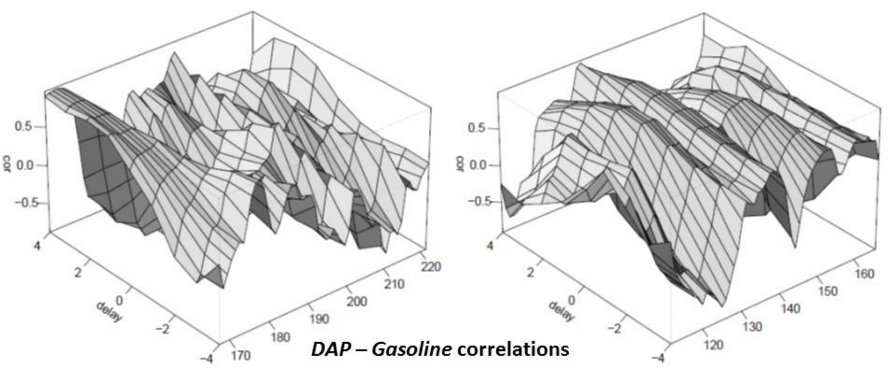

Fig. 5. Moving correlations *DAP – Gasoline.*

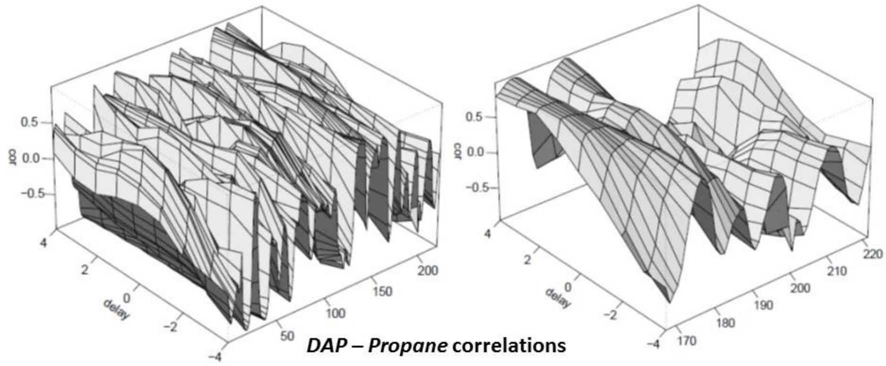

Fig. 6. Moving correlations *DAP – Propane*.

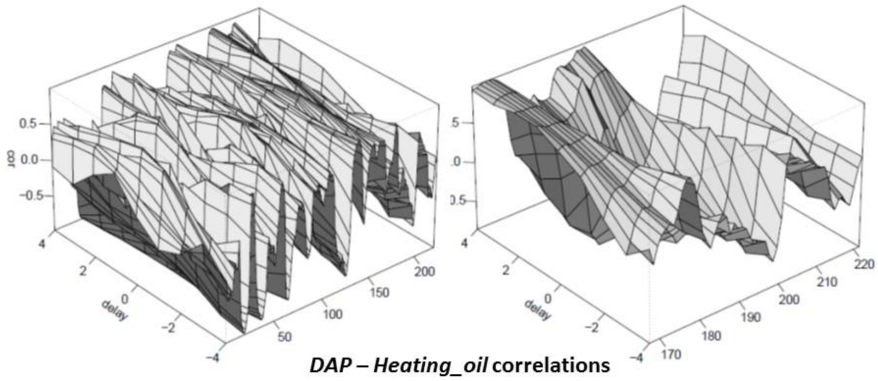

Fig. 7. Moving correlations *DAP – Heating_oil*.

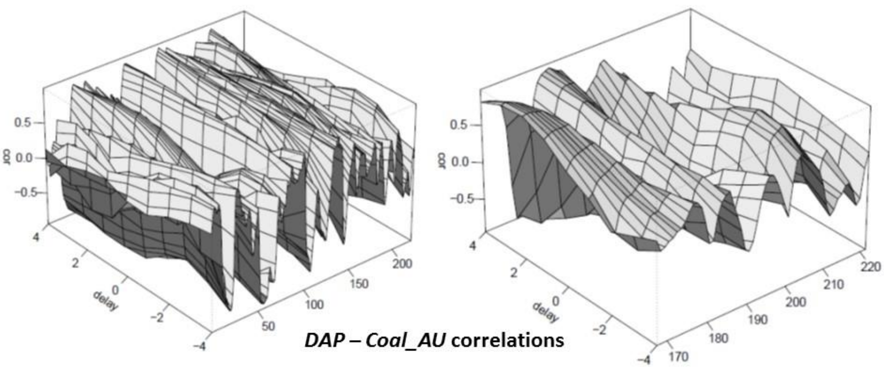

Fig. 8. Moving correlations *DAP – Coal_AU*.

For all of the pairs of variables, the correlation is very unstable in longer period – the effect of changes in macroeconomics together with different fluctuations. Even observing last 5 years (charts on the right-hand side), the trends of correlation change several times. Therefore current dependencies and time range of observations must be taken into account while building the regression model. Usually such a question is not examined and only formal assumptions appropriate for the selected statistical method are checked (like autocorrelation and normality tests, homoscedasticity and sometimes stationarity – for time series). Moreover, the quality of regression model is estimated based on chosen *ex post* and *ex ante* measures without deeper analysis of used data. However, another approach is possible in which simplified statistical inference may be limited to prediction of future correlation only with very simple nonlinear regression. But such approach cannot be applied directly for quantitative forecasts because of existing interdependencies between predictors. Most often such variables coming from micro- or macro-economic environment are not totally independent. Experiment presented below shows how unreliable may be using regression methods without a deeper understanding of the phenomena having an impact on the predictors.

2.4. Regression models built based on different time range of predictors

An experiment was carried out for the same set of variables together with their squares (taking into account existing non-linearity) – with *DAP* as a dependent variable. It was assumed to lag predictors (2 periods) – in order to calculate short-term forecast. OLS method was chosen as the most convenient for analyses, because model parameters directly determines impact of each predictor. The investigations were performed for the following periods:

1. full range: November 1995 till October 2015,
2. November 1995 till October 2000,
3. November 2000 till October 2005,
4. November 2005 till October 2010,
5. November 2010 till October 2015,
6. last 18 months (the minimum number of observations necessary for OLS method),
7. last 6 months – only predictors with monotonic correlation during this period were chosen.

First, the examination of residuals with the tests: autocorrelation of residuals, normality of residuals and homoscedasticity of residuals was carried out – assumptions fulfilled.

The results, containing parameters of regression model (if statistically significant, at significance level 0.05) are presented in Table 1, coefficients of determination R^2 and mean square errors MSE are placed in Table 2. These regression models were compared with two simple benchmark models: naïve forecast (*naïve* case) and naïve method with drift (*drift* case) taking into account *ex post* measures: mean absolute error *MAE* and mean absolute percentage error MAPE – see Table 3.

The last model is very interesting – with only one predictor. Additionally – it is very logical and appropriate model for the customers, because quotation of product depends on quotation of main raw material. Moreover, such an equation can be easily used to define widely accepted price formulas. Other models are proper from the formal point of view - but there is a danger of overfitting and probable different impact predictors in further periods. However, it should be pointed out that the forecast with simple benchmark method – naïve with drift, is almost as accurate as the forecast with last regression model (*case 7*), although perception of such benchmarks by the market may be negative as not dependent on any benchmark quotation.

variable	**case 1**	**case 2**	**case 3**	**case 4**	**case 5**	**case 6**	**case 7**
Coal_AU	0.356	0.464	0.246	0.415	-	16.91	-
Coal_AU^2	0.230	-	-	-	1.052	-16.62	-
Brent	0.629	-	-	0.707	-	-	-
Brent^2	-2.69	-	-	-2.035	-1.574	-	-
Gasoline	-	-	-	-	-	-	-
Gasoline^2	0.439	-0.533	-	-	1.300	-	-
Heating_oil	-	-	-	-	-	-5.524	-
Heating_oil^2	1.948	-	0.435	1.822	-	7.314	-
Natural_gas	-	-	-	-	-	-	-
Natural_gas^2	-	0.230	-	-	-	-1.860	0.964
Propane	-1.118	-1.670	-	-	-	-	-
Propane^2	0.934	1.427	-	0.236	-	-	-
Urea	0.189	-	1.988	0.909	-	-	-
Urea^2	-	0.330	-1.567	-0.605	-	-	-

Table 1. Parameters of regression models.

measure	**case 1**	**case 2**	**case 3**	**case 4**	**case 5**	**case 6**	**case 7**
R^2	0.967	0.965	0.955	0.983	0.849	0.928	0.964
MSE	0.223	0.024	0.047	0.222	0.156	0.030	0.013

Table 2. *Ex ante* measures of regression.

measure	**case 1**	**case 2**	**case 3**	**case 4**	**case 5**	**case 6**	**case 7**	**naive**	**drift**
MAE	0.187	0.073	0.089	0.211	0.109	0.079	0.022	0.052	0.020
MAPE	0.151	0.069	0.067	0.147	0.098	0.065	0.018	0.038	0.019

Table 3. *Ex post* measures of regression.

Conclusion

From the managers point of view, huge current instability of market behaviour means that each longer forecasts should be treated with great caution. In short time, however, even the knowledge of trends can lead to competitive advantage. Therefore the model based on known correlation, even for accepted non-stationary cases – and prediction of coming dependencies may be more useful than “standard” regression model. Moreover, its simplicity may be – among other things – helpful for price formulas. It must be, however, taken into consideration that such a prediction is rather qualitative due to existing cross-dependencies between predictors and risk of spurious correlation. Quantitative approach, more difficult but also more useful, will be developed in further studies.

References

Borio, C. (2014). The financial cycle and macroeconomics: what have we learnt? *Journal of Banking and Finance, 45*, 182-198.

Caballero, R. J. (2010). Macroeconomics after the crisis: time to deal with the pretense-of-knowledge syndrome. *Journal of Economic Perspectives*, *American Economic Association, 24*(4), 85-102.

Cariolle, J., Goujon, M. (2015). Measuring macroeconomic instability: a critical survey illustrated with exports series. *Journal of Economic Surveys, 29*, 1-26.

Gatnar, E. (2008). *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*. Warszawa: Wydawnictwo Naukowe PWN.

Hoffmann, A. O. I., Post, T., & Pennings, J. M. E. (2013). Individual investor perceptions and behavior during the financial crisis. *Journal of Banking and Finance, 37*(1), 60-74.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54*(1-3), 159-178.

Powell, J. L., & Stoker, T. M. (1985). Estimation of complete aggregation structures. *Journal of Econometries, 30*, 317-344.

Phillips, P. C. B., & Perron, P. (1988). Teting for a Unit Root in Time Series Regression. *Biometrika, 75*(2), 335-346.

Stock, J., & Watson, M. (2010). Modeling inflation after the crisis. Paper prepared for the *Federal Reserve Bank of Kansas City Symposium, “Macroeconomic Policy: Post-Crisis and Risks Ahead”*, Jackson Hole, Wyoming, August 26-28.