Brunel University London
Department of
Economics and Finance

Working Paper No. 13-03

Economics and Finance Working Paper Series

John Hunter and Seyedeh Asieh Tabaghdehi

Cointegration and US Regional Gasoline Prices: Testing market efficiency from the stationarity of price proportions

February 2013

http://www.brunel.ac.uk/economics

Cointegration and US Regional Gasoline Prices: Testing market efficiency from the stationarity of price proportions

John Hunter, Seyedeh Asieh Tabaghdehi*
Department of Economics and Finance, Brunel University, School of Social Science, Uxbridge, Middlesex, UB8 3PH, UK

Abstract: It is well known that oil price shocks are a major concern to the health of the global economy. Unstable oil prices have a significant negative impact on consumer confidence and business decision making. As a result economic recovery may be longer and more complicated. Controlling the global oil price may not be possible, but a main concern of this research relates to energy market efficiency and as to whether relative price differences respond in an appropriate way across the region of one country. Here, the different geographic areas of a country are analysed to see whether they belong to the same market and result in the relative prices being stationary. The gasoline market in different regions of the US is analyzed. It is believed that if the gasoline market is sufficiently active in the US, then as a result of arbitrage, long-run gasoline prices by region should follow each other. In an efficient market, a price shock in one region would be reflected in all other prices. This proposition is tested in an effective manner by a barrage of stationarity tests. This is pertinent as such tests have been applied in antitrust cases in Italy and the Netherlands to determine whether there might be market imperfections or in association with more heuristic information, possible collusion.

JEL Classification: C32, D18, D40

Keywords: Gasoline, Stationarity, Cointegration, ARCH, Price differential, Market Efficiency, Arbitrage, Collusion

---

* We would like to thank Ray Barrell, Iain Clacher, Matthew Greenwood Nimmo, Guy Liu, Kevin Reilly, Cheryl Roberts, Yongcheol Shin and contributors at the Micro-econometrics and Public Policy Conference at the National University of Ireland, Galway for their comments. Corresponding author: Seyedeh.Tabaghdehi@brunel.ac.uk.

1- **INTRODUCTION**

Oil prices are currently of particular concern for the health of the global economy, oil price anomalies are likely to harm consumer confidence and business decision making. It has been suggested recently that oil prices have shifted from the control of OPEC to the global market for oil and as a result in recent times it can be viewed as being less susceptible to political intervention.

It may be possible to determine whether the market is efficient and this can be observed via the finding of arbitrage in the long-run. According to the US Energy Information Administration (EIA) report (2007), US refineries produced over 90% of the gasoline used in the US, but less than 40% of the crude oil used to produce the gasoline comes from the US.

Much of the gasoline produced in the US approximately 45% comes from refineries on the Gulf Coast (including Texas and Louisiana). However, a number of factors could cause gasoline price variations across regions of a country like the US: tax differentials; the cost of crude oil; refining costs and profits; distribution and marketing costs and profits; distance from supplier; supply disturbances; retail competition; environmental programs. So where fuel is produced and how it is distributed, may have a significant effect on the price of gasoline sold to the consumer. The focus of attention here is to evaluate whether the market for petroleum products is efficient in a long-run sense and this occurs by determining the extent to which relative prices can be differ across the regions of one country.

[Figure 1 goes here]

Figure 1 shows the distribution of gasoline taxes across the United States that is combined local, state and federal. Red areas indicate high tax levels (greater than 47.70 cents per gallon) on gasoline, yellow areas medium tax levels (between 40 and 47.70 cents per gallon) and blue areas signify lower taxes (less than 40 cents per gallon). The US average tax level for gasoline is 47.70 cents per gallon. Unless there is a trend in tax differentials then these discrepancies ought not to affect the long-run and short-run effects that are stable ought to be captured by the intercept in the model.

This research is primarily empirical and econometric with the innovation relating to the methods applied to detect what was termed by Forni (2004) in the context of regional milk prices in Italy “a broad market”. The study of Forni was applied to the price of a homogenous product over time to determine the dimension of either the product or the geographic market and advise the regulatory authorities over the merits of further concentration in the industry (Lexecon, 2001). Price data have been analysed in a similar way using panel methods by Giulietti et al (2010) to determine anti-competitive behaviour for energy markets in the UK.

An alternative way of analysing the market for gasoline is as a commodity and instead of viewing that as an issue related to competition we might consider the observation of arbitrage as a sign of informational rather than product market efficiency. When a market is efficient in an informational sense, then price signals that impact the market will give rise to price movements that will lead to the removal of mispricing. In a commodity market, prices adjust to eliminate mispricing and this implies that there is arbitrage across prices at least in the long-run or that we find long-run arbitrage correction.

Why should gasoline prices be dissimilar in different geographic areas in the US? Does this mean the energy market is not efficient and how can it be made more efficient? If the market is efficient, oil prices in different regions of a country should follow each other. When price behaviour alone is observed, then the shocks respond to factors that affect both demand and supply across the industry and when all firms respond to these movements then the market should be efficient in the long-run. Here price shocks in one region of the US are expected to impact all the other prices. No suggestion is being made that arbitrage applies in the short-run, though it ought to be noticed that the structure of the Dickey Fuller test model imposes a common factor restriction (Hendry and Mizon, 1978) that forces the long and short-run responses to be the same and that means here that the market is efficient (Burke and Hunter, 2005, Chapter 3). One conclusion of the observation of inefficiency would be that there is collusion and this was made by London Economics (2002) in their study for Dutch regulator of the Netherlands mobile telecommunications market.

Next this article considers the literature, then the methodology and then the results are analysed. Finally our conclusions are offered.

2- REVIEW OF ESSENTIAL LITERATURE

Considering the differences in taxes seen in plot 1, the gasoline prices are not exactly the same across regions of the US, but in a market that is informationally efficient then price shocks in one region should be reacted to another region as a result of arbitrage. Burke and Hunter (2011) show that when a price in one market reacts to a long-run equilibrium price target (LEPT), then this corresponds to all prices in the market responding in proportion or what has been termed parallel pricing by Buccirossi (2006). Gasoline is a relatively homogenous product and in this paper we evaluate such prices without considering technological differences across primary gasoline outlets across the United States.

The proposition that underlines the idea that the market is efficient is that all prices fully reflect all information across the system. One might consider the possibility of anomalies in the short-run, but in the long-run minor differences ought to be smoothed out. Hence one would anticipate prices should respond to the stochastic behaviour that underlines prices that operate in the market and this behaviour is summarized by a stochastic trend.

Some of the early literature on the nature of collusive behaviour suggested this would be indicated by strong correlations between prices (Maunder, 1972). Such price correspondence was seen as a signal of pricing decisions being made in concert and this would be suggestive of collusion especially with imperfect competition. Posner (1976) considered current price as a key factor in estimating the impact of mergers. A key aim of the most competition agencies is to recognize an active monopoly or detect collusion. To this end, an Antitrust Agency's target is to avoid creating a new firm with the potential to exert market power and increase the over-all price level in the market.

The emphasis has changed more recently as the similarity of price movements can also be considered as a signal of an effectively functioning market. The market efficiency proposition implies that commodity prices ought to fully reflect all information, while the associated concept of arbitrage suggests that prices of the same good are likely to move together. Stigler and Sherwin (1985) suggested that products be grouped in a single market when prices move together and that gives support to price co-movement being linked to efficiency when a market is categorized by homogenous products. Furthermore in a theoretical setting Buccirossi (2006) has made the observation that price setting being independent is not necessarily consistent with behaviour that can be objectively described as collusive. Buccirossi finds in a short-run sense, price responses may be different from unity and in certain types of model this behaviour may not be inconsistent with price competition.

Empirically, a range of econometric and time series methods have been proposed to analyse the time series properties of price series to indicate anti-competitive behaviour (see, Horowitz, 1981, Slade, 1986, and Uri, Howell, and Rifkin, 1985). A review of quantitative methods applied to the analysis of competition cases was prepared by LecG (1999) for the United Kingdom Office of Fair Trading (OFT). The LecG report points out that causality and correlation may need to be used in concert to distinguish between competitive and non-competitive behaviour. Further, the need to consider the notion of non-stationarity is emphasised in relation to a modern treatment and as a result distinguish between the long and the short-run.

The argument used in Buccirossi (2006) suggests that it is not conclusive in a short-run sense that prices ought to move in parallel that leads to the conclusion that this can only be considered a long-run property. It would also seem easier to frame a legal argument in the context of long-run behaviour as short-run pricing anomalies may be seen less as the result of aggressive price leadership (Markham, 1951) and more likely as a result of arbitrage or at most be indicative of barometric pricing (Koutsyiannis, 1975).

This leads Forni (2004) to argue that if two product or geographic areas belong in the same market for the purposes of antitrust legislation, their relative log price ratio must be stationary. Hence, stationarity tests such as the Augmented Dickey-Fuller (ADF) test and KPSS test can be useful in describing the related nature of markets. If one considers the non-stationarity of the log of the price ratio, then this is indicative of the distinct nature of a geographic market. Testing whether price proportions are stationary can be seen as a technically efficient way of determining whether prices cointegrate where the proportionality of prices is imposed by the structure of the stationarity/non-stationarity test. Forni considers it to have significant advantages to methods that on one hand consider the elasticity of the residual demand function and on the other cointegration. The test implies that each price within a well defined product market captures a component of the non-stationarity that arises from the stochastic trend driving underlying behaviour. Hence, price proportions in an effectively functioning market ought to be stationary. Forni (2004) mentions that this analysis links strongly with the literature on testing stationarity of the real exchange rate, and Beirne, Hunter and Simpson (2007) consider a joint analysis using univariate and panel methods.

Forni (2004) defined the notion of what he termed a broad market definition to characterise market efficiency in relation to a highly homogenous product, milk. Pricing behaviour across Italy is only limited by one significant natural physical break related to the Straits of Messina lying between Calabria/Puglia and Sicily. With the exception of regulation that forces milk sales to be limited to four days from production and suggests some constraint on sales between the far north and south of the country little else ought to limit arbitrage.

Burke and Hunter (2011) consider the idea of long-run equilibrium price targeting (LEPT) to describe the price setting behaviour of firms. If prices in a market are driven by arbitrage, then subject to the existence of an appropriate number of long-run relations, then arbitrage is consistent with firms following LEPT. In the context of firm specific prices as arose in the analysis of the Dutch market for mobile termination charges by London Economics (2002), then it is in principle possible to further distinguish between parallel pricing and aggressive price leadership when one of a sequence of prices is weakly exogenous (Hunter and Burke, 2007, and Kurita, 2008). In terms of the market data analysed here, weak exogeneity implies that the price process in the other regions may be driven by the weakly exogenous price. However, this price can only be tied down to a single company via the ownership of the refining capacity as compared with oil production.

In the gasoline market it is suggested that the existence of refiners in a region can create differences in price levels. Hence, a further concern of this research is not the existence of such differences but the stability of price relations over time. It is suggested that the price of gasoline in different geographic areas of the same market, should not be different from each other in the long run otherwise there is an arbitrage opportunity. The supporting argument is the statistical concept of stationarity as applied to price proportions.

As was stated above testing stationarity of price proportions is an inferentially effective way to test for cointegration between log prices and when these series are stationary then the result is consistent with long-run arbitrage. This is especially pertinent when as is the case for Forni (2004) and London Economics (2002) the time series sample is small. Here this approach is followed, but unlike Forni we limit ourselves to a subset of price proportions. First the upper triangle of stationarity tests for equations for which the lag order has been securely selected ought to be the same as the reverse calculation associated with the lower triangle.[1] Second, the stationarity tests impose a restriction that implies that arbitrage is imposed to the short and the long-run as the variable tested is a price proportion and this is the dependent variable in the stationarity tests (ADF or KPSS). This implies that a sequence of restricted dynamic models is being estimated that correspond with the stationarity tests of price proportions, but this calls into question the number of coherent price models that might be computed. A similar problem arises when the real exchange rate is tested for stationarity using cross rates (Smith and Hunter, 1985, and Hunter and Simpson, 2004). This implies that there is a limit to the number of dynamic price models related to a fixed number of single equation error correction models from which stationarity might also be tested.

Subject to the above caveats, the approach of Forni (2004) is followed to suggest that gasoline prices be proportional across the US and that this would confirm a broad market or a single market definition in US gasoline market, and thus determine a link to long-run arbitrage and efficiency.

---

[11] This result can be readily shown to apply when the underlying equations relate to the Dickey-Fuller test and it can be shown analytically that the intercept is the reverse coefficient and the coefficient on the lag level is the same irrespective of the way round the regression is run.

Hosken and Taylor (2004) in their extended discussion of Forni (2004) suggest results from stationarity test could deceive the market analyst under two conditions. Firstly when the two series respond to a single common shock and secondly when the original price series are both stationary. The former suggests that the analyst be aware of the impact of a small number of large shocks one indication of which is non-normality. While, the latter problem does not arise in relation to the data used here, but should be detected by testing the univariate series for stationarity.

In addition to analysing the problem via the application of univariate stationarity tests the problem as compared with Forni (2004) is further analysed in a panel context. This is especially pertinent here as it might be imagined that the long-run characteristic of regional gasoline price series should have features in common. It would seem to make sense to compare the univariate analysis with a panel study. One feature that relates to price series is volatility and when the variance process is highly persistent then this may impact the extent to which the computed statistics obtain their asymptotic limit. Boswijk (2001) suggests that a maximum likelihood or least squares correction may be applied to the Dickey Fuller test when the mean and variance processes are independent. While Rahbek et al (2002) found that convergence to the asymptotic distribution might be relatively slow when the variance process has powerful autoregression. The same may not be the case were some extreme distributions selected to explain the data such as the Cauchy or stable Paretian distributions.

3- DATA AND METHODOLOGY

In this paper the stationarity properties of the US gasoline market is analysed using weekly oil price proportions across eight regions of the US: West Coast, Central Atlantic, East Coast, Gulf Coast, Lower Atlantic, Midwest, New England, Rocky Mountains. The sample starts in the first week of May 1993 to the first week of May 2010.[2]

[2] The number of observations is 900. The data have been obtained from energy information administration website (www.eia.doe.gov). The oil price series are all stationary in first difference of their natural logarithm or might be considered to be I(1) in the log price, results available from the authors on request.

To this end we test for the stationarity of the log price differential to determine the definition of the market to consider whether the market can be viewed as being efficient. The approach adopted in this research uses the ADF tests due to Dickey and Fuller (1979), the ADFGLS test of Elliot, Rothenberg, and Stock (1996) and where appropriate a GLS corrected ADF test to account for GARCH/ARCH (Beirne et al, 2007). The panel methods of Hadri (2000), and Im, Pesaran and Shin (2003) are used to support the univariate analysis.

4- RELATION BETWEEN METHODOLOGY AND LITERATURE

To measure market definition in the US gasoline market stationarity tests are applied to see whether there is a long-run dependence between prices of different market segments. Here univariate tests are used to examine whether a combination of log price differentials are stationary and the acceptance of the stationary behaviour of the series confirms consistency with the appropriate market definition.

Finding inefficiencies in the long-run suggests that they relate to market failure rather than mistakes. Whereas arbitrage implies prices move to clear markets and that the market is efficient. Hence, firms prices should move in line or one firm's price should react to the others. If one considers the time series model associated with the ADF test as a reduced form equation of relative price behaviour, then the residuals combine the demand and supply shocks associated with regional and national price movements once market adjustments are captured by the dynamics in the model that underlines the ADF test. Given the underlying non-stationarity of each price, then a stochastic trend explains the evolution of the shocks to each price series and the notion that the trend is common is encapsulated in the error correction term associated with the log price differential being stationary. The underlying econometric hypothesis in the long-run is theoretically consistent to what has been termed parallel pricing (Buccirossi, 2006). Forni (2004) suggests when stationarity is accepted for a number of sub-markets across a country then the market might be viewed as being broad.

The early literature analysing prices considered the observation that they were correlated as a sign of collusive behaviour. While the literature that considers elasticity and consumer loss often requires explicit modelling of demand and supply or cost (see Forni, 2004). One might formulate a dynamic demand and supply systems when information is also available on quantities to compute price elasticities or some notion of consumer surplus, but with uncertainty over quality or price, then a more appropriate measure of loss is termed "Consumer Detriment". Unfortunately, measures of detriment depend on some computation of average cost or the mark-up (Hunter et al, 2001). This leads Forni (2004) to suggest that the observation that finding arbitrage is indicative of a broad market provides an effective means to detect market failure and inappropriate pricing.

[Figure 2 goes here]

STATIONARITY

The log of weekly spot prices for gasoline in West Coast, Central Atlantic, East Coast, Gulf Coast, Lower Atlantic, Midwest, New England, Rocky Mountains are denoted by p_{WC}, p_{CA}, p_{EC}, p_{GC}, p_{LA}, p_{MW}, p_{NE}, and p_{RM}. Figure 2 shows 28 combinations of the differentials in log[2] prices in the US between all the regions prices.

As is explained below tests of stationarity applied to linear combinations of I(1) series relate to restricted cointegrating relations. Hence, the approach is efficient in terms of testing for a combination consistent with long-run efficiency as it fixes the intercept to zero and the slope to unity.

Augmented Dickey Fuller (ADF) test:

The Augmented Dickey Fuller (ADF) test (Dickey and Fuller, 1979) uses a parametric time series regression to eliminate serial correlation. The test is applied to a p^{th} order autoregressive model expressed in differences and levels:

$$\Delta y_t = \pi_0 + \psi y_{t-1} + \sum_{k=1}^{l-1} \pi_k \Delta y_{t-k} + \varepsilon_t . \qquad (1)$$

Testing whether y is a stationary variable considers the proposition that $\psi < 0$. The conventional regression assumptions are required for the model to be well formulated (Davidson and MacKinnon, 2004) and this is important for the correct formulation of the ADF test, under the null of non-stationarity (H_0: $\psi=0$). Acceptance of the null implies that (1) is a p-1th order autoregressive model that relates to series that are stationary in first difference form. Accepting the one sided alternative hypothesis H_1: $\psi<0$ implies y_t is stationary.

To confirm that what we are really considering a form of cointegration it is useful to transform the model underlying the ADF test into a dynamic model in the differential in log prices:

$$\Delta(p_{it}-p_{jt})=\pi_0+\psi(p_{it-1}-p_{jt-1})+\sum_{k=1}^{l-1}\pi_k\Delta(p_{it-k}-p_{jt-k})+\varepsilon_t \qquad (2)$$

Alternatively, (2) might be better seen as an error correction model of Δp_{it}:

$$\Delta p_{it}=\pi_0+\psi(p_{it-1}-p_{jt-1})+\sum_{k=1}^{l-1}\pi_k\Delta p_{it-k}+\sum_{k=1}^{l-1}\delta_k\Delta p_{jt-k}+\varepsilon_t.$$

Any application of a stationarity test to price proportions is also a test of whether the prices cointegrate with a long-run coefficient of unity.

Dickey-Fuller GLS (DFGLS) test

Elliot, Rothenberg, and Stock (1996) show that the ADF regression including a constant, or a constant and a linear time trend can be corrected for the effect of these variables via a preliminary regression. This gives rise to the adjusted model:

$$\Delta y_t^d=\psi y_{t-1}^d+\sum_{k=1}^{l-1}\pi_k\Delta y_{t-k}^d+\varepsilon_t. \qquad (3)$$

Comparing this equation with (2), y_t has been replaced by a GLS corrected variable (y_t^d). The DF-GLS test is more efficient as it eliminates nuisance parameters plus it takes account of initial values that affect the test (Beirne et al, 2007).

Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

Forni (2004) also applied the KPSS test in the analysis of regional data. The null associated with stationarity might be seen to correspond with natural justice so that the test is computed under the proposition that the gasoline market in the US is efficient.

The model can be expressed in terms of a single equation for $t=1\ldots T$:

$$y_t = r_t + \varepsilon_t. \qquad (4)$$

The underlying model breaks down into a random walk and an error component:

$$r_t = r_{t-1} + u_t$$

where r_0 is unknown, the u_t are iid residuals across time. In general, $\sigma_u^2 \geq 0$ and the test is a variance ratio test $\lambda = \sigma_u^2 / \sigma_\varepsilon^2$ with H_0: $\lambda=0$ against H_1: $\lambda>0$.

Under the null $\sigma_u^2 = 0$ and r_t is constant and the Lagrange Multiplier (LM) statistic is based on the residuals from the following regression which we find convenient to specify for the i^{th} price proportion y_{it} related to a common exogenous variable x_t:

$$y_{it} = b\, x_t + \varepsilon_{it}.$$

The LM statistic derived under the assumption that the errors are iid is computed for the i^{th} series using the following relation:

$$LM_i = \frac{1/T^2 \sum_{t=1}^{T} S_{it}^2}{\sigma_i^{*2}}$$

$$S_{it} = \sum_{j=1}^{t} e_{ij},\ \sigma_i^{*2}(z) = \gamma_o + 2\sum_{s=1}^{l} \kappa(l,z)\gamma_s \text{ and } \gamma_s = \frac{1}{T}\sum_{t=s+1}^{T} e_{it}e_{it-s}.$$

Where, γ_0 is the constant variance term, z=$s/l+1$ is the bandwidth, l is the lag truncation and γ_s the autocorrelations that are weighted in the variance equation by the kernel $\kappa(l,z)$. Here the Bartlett, Parzen and Quadratic Spectral functions are applied with different bandwidths.

ANALYSIS OF THE TESTS UNDER THE ALTERNATIVE AND THE NULL OF STATIONARITY

It can be seen from table 1, that the null hypothesis of the unit-root in ADF and DF-GLS tests on virtually all the log differential prices has been rejected at the 5% significance level and that suggest that the corresponding combinations are stationary. This is in contrast to the results of Forni (2004) where the alternative was often rejected at the 5% level. Rejection of the null indicates that the US defines a broad geographic market for gasoline. These results suggest that the market is efficient as out of 28 tests it can be observed from table 1 that only in one case the Gulf Coast and the Lower Atlantic is stationarity rejected at the 5% level.

[Table 1 goes here]

It is an irony that rejection of stationarity occurs when the test is applied under that null as the KPSS test is significant in a number of cases so stationarity cannot be readily accepted. The null of stationarity can only be accepted at the 5% level in the case of Central Atlantic and Mid West, Central Atlantic and West Coast, East Coast and Mid West, Gulf Coast and West Coast, Mid West and West Coast, New England and the Rockies. The West Coast defines the broadest market being integrated with the Central Atlantic, Gulf Coast and Mid West and at the 1% level New England.

Finding a distinct or narrower market based on the analysis derived from the KPSS test in such regions contradicts the findings obtained from ADF and DF-GLS tests. There is some correspondence in these results with those of Forni (2004), except when the impact of the trend is considered.[3]

It has been suggested that the KPSS test performs poorly when the series are stationary, but the autoregressive coefficient is relatively large. While, Caner and Killian (2001) suggested that the KPSS test is not powerful especially in the light of possible moving average behaviour when compared with the ADF and DF-GLS tests. However, Forni (2004) found far less uniform results, though with many fewer observations observed at a lower frequency. Unlike Forni (2004) these results are consistent with the finding of a broad market when the ADF test can be relied on, but the results based on the KPSS test need to be viewed with some circumspection.

PANEL TESTS OF THE NON-STATIONARY/STATIONARY NULL AND COHERENT UNIVARIATE TESTS

In this section the problem is limited to a panel problem and a similar set of univariate tests. In the context of a system of error correction models then the analysis considers *N* price equations as arises when considering a vector autoregressive (VAR) model or a sequence of single equation error correction models. Forni (2004) provided good reasons as to why univariate methods were appropriate for testing long-run arbitrage. Neither the VAR nor the Engle and Granger (1987) procedure simultaneously tests for cointegration and the

[3] These findings may relate to cases where the notion of LEPT due to Hunter and Burke (2007) may better explain the stochastic trend than one series alone. This problem may arise when new information becomes fragmented across a number of price series; especially when related to less influential regions.

restrictions on the coefficients. Hence, testing parallel pricing and stationarity/cointegration is unified and this approach can be seen as being efficient.

[Figure 3 goes here]

A further issue that arises in the context of testing stationarity of real exchange rates relates to cross rates and triangular arbitrage (Smith and Hunter, 1985). This follows from selecting alternative base currencies to determine whether the real exchange rate is stationary. In the exchange rate case the sequence of results that arise from the underlying dynamic exchange rate models are dependent on each other when the different models are coherent (Hunter and Smith, 1982). A similar issue arises when testing stationarity of price proportions using equivalent dynamic models or might be viewed as choosing different regional prices as numeraire. Hence, *½N(N-1)* price proportions relate back to *N* underlying price equations. To this end it has been decided to consider a smaller sub-set of prices, and base this on further information. The remaining analysis is undertaken paying attention to the US regional gasoline infrastructure represented in Figure 3 and as a result the following pairs of log price differentials are selected for further analysis: $\mathbf{y_1}=\log(P_{New\ England}/P_{Mid\text{-}West})$, $y_2=\log(P_{Mid\text{-}West}/P_{Central\ Atlantic})$, $y_3=\log(P_{Mid\text{-}West}/P_{East\text{-}Coast})$, $y_4=\log(P_{Lower\ Atlantic}/P_{Gulf\ Coast})$, $y_5=\log(P_{Mid\text{-}West}/P_{Gulf\ Coast})$, $y_6=\log(P_{Gulf\ Coast}/P_{Rocky\ Mountain})$, $y_7=\log(P_{Gulf\ Coast}/P_{West\ Coast})$, $y_8=\log(P_{Mid\text{-}West}/P_{Rocky\ Mountain})$, $y_9=\log(P_{Rocky\ Mountain}/P_{West\ Coast})$.

The sequence of plots in Figure 4 considers all the selected price proportions. It would appear from visual inspection that y_1, y_2, y_3 y_6 and y_9 are mean reverting and seemingly stationary, but y_4 and y_5 seem to exhibit upward and slight downward trends respectively and as a result appear non-stationary. While y_7 and y_8 do not have any apparent trend, but the behaviour may again appear closer to a random walk without drift and thus potentially non-stationary.

[Figure 4 goes here]

To further evaluate the argument in the above section the ADF, DF-GLS and KPSS test are applied to the log price differentials set out above and the results revealed in table 2.[4] According to these results it can be confirmed for the ADF and DF-GLS tests the null hypothesis of a unit root has been rejected in all cases apart from y_4. The analysis is similar to that considered before. However, the results for the KPSS tests in table 2 imply that the null is accepted at the 1% level for y_2, y_3, y_7 and y_8. However, it is still rejected at the 5% level in cases: y_1, y_4, y_5, y_6, y_7 and y_9.

A further extension to the Forni method especially when the time dimension is limited might be to consider the panel tests due to Im et al (2003) and Hadri (2000). One reason for this might be to investigate the validity of the KPSS test and to this end the small sample corrected panel test due to Hadri (2000) is applied.

[Table 2 goes here]

It has been suggested that the panel case may yield improvements, because there may be benefits to pooling the data in the case where the series are strongly related. A panel approach has also been recently applied by Giulietti et al (2010) to analyse electricity prices for the UK.

The Hadri Panel Unit Root Test

The model underlying the panel version of the test under the null of stationarity can be expressed in terms of a single equation from a set of panel relations for $t=1...T$ and $i=1...N$ the dimension related to each region with

$$y_i = X_i B_i + e_i.$$

Where, $y_i' = [y_{i1}...y_{iT}]$, $e_i' = [e_{i1}...e_{iT}]$ and X_i is a $Tx1$ unit vector or in the trend case includes both the unit vector and a trend. The LM test statistic is the panel version of the KPSS test considered above:

$$LM_u = \frac{1}{N}\sum_{i=1}^{N} LM_i.$$

[4] There are $½N(N\text{-}1)=28$ price proportions to analyse when the upper triangle of price correspondences are excluded; this is a further difference between our analysis and that of Forni. The combinations are further reduced to take account of lines of communication across the US. A further reason for this is that when comparison is made between these types of test and more conventional error correction model and cointegration studies, the 28 inter-relations cannot be independent as they all derive from nine underlying dynamic price equations.

It is shown in Hadri (2000) for the panel version of the LM test that the following finite sample correction follows a standard normal distribution in the limit:

$$Z_u = \frac{\sqrt{N}(LM_u - \xi_u)}{\zeta_u}.$$

Hadri (2000) computes the Bartlett correction terms ξ_u=1/6 and ζ_u^2 = 1/45. For T>50 Hadri (2000) provides evidence based on Monte Carlo simulation that shows that the empirical size of the test is approximately .054 and with λ in the range [$.1,\infty$] the test has maximum power. It is robust to non-normality and performance is not reliant on the size of the cross section. Results for the Hadri panel unit root test based on different bandwidth and kernel are presented in table 3. The test is compared with a one sided critical value at the 5% level of 1.64, but only with the Parzen kernel for the longest bandwidth do we find that for 3 out of the eight cases considered can the null of stationarity be accepted. Hence, the Hadri results do not contradict those of the KPSS test.

[Table 3 goes here]

The size of a typical autocorrelation coefficient related to the ADF test statistic suggests the data are strongly autoregressive. The strength of this correlation explains why their behaviour might be seen to be distinct from the model underlying the KPSS test that considers a constant plus random component drive the series under the null as compared with the distance between the stationary and unit root case that arises with the ADF test. So the poor performance of the Hadri and KPSS tests may be symptomatic of strong or persistent autoregression and this may not be inconsistent with the suggestion by Caner and Killian (2001) that the KPSS test is particularly problematic when the error process is moving average (MA). One interpretation of the analysis of the lags might be that the long autoregression really relates to MA or ARMA behaviour.

The Im, Pesaran and Shin Panel Unit Root Test

For comparison with the panel unit root test under the null of stationarity the test due to Im et al (2003) is applied. The IPS panel unit root test estimates a mean adjusted ADF test statistic that may also be corrected for fixed and random effects. Individual ADF tests are computed for within mean adjusted data and the t-tests computed accordingly. The t-bar test statistic is the average of the individual mean corrected ADF statistics compared with the critical values derived in the tables in Im et al (2003).The result of the IPS test are presented in table 4.

[Table 4 goes here]

In all cases for the panel the alternative of stationarity can be accepted at the 1% level and so on average the series are stationary (Beirne et al, 2007). Hence, the panel and univariate tests are consistent in terms of the stationary and non-stationary null.

5- SUMMARY AND CONCLUSION

Forni (2004) suggests that a broad market might be determined by a confirmatory analysis that follows from univariate stationarity tests based on the null and alternative of non-stationarity producing a common conclusion. In this article, such confirmatory analysis only finds in one case that the market is narrow and that is for the potential relation between the Gulf Coast and the Lower Atlantic.

In terms of finding a broad market based on the Forni approach that is here corrected so the lower triangle of results are defined by symmetry, then the ADF and KPSS results at the 1% level conclude that the West Coast defines the broadest market in association with the Central Atlantic, Gulf Coast, Mid West and New England and via symmetry the reverse applies. The Mid West defines a broad market with the Central Atlantic, East Coast and West Coast; while the Central Atlantic defines an extended market with the Mid West and the West Coast, and New England with the Rockies and West Coast. Furthermore prices all to respond to each other in the following cases: the East Coast and Mid West; the Gulf Coast and West Coast; and the Rockies and New England.

Here the analysis was extended to include panel tests and following a similar analysis of exchange rates by Beirne et al (2007) we conclude that when the majority of univariate unit root tests and panel stationarity tests accept the alternative of stationarity, then the conclusion to be drawn is that on average the series are stationary. This arises when the ADF tests are considered as eight out of the nine univariate tests accept the stationary alternative as does the IPS test for a panel of eight price proportions. So on average when the ADF tests are to be accepted, there is a broad market as in the long-run these price ratios are stationary; the tests based on autoregressive models appear able to distinguish between the highly persistent behaviour related to a random walk as compared with powerful autoregression.

If a similar approach is considered for the tests under the null of stationarity this corresponds in the case of regulation and competition to natural justice. However, the KPSS test results suggest that on average the price series are non-stationary as the null is rejected in five out of nine cases. While the panel test of Hadri with $N>1$ provides very little support that the market is broad. Notice that individual price series are characterized by ARCH style volatility, strong serial correlation and non-normality and as a result the tests under the null of stationarity may not be well disposed to distinguish between strong autoregressive behaviour and a unit root as the process under the null is subject to the intercept just white noise. In relation to the panel form of the test, Hlouskova and Wagner (2006) provide evidence that it may over-reject the null hypothesis when there is powerful serial correlation, but their results relate to a small sample across time. Hadri (2000) suggests that this test is robust to non-normality and the correspondence between the value of the KPSS and Bartlett corrected test would suggest that normality might not be an issue for 900 observations. However, neither the Hadri nor the KPSS test is corrected for ARCH and strong ARCH has been shown to affect the rate of convergence of the Johansen trace test based on the VAR even with a sample of 1000 observations and powerful volatility described by the spectral radius of the ARCH polynomial (Rahbek et al, 2002).

If there is significant doubt in relation to the stationarity test or where there is further evidence of localized market power and collusion, then this would suggest that further concentration of ownership of gas stations or refining capacity should be rejected in regions where the stationary null in terms of price proportions is called into question.

REFERENCES

Beirne J., J. Hunter and M. Simpson (2007) Is the Real Exchange Rate Stationary? The Application of Similar Tests for a Unit Root in the Univariate and Panel Cases, Qualitative and Quantitative Analysis in the Social Sciences, Vol. 1, No.2, pp 55-70.

Boswijk, HP. (2001) Testing for a Unit Root in the presence of near integrated Volatility, Tinbergen Institute Discussion Paper, TI2001-077/4.

Buccirossi, P. (2006) Does Parallel Behavior Provide Some Evidence of Collusion?, Review of Law & Economics, Vol. 2, No 3, pp. 85-101.

Burke S.P. and Hunter J. (2005) Modelling Non-stationary Economic Time Series: A Multivariate Approach. Palgrave, Basingstoke.

Burke S.P. and Hunter J. (2011) Long-run Equilibrium Price Targeting. Quantitative and Qualitative Analysis in the Social Sciences, 5 (1), 26-36.

Dickey D.A. and W.A. Fuller (1979) Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, Vol. 74, pp. 427-431.

Elliott G. T.J. Rothenberg and Stock J.H. (1996) Efficient tests for an autoregressive unit root, Econometrica, Vol. 64, pp. 813-36.

Engle R.F. and Granger C.W.J. (1987) Co-integration and error-correction: representation, estimation and testing, Econometrica, Vol. 55, pp. 251-276.

Forni M. (2004) Using stationarity tests in antitrust market definition, American Law and Economics Review, Vol. 6, pp. 441-464.

Giulietti M. J. Otero and M.Waterson (2010) Pricing behaviour under competition in the UK electricity supply industry, Oxford Economic Papers. Vol. 62, pp. 478-503.

Hendry, D.F. and Mizon, G.E. (1978) Serial correlation as a convenient simplification not a nuisance: a comment on a study of the demand for money by the Bank of England. Economic Journal, 88 549-63.

Hosken D. and C.Y. Taylor (2004) Discussion of Using stationarity test in antitrust Market Definition, American Law and Economics Review, Vol. 6, No. 2, pp. 465-475.

Horowitz I. (1981) Market definition in antitrust analysis: A regression based approach, Southern Economic Journal, Vol. 48. pp. 1-16.

Hunter, J. and S.P Burke (2007) Common trends, Cointegration and Competitive Price Behaviour, paper presented at the RES conference, Warwick, 2007 and the Econometrics Study Group Conference, Bristol, June, 2007.

Hunter J., C. Ioannidis, E. Iossa and Skerratt, L. (2001) Measuring, Consumer Detriment Under Conditions of Imperfect Information, Office of Fair Trading, Economic Research Paper 21.

Hunter J and M. Simpson (2004) The specification of cross exchange rate equations used to test Purchasing Power Parity, presented at the EEFS conference in Coimbra, Portugal and Brunel Discussion Paper, 04-22.

Hunter J. and R.P. Smith (1982) Cross Arbitrage and the Specification of Exchange Rate Models, paper presented at the NBER Summer School, July 1982.

Im K.S., M.H. Pesaran and Y. Shin (2003) Testing for unit root in heterogeneous panels, Journal of Econometrics, Vol. 115, pp. 53-74.

Koutsyiannis A. (1994) Modern Microeconomics, MacMillan, London.

Kurita T. (2008) Common stochastic trends and long-run price leadership in the US gasoline market, Working paper WP-2008-003, Faculty of Economics, Fukuoka University.

LecG Ltd (1999) Quantitative techniques in competition analysis, Research paper 17, OFT 266

Lexecon (2001) Market Definition: How stationarity tests can improve accuracy, www.lexcon.co.uk.

London Economics (2002) Study of Sectoral Price Inter-dependence, paper prepared for the Nederland's Mededingingsautoriteit.

Markham J. (1951) The Nature and Significance of Price Leadership, American Economic Review, Vol. 41, pp. 891-905.

Maunder P. (1972) Price Leadership an Appraisal of its Quality in some British Industries, The Business Economist, Vol. 4, pp. 132-140.

Posner R.A. (1976) Antitrust law: an economic perspective, University of Chicago press, Chicago.

Slade M.E. (1986) Exogeneity tests of market boundaries applied to petroleum products, Journal of Industrial Economics, Vol. 34, pp. 291-303.

Smith R.P. and J. Hunter (1985) Arbitrage and Specification in Exchange Rate Models, Economics Letters, Vol. 18, pp. 375-376.

Stigler G.J. and Sherwin R.A. (1985) The Extent of a Market, Journal of Law and Economics, Vol. 28, pp. 555–585.

Uri N., D.J., Howell and J.R. Rifkin (1985) On defining geographic markets, Applied Economics, Vol. 17, pp. 959-977.

Figure 1- Gasoline Taxes across the US[5]

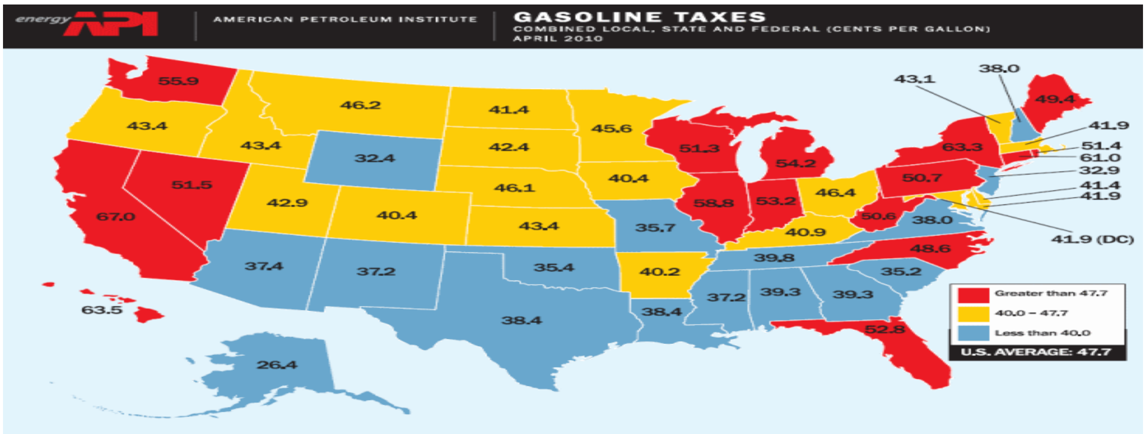
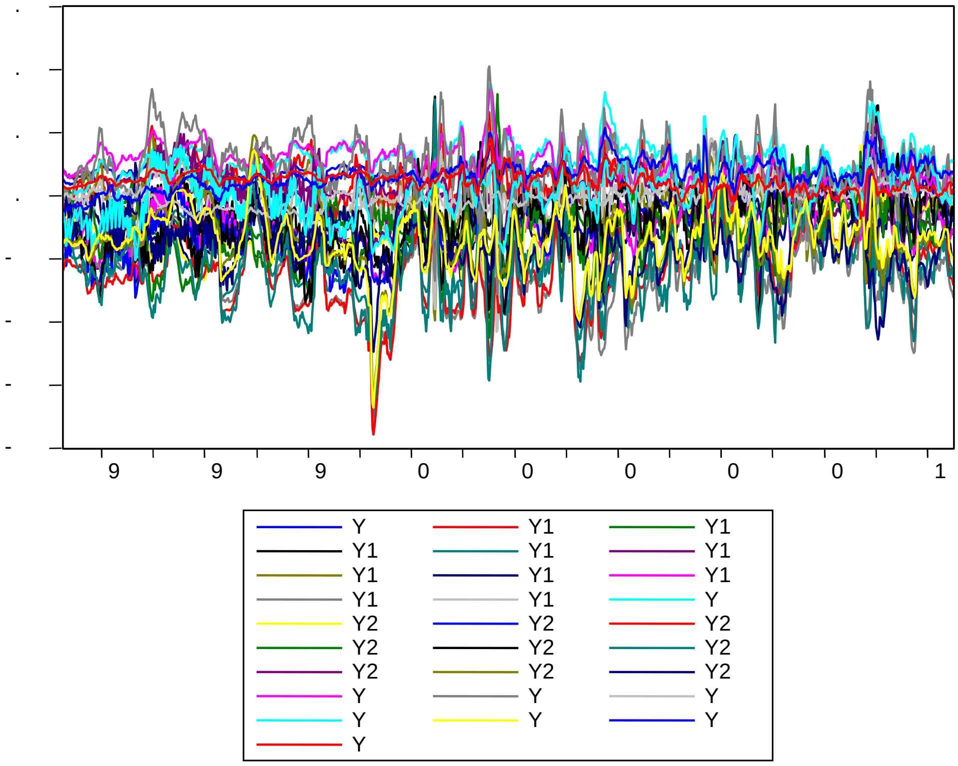


Figure 2- Plot of log price differential in CA, EC, GC, LA, MW, NE, RM, WC



Note: Representing plots of log differential in prices of gasoline in CA, EC, GC, LA, MW, NE, RM, WC; $\mathbf{Y_1}$= log (P_{CA})- log (P_{EC}), $\mathbf{Y_2}$= log(P_{CA})- log (P_{GC}), $\mathbf{Y_3}$= log (P_{CA})- log (P_{LA}), Y_4=log (P_{CA})- log (P_{MW}), Y_5= log (P_{CA})- log (P_{NE}), Y_6= log(P_{CA})- log (P_{RM}), Y_7= log (P_{CA})- log (P_{WC}), Y_8= log (P_{EC})- log (P_{GC}), Y_9= log (P_{EC})- log (P_{LA}), Y_{10}= log (P_{EC})- log (P_{MW}), Y_{11}= log(P_{EC})- log (P_{NE}), Y_{12}= log (P_{EC})- log (P_{RM}) , Y_{13}= log(P_{EC})- log(P_{WC}), Y_{14}= log(P_{GC})- log (P_{LA}), Y_{15}= log (P_{GC}) log- log(P_{MW}), Y_{16}= log(P_{GC})- log (P_{NE}), Y_{17}= log (P_{GC})- log (P_{RM}) , Y_{18}= log (P_{GC})- log(P_{WC}), Y_{19}= log (P_{LA})- log (P_{MW}), Y_{20}= log(P_{LA})- log (P_{NE}), Y_{21}= log (P_{LA})- log (P_{RM}) , Y_{22}= log (P_{LA})- log (P_{WC}), Y_{23}= log (P_{MW})- log (P_{NE}), Y_{24}= log (P_{MW})- log (P_{RM}) , Y_{25}= log (P_{MW})- log (P_{WC}) , Y_{26}= log (P_{NE})- log (P_{RM}), Y_{27}= log (P_{NE})- log (P_{WC}) , Y_{28}= log (P_{RM})- log (P_{WC}).

[5] The above diagram was obtained from www-static.shell.com.

Figure 3- Map of US regional gasoline infrastructure[6]

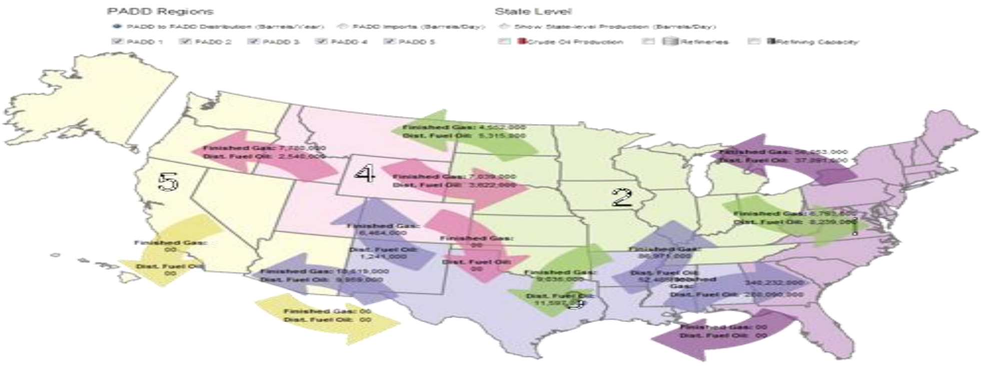
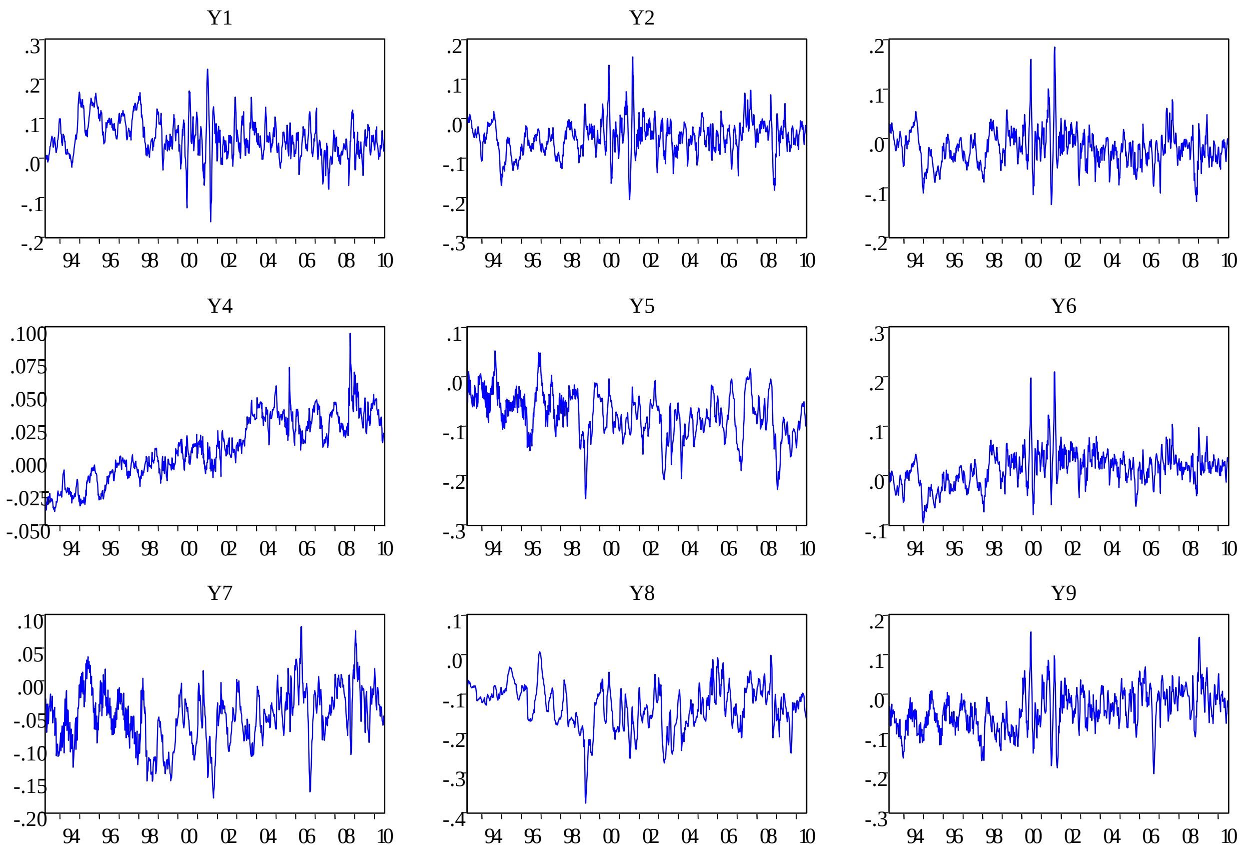


Figure 4- Plot of gasoline log price differential in different regions of US based on the regional gasoline infrastructure



[6] The above diagram was obtained with permission of the National Association of Convenience Stores, from the 2012 NACS Retail Fuels Report.

Table 1- Summary of ADF tests, DF-GLS tests and KPSS tests on the log differential of gasoline prices - with intercept and no trend.

	P_{LCA}	P_{LEC}	P_{LGC}	P_{LLA}	P_{LMW}	P_{LNE}	P_{LRM}	P_{LWC}
P_{LCA}		-3.724537** (-3.706698**) [3.452390**]	-6.171970** (-1.990969*) [0.915151**]	-4.989478** (-4.707481**) [4.902073**]	-4.928462** (-2.335932*) [0.446911]	3.829677** (-3.747953**) [5.438980**]	-4.541653** (-3.167151**) [1.165679**]	-5.441773** (-5.411565**) [0.267441]
P_{LEC}			-4.080439** (-1.077737) [4.317563**]	-4.648051** (-3.939956**) [2.451975**]	-4.721538** (-2.055378*) [0.182702]	-3.700019** (-3.640446**) [7.843325**]	-4.011872** (-3.121466**) [2.022444**]	-4.808598** (-4.800208**) [0.749265**]
P_{LGC}				-2.222776 (-0.303123) [3.431025**]	-3.360986* (-2.815583**) [1.681035**]	-5.177154** (-1.868563) [0.861199**]	-6.013833** (-5.089623**) [4.908352**]	-3.784865** (-2.275466*) [0.378705]
P_{LLA}					-3.870944** (-2.352104*) [0.962882**]	-3.838434** (-3.636543**) [4.868950**]	-3.917549** (-3.512071**) [2.318920**]	-4.068736** (-4.051737**) [0.921917**]
P_{LMW}						-3.805266** (-2.132938*) [1.493587**]	-4.430335** (-4.251056**) [2.443409**]	-4.741388** (-3.287946**) [0.253900]
P_{LNE}							-5.436560** (-3.795707**) [0.365536]	-5.108161** (-5.030299**) [0.542706*]
P_{LRM}								-5.814599** (-4.529151**) [1.036716**]
P_{LWC}								

Note: Values without the brackets presents ADF/OLS t-statistic, values in ( ) shows DF-GLS/OLS t-statistic, and values in [ ] indicates KPSS LM-statistic. ADF test critical value at 1% level is -3.437483 and at 5% level is -2.864578. DF-GLS test critical values at 1% level is -2.567566, at 5% level is -1.941180. The KPSS test critical value at 1% level is 0.739, at 5% is 0.463. ** Significant at the 99% confidence level, and* at the 95% confidence level.

Table 2- Summary of ADF tests, DF-GLS tests and KPSS tests on the log differential of gasoline prices (with intercept and no trend)

Log price differential	ADF/ OLS t-statistic	DF-GLS/ OLS t-statistic	KPSS LM-statistic
$X_1(25)$ [7]	-3.805266 *	-2.526560 *	1.038708**
$X_2(25)$	-4.928462 *	-2.335932 *	0.413221
$X_3(25)$	-4.721538 *	-2.055378 *	0.166735
$X_4(23)$	-2.222776	-0.303123	2.866131**
$X_5(20)$	-5.814599 *	-4.52915*	0.887255**
$X_6(16)$	-3.360986*	-2.81558*	1.226137**
$X_7(20)$	-5.213816*	-3.99106*	0.635492*
$X_8(24)$	-3.784865**	-2.275466*	0.323487
$X_9(16)$	-4.430335*	-4.251056*	2.443409**

Note: ADF Test Critical value at 1% is -3.44, at 5% is -2.86. DF-GLS test Critical value at 1% is -2.57, at 5% is -1.94. KPSS test Critical value at 1% is 0.74, at 5% is 0.46. * Significant at the 95% confidence level and ** significant at the 99% confidence level.

[7] Using the same process as was presented in the previous section the lag orders have been selected via inspection of the correlogram. For the nine price differentials $y_i(j)$, then (j) represents the lag order.

Table 3- panel unit root test based on different bandwidth and kernel with individual intercept

	Kernel	*Z-Statistic with max lag bandwidth I*	*P-value*	*Z-Statistic with3xmax lag bandwidth*	*P-value*	*Z-Statistic with long bandwidth*	*P-value*
$X_1(25)$ $X_2(25)$	**Bartlett**	7.01801	0.0000	3.39543	0.0003	3.16228	0.0008
	Parzen	7.97431	0.0000	4.28954	0.0000	1.67134	0.0473
	Quadratic spectral	6.15763	0.0000	2.77395	0.0028	4.54125	0.0000
$X_9(16)$ $X_7(20)$	**Bartlett**	12.3978	0.0000	6.36771	0.0000	3.16228	0.0008
	Parzen	14.3860	0.0000	7.83900	0.0000	**1.62509**	0.0521
	Quadratic spectral	10.4138	0.0000	5.17141	0.0000	4.34528	0.0000
$X_9(16)$ $X_6(16)$	**Bartlett**	19.0721	0.0000	9.56826	0.0000	3.16228	0.0008
	Parzen	21.7708	0.0000	11.3763	0.0000	1.68960	0.0456
	Quadratic spectral	16.2368	0.0000	7.84960	0.0000	4.19906	0.0000
$X_4(23)$ $X_8(24)$	**Bartlett**	7.99594	0.0000	3.51012	0.0002	3.16228	0.0008
	Parzen	9.86467	0.0000	4.18272	0.0000	**1.59771**	0.0551
	Quadratic spectral	6.55805	0.0000	2.78203	0.0027	5.00252	0.0000
$X_5(20)$ $X_8(24)$	**Bartlett**	4.18928	0.0000	2.09300	0.0182	3.16228	0.0008
	Parzen	4.93103	0.0000	2.41350	0.0079	**1.63134**	0.0514
	Quadratic spectral	3.39104	0.0003	1.66684	0.0478	5.66465	0.0000
$X_3(25)$ $X_6(16)$	**Bartlett**	6.14625	0.0000	2.53044	0.0057	3.16228	0.0008
	Parzen	7.44649	0.0000	3.28497	0.0005	1.74157	0.0408
	Quadratic spectral	5.22775	0.0000	1.94076	0.0261	5.35815	0.0000
$X_1(25)$ $X_2(25)$ $X_3(25)$ $X_4(23)$ $X_5(20)$ $X_6(16)$ $X_7(20)$ $X_8(24)$ $X_9(16)$	**Bartlett**	17.4965	0.0000	8.36028	0.0000	6.70820	0.0000
	Parzen	20.3252	0.0000	10.3208	0.0000	3.53917	0.0002
	Quadratic spectral	14.8680	0.0000	6.78183	0.0000	9.75988	0.0000

Table 4- Im, Pesaran and Shin Unit Root Test with individual intercept

Log differential price	*IPS/ OLS t-statistic*	*P-Value*
$X_1(25)$ and $X_2(25)$	-4.55156	0.0000
$X_9(16)$ and $X_7(20)$	-5.62423	0.0000
$X_9(16)$ and $X_6(16)$	-3.81476	0.0001
$X_4(23)$ and $X_8(24)$	-2.34972	0.0094
$X_5(20)$ and $X_8(24)$	-4.58201	0.0000
$X_3(25)$ and $X_6(16)$	-4.01161	0.0000
$X_1(25)$, $X_2(25)$, $X_3(25)$, $X_4(23)$, $X_5(20)$, $X_6(16)$, $X_7(20)$, $X_8(24)$, $X_9(16)$	-8.04934	0.0000

The selected lag number for each pair is equal to the maximum lag order within the series.