SFB
823

Discussion Paper

Time lags in the pass-through of crude-oil prices: Big data evidence from the German gasoline market

Manuel Frondel, Colin Vance,
Alex Kihm

Nr. 24/2015

SFB
823

Time Lags in the Pass-Through of Crude-Oil Prices: Big Data Evidence from the German Gasoline Market

Manuel Frondel, Rheinisch-Westfälisches Institut für Wirtschaftsforschung and Ruhr University Bochum

Colin Vance, Rheinisch-Westfälisches Institut für Wirtschaftsforschung and Jacobs University Bremen

Alex Kihm, Rheinisch-Westfälisches Institut für Wirtschaftsforschung and Fairr.de

Abstract. This note investigates the pass-through of global Brent oil notations to fuel prices across the oligopoly of retail majors in Germany. We assemble a high-frequency panel data set that encompasses millions of price observations and allows us to distinguish effects by brand. Upon establishing a cointegrating relationship between fuel and crude-oil prices using daily data, we estimate an error-correction model (ECM) and find that (1) the pass-through of oil prices critically depends on the number of time lags included in the ECM, (2) strict adherence to classical information criteria for determining lag length yields extremely long pass-through durations, and (3) the estimated impulse response functions are virtually identical across brands, irrespective of the lag count, suggesting a high degree of competition among brands.

JEL classification: D12, Q41.
Key words: Retail Markets, competition, error-correction model.
Correspondence: Manuel Frondel, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Hohenzollernstr. 1-3, D-45128 Essen. frondel@rwi-essen.de.

Acknowledgements: We are grateful for invaluable comments and suggestions by Christoph M. Schmidt and Reinhard Madlener. This work has been supported by the NRW Ministry of Innovation, Science, and Research (BMBF) within the framework of the project "Rebound effects in NRW" and by the Collaborative Research Center "Statistical Modeling of Nonlinear Dynamic Processes" (SFB 823) of the German Research Foundation (DFG), within the framework of Project A3, "Dynamic Technology Modeling".

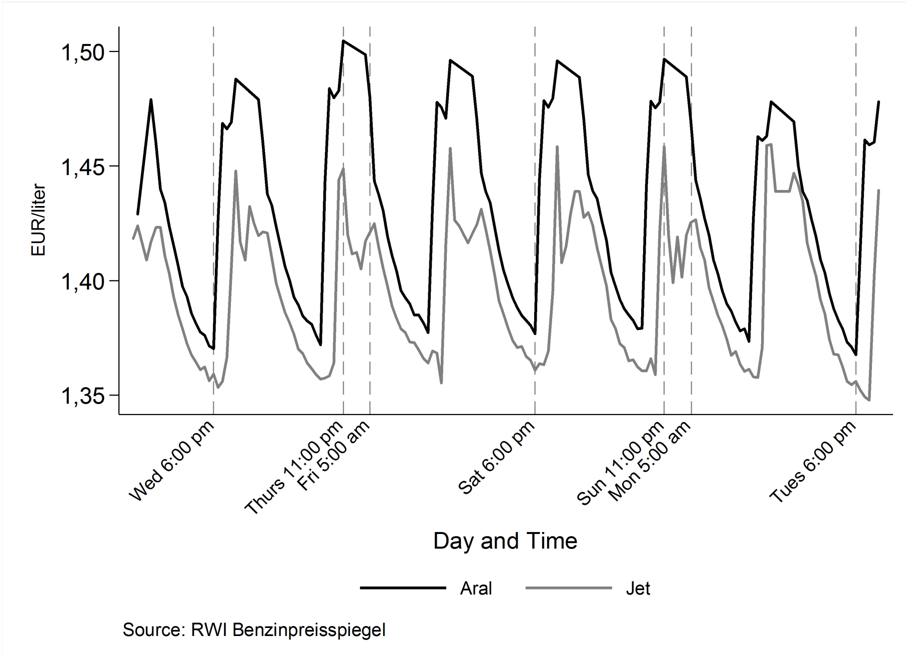
1 Introduction

Drawing upon a huge panel data set originating from a recently established census of retail prices covering virtually all fuel stations in Germany, this note investigates the pass-through of global Brent oil notations to gasoline prices, thereby distinguishing between retail majors, minors, and independents. Gasoline markets are well-known to exhibit retail price evolutions that resemble the Edgeworth price cycle equilibria formalized by MASKIN and TIROLE (1988), which can have implications for the speed of gas price responses (LEWIS, NOEL, 2011). Such cycles have been found for the US (LEWIS, 2009; DOYLE, MUEHLEGGER, SAMPAHANTHARAK, 2010), Canada (ECKERT, 2003; NOEL 2007a,b), and Australia (WANG 2008), with a typical cycle lasting one to two weeks (LEWIS, NOEL, 2011:672).

Fluctuations in German fuel prices are likewise characteristic of an Edgeworth Cycle, but one that takes place over a 24 hours period, rather than weeks. Figure 1 presents this pattern for E5 gasoline and the retailers Aral and Jet, but is also representative for the other fuel types and retailers. The fuel price reaches a trough each day at about 6:00 p. m. , after which it rises rather sharply until 11:00 p. m. , stagnating until 5:00 a. m. , and thereafter falling gradually over the course of the day until 6:00 p. m. When averaging the prices on a daily basis, however, the evidence for a cyclical Edgeworth pattern vanishes.

Moreover, using an error-correction based cointegration test for panel data (WESTERLUND, 2007; PERSYN, WESTERLUND, 2008), a cointegrating relationship between fuel and Brent prices is not rejected with the daily data, contrasting with a rejection of cointegration using the hourly data. In what follows, we use this result to apply the standard error-correction model (ECM) of ENGLE and GRANGER (1987) to the daily data to investigate both the critical role of the lag order in the pass-through of crude-oil prices and the degree of competition among major brands. To compare the price pass-through speed of different brands, impulse response functions (IRFs) are subsequently estimated.

Figure 1: **Intra-Day Price Cycles for E5 Gasoline in Germany**.



Three main results emerge. First, we find that the estimated pass-through of oil prices critically depends on the number of time lags included in the ECM. Second, strict adherence to classical model selection criteria, such as AKAIKE's (1973) and SCHWARZ' (1978) information criteria, yields an extremely long pass-through period, leading us to advocate discretionary limits on the number of lags included. Lastly, irrespective of the number of lags included in the model, the differences in the associated IRFs across brands is negligible, which is interpreted as evidence for a competitive retail market.

The following section describes the panel data set. Section 3 provides a description of the estimation method, followed by a derivation of the formula for the impulse response function (IRF). The presentation and interpretation of the results is given in Section 4. The last section summarizes and concludes.

2 Data

The German retail market for gasoline and other fuels is dominated by an oligopoly of five vertically integrated oil companies that have a large network of stations and direct access to refining capacities: Aral, Shell, JET, Esso and Total (Table 1). These players have long been a source of scrutiny by Germany's Cartel Office (BUNDESKARTELLAMT, 2011:20-21). Increasing concern about collusion culminated in the establishment of the so-called Market Transparency Unit for Fuel and an on-line portal that posts fuel prices in real-time from each of Germany's roughly 14,000 filling stations.[1]

Since September 2013, stations are legally obligated to post every price change, the precise time stamp, the geographic coordinates of the station, the opening hours, and the brand. To access this data, we wrote a script that continuously retrieves entries on the site and stores these on a server. From the raw data, we create a balanced panel of daily prices for E5 and E10 gasoline, as well as diesel, charged by each station covering the period from May 17, 2014, to March 14, 2015, and resulting in millions of price observations altogether.

Table 1: Mean Gasoline Prices (E5) Across Retailers in Germany (May 17, 2014 - March 10, 2015)

	Mean (€/Liter)	Std. Dev.	# Stations	# Days
Aral	1.505	(0.118)	2,270	298
Esso	1.490	(0.006)	1,023	298
Jet	1.465	(0.110)	575	298
Shell	1.508	(0.119)	1,774	298
Total	1.498	(0.116)	714	298
Minors and independents	1.470	(0.115)	6,511	298

Note: Average Brent Oil prices amounted to 0.42 €/liter over the same time interval.

[1]For more information on the Market Transparency Unit for Fuel (Markttransparenzstelle für Kraftstoffe, MTS-K), see `http://www.bundeskartellamt.de/EN/Economicsectors/MineralOil/MTU-Fuels/mtufuels_node.html`.

For this period and the example of E5 gasoline, mean prices across brands are presented in Table 1. The highest average price, at 1.508 €per liter, is to be observed for Shell, whereas Jet exhibits the lowest average price of 1.46 €/liter. Prices are in nominal terms and include a 65 cents excise tax, as well as a 19% value-added tax. Following standard practice, we estimate the ECM on the before-tax gas prices, using daily data on Brent oil prices published by the U.S. Energy Information Administration (EIA).

3 Methodological Issues

To model the transmission of crude-oil prices, PC, to gasoline prices, PG, we follow BACHMEIER and GRIFFIN (2003). These authors abstract from determinants other than crude-oil prices, arguing that crude oil is the principal input to gasoline production and that the purpose of their model is simply to examine the transmission of crude-price shocks to gasoline prices. Furthermore, we exploit the fact that average daily gasoline prices do not exhibit Edgeworth cycles, thereby allowing us to employ a standard ECM (BACHMEIER, GRIFFIN, 2003:773):[2]

$$\Delta PG_t = \sum_{i=0}^{k} \beta_{ci} \Delta PC_{t-i} + \sum_{i=1}^{n} \beta_{gi} \Delta PG_{t-i} + \theta z_{t-1} + \varepsilon_t, \tag{1}$$

where β_{ci} and β_{gi} measure the short-run impact of crude oil prices and lagged gasoline prices, respectively, θ is the long-run equilibrium parameter and

$$z_t = PG_t - \gamma_0 - \gamma_1 PC_t \tag{2}$$

measures the long-run disequilibrium between gasoline and crude-oil prices. γ_1 reflects the long-run effect of a permanent change in crude-oil prices. As we have empirically found that the PC and PG time series are cointegrated, the long-run relationship follows a stationary process, as well as the other regressors in (1), which are found to

[2]Using a Markov switching regression framework, LEWIS and NOEL (2011:672) argue that in markets that exhibit price cycles, distributed lag models, such as the ECM, are unable to capture the large and periodic changes in retail margins.

be integrated of order one. Hence, inference on functions of the coefficients, such as the impulse response function (IRF), is standard.

The impulse response – or cumulated adjustment – function, recursively defined by $IRF_t := PG_t - PG_{t-1} + IRF_{t-1} = \Delta PG_t + IRF_{t-1}$, measures the t-period cumulative response in gasoline prices to a one-time, but permanent unit change in the price of crude oil at $t = 0$: $PC_t = 1$ for $t = 0, 1, 2, ...$ Our derivation of the IRF leads to a formula very similar to that presented by BORENSTEIN, CAMERON, and GILBERT (1997). For starters, for $t = 0$, we obtain

$$IRF_0 = PG_0 - PG_{-1} + IRF_{-1} = \widehat{\beta}_{c0}(PC_0 - PG_{-1}) + \widehat{\beta}_{g1}(PG_{-1} - PG_{-2}) + \hat{\theta} z_{-1} = \widehat{\beta}_{c0},$$

because $IRF_{-1} = 0 = PC_{-1} = PG_{-1} = PG_{-2}, z_{-1} = 0$, as the one-unit shock occurs in $t = 0$. For $t = 1$ and $k, n \geq 1$, it is

$$\begin{aligned} IRF_1 &= PG_1 - PG_0 + IRF_0 = \widehat{\beta}_{c0}\Delta PC_1 + \widehat{\beta}_{c1}\Delta PC_0 + \widehat{\beta}_{g1}\Delta PG_0 + \hat{\theta} z_0 + IRF_0 \\ &= \widehat{\beta}_{c1} + \widehat{\beta}_{g1} IRF_0 + \hat{\theta}(IRF_0 - \gamma_1) + IRF_0, \end{aligned}$$

because $\Delta PC_0 = PC_0 - PC_{-1} = 1 - 0 = 1$ and $\Delta PC_1 = PC_1 - PC_0 = 1 - 1 = 0$, as the unit change in $t = 0$ is permanent, and $\Delta PG_0 = IRF_0$. Furthermore, z_0 results from $z_0 = z_0 - z_{-1} = \Delta PG_0 - \gamma_1 \Delta PC_0 = IRF_0 - \gamma_1$, as $\Delta PC_0 = 1$ and $\Delta PG_0 = IRF_0$.

Likewise, for $t = 2$ and $k, n \geq 2$, because of $\Delta PC_2 = \Delta PC_1 = 0$ and $\Delta PC_0 = 1$, we get

$$\begin{aligned} IRF_2 &= PG_2 - PG_1 + IRF_1 = \widehat{\beta}_{c0}\Delta PC_2 + \widehat{\beta}_{c1}\Delta PC_1 + \widehat{\beta}_{c2}\Delta PC_0 + \\ &\quad \widehat{\beta}_{g1}\Delta PG_1 + \widehat{\beta}_{g2}\Delta PG_0 + \hat{\theta} z_1 + IRF_1 \\ &= \widehat{\beta}_{c2} + \widehat{\beta}_{g1}(IRF_1 - IRF_0) + \widehat{\beta}_{g2} IRF_0 + \hat{\theta}(IRF_1 - \gamma_1) + IRF_1, \end{aligned}$$

since, by definition, $\Delta PG_1 = IRF_1 - IRF_0$ and $\Delta PG_0 = IRF_0$. In addition, $z_1 - z_0 = \Delta PG_1 - \gamma_1 \Delta PC_1 = IRF_1 - IRF_0$ and, hence, $z_1 = z_0 + IRF_1 - IRF_0 = IRF_0 - \gamma_1 + IRF_1 - IRF_0 = IRF_1 - \gamma_1$. Note that the formula for z_1 can be generalized by recursive induction to $z_t = IRF_t - \gamma_1$ for all $t \geq 0$.

In sum, as has been motivated by calculating IRF_t for $t = 0, 1, 2$, the general formula for $t = j$ reads:

$$IRF_j = \widehat{\beta}_{c_j} + \sum_{i=1}^{j} \widehat{\beta}_{gi}(IRF_{j-i} - IRF_{j-i-1}) + \hat{\theta}(IRF_j - \gamma_1) + IRF_{j-i}. \tag{3}$$

It bears noting that $\widehat{\beta}_{c_j} = 0$ if $j > k$ and $\sum_{i=1}^{j} \widehat{\beta}_{gi}(IRF_{j-i} - IRF_{j-i-1}) = \sum_{i=1}^{n} \widehat{\beta}_{gi}(IRF_{j-i} - IRF_{j-i-1})$ if $j > n$. Finally, the long-term equilibrium $IRF := \lim_{k\to\infty} IRF_k$ is given by $IRF = \gamma_1$, as can be seen from formula (3) by setting $IRF_j = IRF$ for all j.

4 Empirical Results

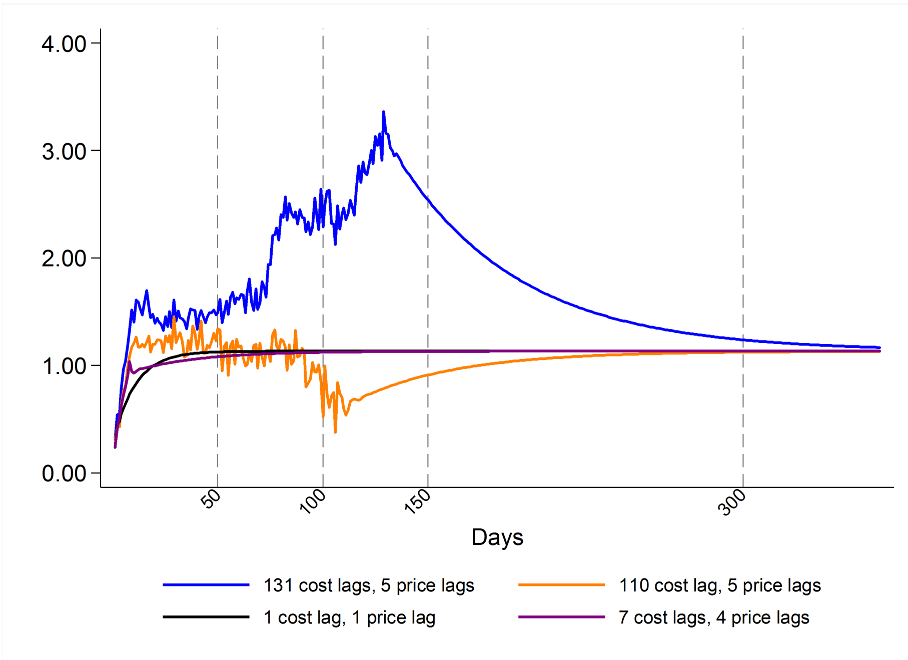
An important step in estimating an ECM is the specification of the lag lengths k and n: employing too few lags risks biased estimates, while including too many lags compromises precision and may lead to an over-fitted model that generalizes poorly. Various techniques have been employed for determining lag length, including direct testing of the statistical significance of the lagged terms (BORENSTEIN, CAMERON, GILBERT, 1997), expert discretion (LEWIS, 2011) and, perhaps most commonly, the application of information criteria (BACHMEIER, GRIFFIN, 2003), such as the Akaike and Bayes Information Criterion (AIC and BIC, respectively).

As HAN, PHILLIPS, and SUL (2015) demonstrate, the application of the BIC in the context of dynamic panel models can be problematic, leading to considerable overestimation of the lag order. These authors propose alternative model selection methods, two of which modify the BIC by increasing the penalty, whereas another approach, called the truncated sample method, truncates the sample based on the highest lag order, with the consequence that the comparison of the BIC references the same sample.

We have explored alternative techniques for determining lag lengths, finding that all methods using information criteria, including those suggested by HAN, PHILLIPS, and SUL, result in extremely long – and seemingly implausible – lag orders for the cost variable, i. e. the Brent crude oil price. Moreover, the shape of estimated IRFs is found to be highly sensitive to the lag lengths. The degree of variation is illustrated by

Figure 2, presenting select IRFs for the panel of Aral stations. The longest pass-through duration, estimated at about 350 days, results from a model with 5 lags of retail prices and 131 lags of Brent prices, determined using the truncated sample method.

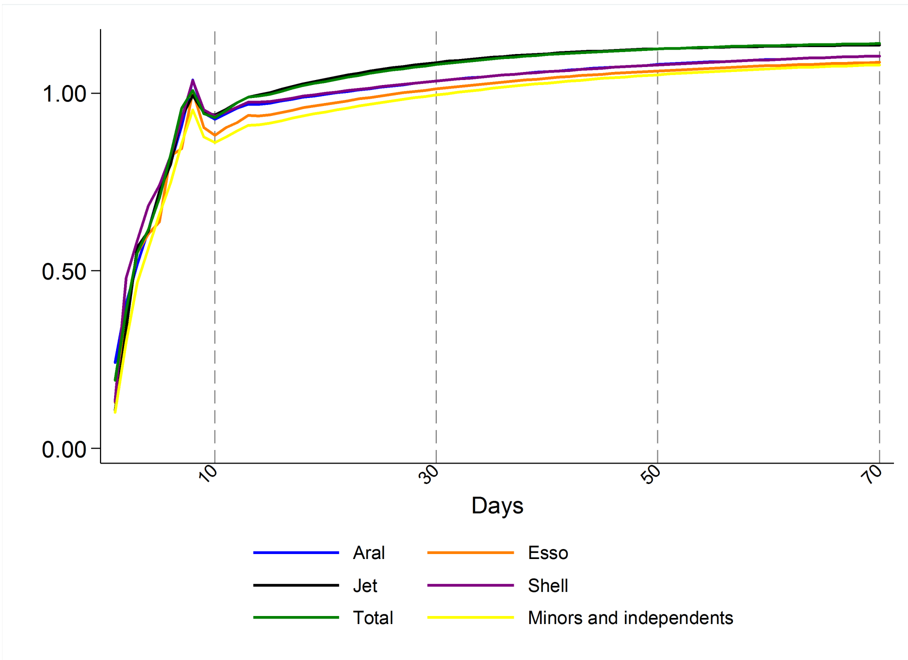
Figure 2: **Impulse Response Functions by Lag Length for Aral**.



Reducing the oil price lag to 110, where the BIC reaches a local minimum, results in a markedly different path whose pass-through time is considerably shorter, at about 200 days. We have also estimated two IRFs based on ECM specifications taken from the literature, yielding much shorter, more plausible pass-through times: First, a parsimonious variant specified by BACHMEIER and GRIFFIN (2003) using the BIC, includes one lag of the oil price and one retail price lag, resulting in a pass-through of 30 days. A second specification includes 4 retail price lags and 7 oil price lags, a selection used by LEWIS (2011) in citing its similarity with previous studies. This results in a longer pass-through of about 60 days.

Notwithstanding the heterogeneity evident in Figure 2, we find a high degree of stability in the estimated IRFs across brands. Figure 3 presents the IRFs generated by the model with 4 price lags and 7 cost lags, documenting that the trajectories are statistically indistinguishable. We have explored a multitude of other specifications, finding that the different brands always follow a similar convergence path, irrespective of the specified lag orders. This result may reflect price setting close to marginal costs, so that stations have limited leeway in absorbing oil price shocks and follow a highly similar path of adjustment with their competitors.

Figure 3: **Impulse Response Functions by Brand, 7 Cost Lags, 4 Price Lags**.



5 Summary and Conclusion

Drawing upon a huge panel data set entailing millions of fuel price values that originate from a recently established census of retail prices covering virtually all fuel stations in Germany, this note has investigated the pass-through of Brent oil prices, the primary cost factor not only for German fuel retailers. After deriving and estimating impulse response functions for standard error-correction models, we have explored the consequences of different lag specifications – selected on the basis of classical information criteria – for the estimated pass-through time.

Along the lines of LEWIS and NOEL (2011: 674), we find that statistical procedures to determine the proper lag length do not work well in our application. Even when using a penalized variant of the Bayes Information Criterion, as suggested by HAN, PHILLIPS, and SUL to handle dynamic panel models, we obtain a model specification that results in an extremely long pass-through time of nearly one year. Following shorter lag specifications that are established in the literature results in an estimated pass-through time of 6 to 8 weeks, which is within the range identified in previous studies (e.g. BORENSTEIN, CAMERON, and GILBERT, 1997; BACHMEIER, GRIFFIN, 2003; LEWIS, NOEL 2011). Most notably, we find that the IRF trajectories are highly similar across brands for given lag lengths, a likely reflection of competition.

References

AKAIKE, H. (1973) Information Theory and an Extension of the Likelihood Ratio Principle. In: *Second International Symposium of Information Theory*. Editors: B. N. PETROV and F. CSAKI. Minnesota Studies in the Philosophy of Science. Budapest: Akademinai Kiado.

DOYLE, J. , MUEHLEGGER, E. , SAMPAHANTHARAK, K. (2010) Edgeworth Cycles Revisited. *Energy Economics*, **32(3)**, 651-660.

ECKERT, A. (2003) Retail Price Cycles and the Presence of Small Firms. *International Journal of Industrial Organization*, **21(2)**, 151-170.

ENGLE, R. F. , GRANGER, C. W. J. (1987) Co-Integration and Error-Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2), 251-276.

FRONDEL, M., VANCE, C. (2010) Fixed, Random, or Something in Between? A Variant of Hausman's Specification Test for Panel Data Estimators. *Economics Letters*, **107**, 327-329.

HAN, C. , PHILLIPS, P. C. B. , SUL, D. (2015). Lag Length Selection in Panel Autoregression, Econometric Reviews, forthcoming.

HAUCAP, J. , HEIMESHOFF, U. , SIEKMANN, M. (2015) Price Dispersion and Station Heterogeneity on German Retail Gasoline Markets. DICE Discussion Paper No. 171, Düsseldorf Institute for Competition Economics (DICE).

LEWIS, M. (2009) Temporary Wholesale Gasoline Price Spikes Have Long-Lasting Retail Effects: the Aftermath of Hurricane Rita. *Journal of Law and Economics*, **52(3)**, 581-605.

LEWIS, M. , NOEL, M. (2011) The Speed of Gasoline Price Response in Markets With and Without Edgeworth Cycles. *Review of Economics and Statistics*, **93(2)**, 672-682.

MASKIN, E. , TIROLE, J. (1988) A Theory of Dynamic Oligopoly II: Price, Competition, Kinked Demand, and Edgeworth Cycles. *Econometrica*, 56(3), 581-599.

NOEL, M. (2007a) Edgeworth Price Cycles, Cost Based Pricing and Sticky Pricing in Retail Gasoline Markets. *Review of Economics and Statistics*, **89(2)**, 324-334.

NOEL, M. (2007b) Edgeworth Price Cycles: Evidence from the Toronto Retail Gasoline Market. *Journal of Industrial Economics*, **55(1)**, 69-92.

PERSYN, D. , WESTERLUND, J. (2008) Error-Correction-Based Cointegration Tests for Panel Data. *Stata Journal*, **8** (2), 232-241.

SCHWARZ, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics* 6, 461-464.

WANG, Z. (2008) Collusive Communication and Pricing Coordination in a Retail Gasoline Market. *Review of Industrial Organization*, **32(1)**, 35-52.

WESTERLUND, J. (2007) Testing for Error Correction in Panel Data. *Oxford Bulletin of Economics and Statistics*, **69**, 709-748.