Nonlinearities in the US corn-ethanol-oil price system

Teresa Serra, David Zilberman, José M. Gil, and Barry K. Goodwin

Abstract

We use a smooth transition vector error correction model to assess price relationships within the US ethanol industry. Daily ethanol, corn and oil futures prices observed from mid-2005 to mid-2007 are used in the analysis. Results indicate the existence of an equilibrium relationship between ethanol, corn and oil prices. However, only ethanol prices adjust, in a non-linear fashion, to deviations from this long-run parity. Generalized impulse response functions indicate that a shock to both oil and corn prices causes a change in ethanol prices of the same sign. Ethanol responses usually reach a peak after about 10 days of the initial shock and fade away within 35 days.

Key words: Biofuels, United States, Cointegration, Threshold

Topic: Market and Demand Analysis


Copyright 2008 by [authors]. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.
Introduction

Ethanol currently represents the major liquid biofuel produced around the world. Global ethanol production has been growing at an impressive rate in recent years, doubling its amount during the first half of the 2000s. In 2005 ethanol output totaled 12,150 million gallons, with Brazil and the US producing, each, around a 35% of this quantity (Renewable Fuels Association 2006). Although ethanol can be produced from a wide range of feedstocks, sugar cane in Brazil and corn in the US represent the most relevant ones.

Rising interest in ethanol as a renewable source of energy can be attributed to two main causes. First, increasing crude oil prices that have created an incentive to use alternative energy sources. Second, the worldwide introduction of climate change and energy security-related policies supporting the use of ethanol. These changes have brought about an increase in both ethanol and biofuel-related demand for agricultural products. This has raised interest in analyses assessing the impacts of biofuels’ expansion on agricultural markets (see Rajagopal and Zilberman 2007 for a thorough review of the economic studies on biofuels).

Agricultural prices have traditionally been affected by energy prices. However, until recently, such price link was limited to the impacts of fossil fuel prices on production costs. With increasing energy prices and improved bioenergy technologies, higher energy prices are also affecting agricultural prices through increased demand for agricultural produce. Previous research has suggested that ethanol, oil and feedstock prices may be related in a nonlinear fashion. When analyzing the Brazilian ethanol market Schmidhuber (2006) notes that, since the energy market is large relative to feedstock markets, feedstock prices are endogenous to changes in energy prices. He also argues that due to transactions costs, in the short-run feedstock prices may under or overshoot their energy parity price, but they eventually should move back to the long-run equilibrium.
In the same line of research, Balcombe and Rapsomanikis (2007) argue that feedstock, ethanol, and oil prices may be perceived as being non-related if prices move within a threshold, but price transmission mechanisms may be activated outside the threshold. In their empirical analysis they prove the existence of nonlinear price adjustments of sugar and ethanol to oil in Brazil, though linear price adjustments between ethanol and sugar are found. While price adjustments in Brazilian markets have been analyzed, to our knowledge no previous study has attempted to address the relationship between corn, ethanol and oil prices in the US. Our study attempts at assessing this issue.

The bioenergy market in the US

The US has a long history in the production of ethanol as a fuel alternative, it is the first major producer of biofuels, and uses maize as the principal feedstock. The Federal Government promotes ethanol production through tax incentives. In addition to the federal blending credit subsidy, there are other federal and state subsidies. Koplow (2006) estimates the total subsidy for ethanol in 2006 to be in the range between USD 1.42 and USD 1.87 per gallon of gasoline equivalent.

In 1990, through the Clean Air Amendments, the US Congress acknowledged that changes in fuels and their composition can reduce pollution and required gasoline to contain fuel oxygenates that include ethanol. A further incentive to ethanol production in the US came from the US Energy Policy Act of 2005 that mandated the Environmental Protection Agency (EPA) the promulgation of regulations to ensure that gasoline sold in the US contains a minimum volume of renewable fuel. The Renewable Fuel Standard (RFS) Program is expected to increase this volume from 4 billion gallons in 2006 to 7.5 by 2012 (EPA 2007).

In spite of tax exemptions, until recently the US ethanol industry played only a minor role as a supplier of fuel additives. The market for oxygenate additives was
dominated by MTBE (methyl-tertiary-butyl ether) which had been traditionally used as a source of oxygen. However, the toxicity of this compound led many states to ban its use. Since ethanol is the main MTBE competitor, producers switched to ethanol causing the demand for ethanol to expand nationally.

While during most of the history of the US federal ethanol subsidy, crude oil prices have ranged between USD 20 and USD 30 per barrel, recent prices are around USD 60. Ethanol investments have been very profitable at these prices and this has attracted new capital into the industry. The market grew by 1 billion gallons in 2006 and it is expected to grow an additional 3 billion in 2007. This has caused a tremendous impact on corn prices.

As noted, the relationship between energy and feedstock prices within the US has not been assessed. A high degree of integration between the biofuel and the fossil fuel markets should ensure a strong price link between the two products. There are however different constraints that can limit such integration: bottlenecks in distribution, technical problems in transportation and blending systems, etc. Flexibility both at the demand and supply side should also ensure a tight price link between feedstock and energy prices. Previous research suggests that price relationships within the ethanol industry may be of a nonlinear nature (Balcombe and Rapsomanikis 2007; Schmidhuber 2006; Rapsomanikis and Hallam 2006). Our study attempts at characterizing price relationships between corn, ethanol and oil prices allowing for nonlinearities in the process of price adjustments.

**Methodology**

To capture nonlinearities in the process of price adjustment we use threshold models. As noted by previous literature (see Chan and Tong 1986) threshold vector error correction models are based upon the assumption that the transition from one regime to another is abrupt and discontinuous. However, if transactions costs differ among individuals, then
adjustment is likely to be of a smoother nature. Smooth transition autoregressive type of models (STAR) allow capturing this issue.

Of particular interest to our analysis are the smooth transition vector error correction models (STVECM) that allow for long-run relationships among the variables of the model, as well as for nonlinear adjustments towards long-run equilibrium. We model US corn-ethanol-oil price relationships by using a STVECM. To our knowledge STVECM have not been used to assess price relationships within biofuel markets, which represents a contribution to previous literature. Following van Dij et al. (2002) a k-dimensional STVECM can be expressed as

\[
\Delta P_t = \left( \phi_{1,0} + \alpha_1 z_{t-1} + \sum_{j=1}^{p-1} \phi_{1,j} \Delta P_{t-j} \right) \left( 1 - G(s_{t-d}; \gamma, c) \right) + \\
\left( \phi_{2,0} + \alpha_2 z_{t-1} + \sum_{j=1}^{p-1} \phi_{2,j} \Delta P_{t-j} \right) G(s_{t-d}; \gamma, c) + \varepsilon_t
\]

where \( P_t \) is a \((k \times 1)\) vector of I(1) prices, \( \alpha_i, \ i = 1, 2, \) are \((k \times r)\) matrices representing the speed of adjustment to disequilibriums from the long-run relationships, \( z_{t-1} = \beta' P_{t-1} \) is a \((r \times 1)\) matrix of error correction terms, and \( \beta \) \((k \times r)\) contains the parameters of the cointegration relationships. The short-run dynamics are represented by \( \phi_{i,0}, \ i = 1, 2, \) that are \((k \times 1)\) vectors and \( \phi_{i,j}, \ i = 1, 2, \ j = 1, \ldots, p-1 \) being \((k \times k)\) matrices. \( \varepsilon_t \) is a k-dimensional vector white noise process with a mean zero vector and a \((k \times k)\) \( \Sigma \) covariance matrix. \( G(s_{t-d}) \) is the transition function, assumed to be continuous and bounded between zero and one.

The STVECM can be considered as a regime-switching model allowing for two regimes that are associated with the extreme values of the transition function, one
corresponding to \( G(s_{t-d}) = 0 \) and another to \( G(s_{t-d}) = 1 \), where the transition from one regime to another takes place smoothly. The regime occurring at a certain point in time \( t \) is determined by the transition variable \( s_{t-d} \) and the corresponding value of \( G(s_{t-d}) \). In our analysis the transition variable is assumed to be a lagged residual from an error-correction relationship. Following Teräsvirta (1994) we discriminate between logistic and exponential specifications of the transition function and results indicate that an exponential function better represents our data, which yields an exponential smooth transition threshold vector error correction model (ESTVECM):

\[
G(s_t; \gamma, c) = 1 - \exp \left\{ -\frac{\gamma(s_{t-d} - c)^2}{\sigma^2(s_{t-d})} \right\}, \; \gamma > 0
\]

where \( \sigma^2(s_{t-d}) \) is the variance of the transition variable.

Under the exponential specification, the regimes are associated with small and large absolute values of \( s_{t-d} \) and adjustment is symmetric around parameter \( c \). Parameter \( c \) is the threshold between the two regimes. Function \( G(s_{t-d}) \) takes the value of 0 when \( s_{t-d} = c \) and increases monotonically as \( s_{t-d} \) moves away from this threshold. Parameter \( \gamma \) determines the speed of transition from one regime to another. The normalization of the transition variable and the threshold parameter using \( \sigma^2(s_{t-d}) \) allows making the slope parameter scale free.

Our estimation strategy can be summarized as follows. We first specify the STVECM by conducting linearity tests. Before conducting these tests, we use the linear version of the STVECM to select the appropriate lag length. Specifically, we use both the system AIC and SBC information criteria, as well as the multivariate Portmanteau and
univariate F tests for autocorrelation to ensure that there is no remaining autocorrelation. Once linearity is rejected, we estimate the nonlinear model and test for no remaining error autocorrelation. Finally, the ESTVECM dynamic properties are evaluated by using the impulse response analysis.

An important step towards building a STVECM consists of testing linearity against the nonlinear model. A test of linearity in model (1) is a test of the equality of the error correction and autoregressive parameters in the two regimes, i.e., $\alpha_1 = \alpha_2$ and $\phi_{1,j} = \phi_{2,j}$, $j = 0, \ldots, p - 1$. Testing for linearity is complicated by the presence of unidentified nuisance parameters under the null (see van Dij et al. 2002). An approach to solve the problem was suggested by Luukkonen, Saikkonen and Teräsvirta (1988), who proposed replacing the transition function $G(s_{t-d})$ by a suitable Taylor series expansion. A system likelihood ratio statistic or an F test can be then used to test for linearity. Thus, to test for linearity, we estimate the following auxiliary regression, a first-order Taylor series expansion of the STVECM using the exponential as the transition function:

\begin{equation}
\Delta P_t = M_0 + A_0 z_{t-1} + \sum_{j=1}^{p-1} \beta_{a,j} \Delta P_{t-j} + M_1 s_{t-d} + A_1 z_{t-1} s_{t-d} + \sum_{j=1}^{p-1} \beta_{1,j} \Delta P_{t-j} s_{t-d} + M_2 s_{t-d}^2 + A_2 z_{t-1} s_{t-d}^2 + \sum_{j=1}^{p-1} \beta_{2,j} \Delta P_{t-j} s_{t-d}^2 + \eta_t
\end{equation}

where $\eta_t$ includes the original error vector and the errors that derive from the Taylor series approximation. The corresponding linearity test can be denoted by $H_0 : M_1 = M_2 = A_1 = A_2 = B_{1,j} = B_{2,j} = 0$ and the F statistic is used to test for the null.
As noted above, the variable driving regime switching is a lagged residual from a cointegration relationship. Linearity tests are used to determine the optimal error correction term (both Engle and Granger 1987; and Johansen 1988 error correction terms are considered) and its delay parameter $d$ along the lines suggested by Teräsvirta (1994). Specifically, linearity tests are conducted using different alternatives and the error correction term and its lag are chosen to minimize the p-value of the linearity test. Once the transition function and variable are selected, the parameters of the STVECM are estimated by using nonlinear least squares (NLS). Following previous literature (Legrenzi and Milas 2005), we use $\gamma = 1$ as a starting value for $\gamma$ and values of $s_{t-d}$ close to its mean as a starting value for the threshold parameter $c$. We test the estimated model for no residual autocorrelation following Eitrheim and Teräsvirta (1996).

We then examine dynamic price relationships by computing Generalized Impulse-Response Functions (GI). GI are useful instruments to predict the effects of shocks to a specific series or to the whole system. As proposed by Koop, Pesaran and Potter (1996) and Weise (1999) the GI can be expressed as:

\[
GI(h, \eta, \omega_{t-1}) = E[\Delta P_{t+h}|\epsilon_t = \eta, \omega_{t-1}] - E[\Delta P_{t+h}|\omega_{t-1}]
\]

where $h = 0, 1, ..., H$ represents the forecast horizon, $\epsilon_t = \eta$ is a $(k \times 1)$ vector denoting a specific shock to the system, and $\omega_{t-1}$ contains the histories of the series. From the expression above it is thus clear that the impulse response is defined as the difference between two realizations of $\Delta P_{t+h}$ that have identical histories up to period $t-1$ ($\omega_{t-1}$). In one realization $\Delta P_{t+h}$ is shocked while in the other it is not. The impacts of a shock depend not only on the history of the process, but also on the sign and size of the shock. To compute the GI we follow Weise (1999).
Results

Our empirical analysis uses daily futures prices for corn, ethanol and crude oil observed from July 21, 2005 to May 15, 2007. Information on corn and ethanol futures was obtained from the Chicago Board of Trade (CBOT) and crude oil (light-sweet, cushing, Oklahoma) futures prices were obtained from the New York Mercantile Exchange (NYMEX). While contract months for oil and ethanol futures are the twelve consecutive calendar months (from January to December), contract months for corn are December, March, May, July and September. To make the three series comparable, we use corn contract months to define maturity months and nearby contracts when constructing futures price series.

Following previous research (Jin and Frechette 2004; Booth and Tse 1995), when a futures contract comes to its maturity month, the nearby contract is used to compile the data. Specifically, on the fifteenth day of the month previous to the expiration of the contract (day $t$), the settlement price is computed as the average of the settlement price for days $t$ and $t-1$. On day $t+1$, the settlement price corresponds to the average of prices in $t$ and $t+1$.

Figure 1 plots the price series used in the empirical implementation (all series are divided by their respective means). A preliminary analysis of these series was carried out to assess their time series properties. Standard ADF (1979), Perron (1997) and KPSS (1992) statistics, as well as Elliott et al. (1996) and Elliott (1999) tests confirmed the presence of a unit root in all price series. Standard Johansen (1988) and Engle and Granger (1987) tests also suggest the presence of a (single) cointegration relationship between the prices studied.

The estimation strategy described above was followed. As noted, a first step before conducting linearity tests involves selection of the optimal number of lags using the linear version of the STVECM. Different lag lengths were considered and 10 lags were chosen. Linearity tests were conducted and the transition variable and its lag length were optimized.
Two transition variables and different lags for these variables were initially considered: the error correction term derived from Johansen (1988) and the error correction term obtained from the Engle and Granger (1987) method. In both cases linearity tests would support the use of three lags and for such lag length, the p-value of the test was smaller when using Engle and Granger (1987) error correction term and equal to 9.75E-31 which allows to comfortably reject the null of linearity. We therefore estimate the parameters of the STVECM.

A total of 194 parameters were estimated and more than a quarter are statistically significant. The main parameters of interest in STVECM are the speed of adjustment to the long-run parity($\alpha$), the speed of transition from one regime to another ($\gamma$) and the threshold parameter ($c$). They are presented in table 12 along with the Eitrheim and Teräsvirta (1996) tests for autocorrelation. To preserve space, parameters showing the short-run dynamics of the series are not offered but are available from the authors upon request. The estimate for the threshold parameter $c$ is statistically significant and equal to 0.32. The parameter estimate representing the speed of adjustment is statistically significant and equal to 0.45, indicating that the speed of the transition from one regime to another is not very quick. In figure 2 we offer a plot of the speed of the transition process which confirms the slow adjustment.

Figure 3 plots the values of the transition function together with the transition variable and the estimated threshold. From this figure it is evident that the relationship between the prices considered has been highly volatile during the period of analysis in the sense that frequent regime changes have been registered. The declining trend in $G(s_{t-\delta})$ from the mid-2005 to end-2005 indicates a tendency to stabilize the relationship between the three prices. However, the sharp increase in ethanol prices by mid-2006 and the consequent and subsequent increase in corn prices destabilized the market and brought about a marked instability.
Also, this instability is not surprising in light of the fact that, during the period of analysis, only ethanol prices have been adjusting to correct disequilibriums from the long-run relationship. The ESTVECM parameter estimates (see table 1) suggest that neither oil nor corn prices have contributed to correcting deviations from the equilibrium. These results are compatible with previous research that has found oil prices to be exogenous to ethanol and sugar prices in the Brazilian biofuels market (see Balcombe and Rapsomanikis 2007). The same authors have found feedstock prices to be exogenous to ethanol prices. The strong demand for corn created as a result of the ethanol demand expansion has caused a shortage of corn supply that has prevented corn price from moving back to its long-run parity. In addition, parameter estimates suggest that the adjustment of ethanol prices only takes place for high absolute deviations from the equilibrium relationship \((G(s_{t-d})=1)\), but not for small ones \((G(s_{t-d})=0)\). Table 1 also provides evidence of no residual autocorrelation.

Generalized impulse-response functions (GI) are calculated using parameter estimates of the ESTVECM. Since the only endogenous price in the system is ethanol, we are interested in assessing how ethanol responds to positive shocks to oil and corn. As noted, variable-specific shocks correspond to one positive standard deviation of the ESTVECM residuals. The maximum forecast horizon is set to \(H=40\), i.e., a period of approximately two months. Two points, one where the transition function is close to 1, \(t=25\), and another where the transition function is close to 0, \(t=128\), are selected as the initialization point. Results are presented in figures 4 to 7.

From these graphs several implications for price relationships can be derived. A change in corn price when the system is far from the equilibrium relationship \((G(s_{t-d})=1)\) causes a response from ethanol price in the same direction (figure 4). This result is expected given the fact that feedstock prices represent the lion’s share of ethanol
production costs. The magnitude of the response increases with time and reaches a peak at about 10 days after the shock. Afterward the response decreases in magnitude until it’s not statistically different from zero at around 35.

The shock to corn prices could take place when the system is close to the equilibrium relationship \( G(s_{t-d}) = 0 \). Since price transmission mechanisms are only activated for relatively large deviations from the equilibrium, one would expect to find very small responses from ethanol prices. Figure 6 confirms this hypothesis by showing that ethanol price responses are only statistically significant for some observations up to the sixth day. The theoretical increases that should take place afterwards are not statistically different from zero. The increase in ethanol prices following the shock is very small compared to the increase when the price system is in the disequilibrium regime \( G(s_{t-d}) = 1 \).

Figure 5 shows that a shock to oil price when \( G(s_{t-d}) = 1 \) causes a change in the ethanol price of the same sign. Ethanol price response reaches a peak after approximately 10 days to decay thereafter and become non-significant after 30 days. However, if the shock to oil prices occurs when the system is close to the equilibrium (figure 7), price transmission mechanisms are not activated and ethanol prices experience only a very small increase within the first three days following the shock. The increase that takes place is considerably below the increase experienced under the disequilibrium regime.

**Concluding remarks**

This paper uses a STVECM to assess price relationships within the US ethanol industry. Specifically, we use daily futures observed from July 21, 2005 to May 15, 2007. Cointegration tests support the existence of a (single) long-run relationship between ethanol, corn and oil prices. The STVECM parameter estimates suggest that neither corn
nor oil prices contribute to correcting deviations from the equilibrium, which is compatible with previous research results. Only ethanol prices adjust to their long-run parity in a nonlinear fashion when the error correction term in absolute values is far from the threshold parameter. The adjustment process is slow as indicated by a small parameter estimate for $\gamma$.

Results also show that, during the period analyzed, there have been frequent regime changes, especially since the sharp increase in ethanol prices by mid-2006. This instability may be explained by the fact that fuel and corn prices are responding to random shocks that cannot be accommodated by inventories. Generalized impulse response functions indicate that a shock to both oil and corn prices when the system is far from its equilibrium, cause a change in ethanol prices of the same sign. Ethanol responses usually reach a peak after about 10 days of the initial shock and fade away after around 30-35 days.
References


Table 1. ESTVECM parameter estimates

<table>
<thead>
<tr>
<th>Equation</th>
<th>Parameter</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>LM(10)</th>
<th>p-value</th>
<th>R-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethanol</td>
<td>$s_{t-d}$</td>
<td>-0.03647**</td>
<td>0.00031</td>
<td>0.01644</td>
<td>0.00115</td>
<td>0.29</td>
<td>0.9834</td>
<td>0.2356</td>
</tr>
<tr>
<td>Corn</td>
<td>$s_{t-d}$</td>
<td>-0.68796</td>
<td>2.65437</td>
<td>1.75724</td>
<td>9.27112</td>
<td>1.14</td>
<td>0.3354</td>
<td>0.1685</td>
</tr>
<tr>
<td>Oil</td>
<td>$s_{t-d}$</td>
<td>0.38350</td>
<td>0.11911</td>
<td>0.51502</td>
<td>0.42652</td>
<td>0.37</td>
<td>0.9592</td>
<td>0.1209</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transition function</td>
<td>$\gamma$</td>
<td>0.44698**</td>
</tr>
</tbody>
</table>

$\gamma$ denotes statistical significance at the 5 per cent significance level.

LM(10) is the F variant of the LM test for no remaining autocorrelation described above.
Figure 1. Price Series

Figure 2. Estimated exponential transition function
Figure 3. Evolution of the value of the transition function over time

Notes: the transition function is represented by the solid thick line and is plotted on the left hand side axis. The transition variable (solid thin line), as well as the threshold value (dotted line) are plotted on the right hand side of the axis.
Figure 4. Generalized impulse response functions of STVECM: responses of ethanol prices to a positive shock to corn prices. Regime corresponding to $G(s_{t-d}) = 1$

Note: ■ indicates the response is statistically significant at the 5% level.
Figure 5. Generalized impulse response functions of STVECM: responses of ethanol prices to a positive shock to oil prices. Regime corresponding to $G(s_{t-d}) = 1$

Note: ■ indicates the response is statistically significant at the 5% level.
Figure 6. Generalized impulse response functions of STVECM: responses of ethanol prices to a positive shock to corn prices. Regime corresponding to $G(s_{t-d}) = 0$

Note: ■ indicates the response is statistically significant at the 5% level.
Figure 7. Generalized impulse response functions of STVECM: responses of ethanol prices to a positive shock to oil prices. Regime corresponding to $G(s_{t-d}) = 0$

Note: ■ indicates the response is statistically significant at the 5% level.