Cointegration: implications for the market efficiencies of the high fructose corn syrup and refined sugar markets

ORAL WILLIAMS and DAVID A. BESSLER

Eastern Caribbean Central Bank, St Kitts, West Indies and Texas A & M University, College Station, Texas, USA

The dynamic relationship between the prices of refined sugar and high fructose corn syrup is investigated using cointegration econometrics. The analysis is based on observational data, and although results indicate cointegration between 1984 and 1991 for the most part, alternative explanations for empirical regularities may exist and should be the subject of future research.

I. INTRODUCTION

US domestic sugar prices have been maintained substantially above world prices engendering sugar substitutes (notably high fructose corn syrup, HFCS) to be sold at a discount to sugar – a discount of 10–30% (Greer, 1991; USDA, 1990). Sugar substitutes have been successful in capturing as much as 42% of the US caloric sweetener market by 1984 (Leu, Schmitz and Knutson, 1987). This paper investigates the dynamic relationship between refined sugar price and HFCS prices. Cointegration econometrics is utilized to investigate possible linkages between these markets. The remainder of the paper is divided into four sections. First, an outline of US sugar policy is provided. Section III presents a very brief summary of the cointegration methods applied in the paper. Then, results from application of cointegration techniques to monthly refined sugar and HFCS prices are presented. Finally, Section V discusses the implications of the results in a context that is germane to policy and market efficiency issues.

II. US SUGAR POLICY

US sugar policy has insulated the domestic sugar industry through a combination of subsidies to producers and various trade restrictions. These subsidies have taken the form of direct payments, non-recourse loans, price support and purchases. For several decades trade restrictions such as tariffs, import duties and quotas have effectively limited imports and stimulated domestic production.

With domestic prices maintained above world prices, sugar beet producers and HFCS manufacturers have shared in the economic rents engendered by policy. In the absence of on-going governmental assistance the US sugar industry would undergo considerable structural change (Schmitz, Allen and Leu, 1985). The costs of US sugar policy include the welfare losses inherent in transferring income from consumers to producers at a cost borne by the taxpayer. Several studies have addressed the costs and benefits associated with US sugar policy and alternative policy options (Leu, Schmitz and Knutson, 1987; Nyankori and Hammig, 1991). One omission in these studies pertains to how pricing inefficiencies that arise due to policy affect the transmission of information within and among sweetener markets. Corn sweetener manufacturers benefit from the higher long-term stable support prices which have spurred investment in plant and equipment and research and development in corn wet milling products. Stable support prices have also facilitated the offering of price discounts, relative to sugar, and acquisition of market share by HFCS manufacturers. If the US sugar and HFCS industries conform to perfect capital markets in view of the linkage between them, both markets should account for elements in the separate markets in their respective information sets. Evidence is presented below that the refined sugar market does not respond to perturbations in the HFCS–refined sugar long-run equilibrium. HFCS prices do respond to such perturbations.
III. COINTEGRATION

The maximum likelihood methods of Johansen (1988) and Johansen and Juselius (1990) are used in this paper. They consider the cointegration problem as one of reduced rank regression. Let $X(t)$ be the vector of series under investigation (say refined sugar prices and high fructose corn syrup prices). The procedure is based on the error-correction representation:

$$\Delta x(t) = \mu + \Gamma(1)\Delta x(t - 1) + \ldots + \Gamma(k - 1)\Delta x(t - k + 1) + \pi x(t - 1) + \psi D_t + \varepsilon(t)$$

where $- \sum_{j=1}^{k} \pi(j)$ for $i = 1, \ldots, k - 1$ and

$$\pi = - [I - \pi(1) - \ldots - \pi(k)].$$

Here the $\pi(i)$ are $(p \times p)$ matrices of autoregressive parameters from a VAR in levels of $x(t)$ of lag order $k$, $\mu$ is a constant, $D_t$ is a set of predetermined variables (seasonal dummy variables or intervention dummy variables), $\psi$ the associated parameter(s) on these predetermined variables and $\varepsilon(t)$ is a white noise innovation term.

Equation 1 resembles a vector autoregression (VAR) model in first differences, except for the presence of the lagged level of $x(t - 1)$. Its associated parameter matrix, $\pi$, contains information about the long-run (cointegrating) relationship between the variables in the vector $x$. There are three possibilities of interest: (a) if $\pi$ is of full rank, then $x(t)$ is stationary in levels and a VAR in levels is an appropriate model; (b) if $\pi$ has zero rank, then it contains no long-run information, and the appropriate model is a VAR in differences; and (c) if the rank of $\pi$ is a positive number, $r$, and is less than $p$ (here in our case $p = 2$), there exist matrices $\alpha$ and $\beta$, with dimensions $p \times r$, such that $\pi = \alpha \beta'$; the latter case, $\beta' x(t)$, is stationary, even though $x(t)$ is not.

The treatment of the constant $\mu$ is particularly interesting, as under (c) the constant term can be decomposed into two parts, that in the intercept of the cointegrating relation ($\beta' x(t)$) and that representing a linear trend (see Johansen, 1992 for details). These alternatives lead to sequential hypothesis testing with respect to the rank ($r$) of $\pi$. If there is a linear trend in the model, label this hypothesis $H(r)$, which is unrestricted. If there is no linear trend in the model, label the hypothesis $H(0)$, which is restricted. Johansen (1992) provides the rationale for sequential hypothesis testing to decide jointly for the rank of cointegrating vector ($r$) and whether there is linear trend in the model. He suggests testing the hypothesis in the following sequence: $H(0)^*$, $H(0)$, $H(1)^*$, $H(1)$, . . . ; stop testing the first time we do not reject. These tests are carried out using the ordered ($\lambda_1 > \lambda_2 > \ldots > \lambda_p$) eigenvalues $\lambda_j$ or $\lambda_j^*$, $j = 1, \ldots, p$ (use an asterisk to indicate that the eigenvalues have been calculated without a linear trend in the model). The trace test considers the hypothesis that the rank of $\pi$ is less than or equal to $r$. The trace test with the time trend is given by

$$-2 \ln(Q) = - T \sum_{j=r+1}^{p} \ln(1 - \lambda_j)$$

(2)

An analogous test can be defined using eigenvalues calculated under the hypothesis of no time trend. Johansen and Juselius (1990) provide the asymptotic critical values (see their appendix, table A1 for linear time trend and table A3 without a time trend).

In case (c), the long-run equilibrium between the series is summarized by the vector $\beta$. The rate at which each series adjusts to perturbations in this equilibrium is given by the $\alpha$ matrix. Analysis of the elements of $\alpha$ and $\beta$ is fundamental to our understanding of long-run and short-run relations between economic time series data. Whether some of the elements of $\alpha$ or $\beta$ are zero is of particular interest. Tests of these restrictions ($\alpha = 0$ or $\beta = 0$) can be carried out with the statistic $T = N(\ln(1 - \lambda_{1,3}) - \ln(1 - \lambda_{1,2})$, where $\lambda_{1,j}$, $j = 2, 3$ are the eigenvalues calculated under the hypothesis of one cointegrating vector ($r = 1$) without $(j = 2)$ and with $(j = 3)$ the restriction.

IV. APPLICATION TO REFINED SUGAR AND HFCS DATA

We study monthly price data for refined sugar and HFCS. The data covered prices from January 1975 to May 1991 and were obtained from the Sugar and Sweetener Situation and Outlook reports (USDA, various dates). The data are studied in logarithmic units. These data are plotted in Figure 1. Also included in Figure 1 are the historical means of each series. Notice that both series show few crossings of their historical means over the entire sample. This suggests mean non-stationarity (more will be said about this below).

Our interest is to study the series sequentially over the entire

![Fig. 1. ln of sugar prices and ln of HFCS prices](Image)
period. As HFCS has evolved from less than 5% of the caloric sweetener consumption at the beginning of our data set to near 40% at the end of the period (see Lin and Novick, 1988; USDA, June 1991, Table 65 for consumption totals by type), it is not inconceivable that its pricing relationship to sugar may have changed as well.

Particular dates that stand out as a priori important (before we look at the data) are January 1980, when Coca-Cola first allowed HFCS to be used in place of refined sugar in its caloric soft drinks (see Pendergrast, 1993, p. 337). This initial use of HFCS was only partial, as the percentage of allowance of HFCS in caloric cola soft drinks was initially less than 50%, increased to 75% in 1983, and finally increased to 100% in 1984 (USDA, December 1983, p. 8 and March 1985, pp. 16–17). In addition, a period of interest which we became aware of only after our initial testing of the data is the interval of trading of the HFCS contract on the Minneapolis Grain Exchange, from April 1987 through November 1988. It will become apparent in our analysis that this period is of special interest for understanding the pricing efficiency in the caloric sweetener market.

We follow Juselius (1993) and test for the rank of $\pi$ in Equation 1. If we find that $\pi$ is of reduced rank, $0 < r < p$, then we can test the individual elements of $\beta^*$ against zero in the factorization $ab^* = \pi$. If one particular series is stationary then it will by itself provide for one cointegrating vector. In contrast to the usual approach of first testing each series individually for non-stationarity, using for example the Dickey–Fuller test, this approach puts all testing into one system. The most obvious difference between the usual approach of testing each series individually first and the approach used here is that the former has non-stationarity as the null hypothesis (Dickey–Fuller test, for example); while our approach has stationarity as the null hypothesis (more will be said about this below).

We test the rank of $\pi$ in an error correction representation having one lag of first differences, one lag of levels and eleven monthly dummy variables on the right-hand-side of Equation 1. Tests on the rank of $\pi$ are carried out for $\mu$ both in and not in the cointegrating vector. We selected one lag of the first differences in the above-described error correction model using sequential likelihood ratio tests on lags 0 through 3 in levels vector autoregressions (VAR); the test on lag 1 versus lag 0 resulted in a rejection of the null hypothesis that the coefficients associated with the higher order lag in the VAR are equal to zero at a marginal significance level of 0.00, the test of lag 2 versus lag 1 was rejected at a marginal significance of 0.00, and the test of lag 3 versus lag 2 was rejected at a marginal significance of 0.09. (Lutkepohl, 1985, suggests applying a 0.01 significance level for these tests.) A vector autoregression of order $k$ lags in levels implies an error correction model of order $k - 1$ (Johansen and Juselius, 1990).

In Figures 2 and 3, we show normalized trace tests calculated at each data point over the period 1979:1 through 1991:5. Figure 2 gives statistics $T^*$ and $T$ on the null hypothesis that $r = 0$ (the rank of $\pi = 0$). Figure 3 gives statistics $T^*$ and $T$ on the null hypotheses that $r \leq 1$. Statistics in both figures are normalized by the 5% critical values given in appendix tables A1 and A3 of Johanssen and Juselius (1990). (figure entries greater than 1.0 indicate that we reject the null hypothesis at that data point.)

Recall from above that the sequential testing suggested by Johanssen was in order: $T^*(r = 0)$, $T(r = 0)$, $T^*(r \leq 1)$, $T(r \leq 1)$; and to stop testing the first time we fail to reject the null hypothesis. Here $T^*(r = 0)$ refers to the test of the hypothesis that the rank of $\pi = 0$ and no linear trend exists in the levels data, $T(0)$ refers to the test of the hypothesis that the rank of $\pi = 0$ and there is a linear trend in the levels representation; similar descriptions exist for $T^*(r \leq 1)$ and $T(r \leq 1)$. From Figure 2 it is clear that several ‘structures’
exist over our study period. Before 1980, it is clear that $r \neq 0$, as we reject the null $r = 0$ under the * and no-* forms of the test (our plots of the trace test under both forms of the test are above the line 1.0 for both tests before 1980). After mid-1983, both trace tests on $r = 0$ increase with sample size (through time) suggesting that after mid-1983 $r \neq 0$. The period early 1980 through mid-1983 appears to suggest $r = 0$. Finally, there is a sharp drop and subsequent upturn in both trace statistics 1987–88 (indicating that $r$ may be equal to zero over this ‘window’).

From Figure 3, we see that $r$ is clearly less than or equal to one in the post mid-1983 period; as we fail to reject first the $T^* (r \leq 1)$ statistic at every date. In mid-1979 to early 1980 there is a small window where we reject $r \leq 1$ under the $T$ and $T^*$ forms of the test.

From Figures 2 and 3 it appears that the cointegrating relationships in our data have changed over time. The period before 1980 appears to represent a period in which both series may well be stationary in levels; the period from early 1980 to mid-1983 shows no long-run relationship (no cointegration); and the period post mid-1983 indicates one long-run relationship, with no time trend in the levels of the data. A ‘window’ of no cointegration appears in 1987 and early 1988.

As the sequential tests presented in Figures 2 and 3 may well have combined data over two or more ‘structures’, we re-estimate using data from January 1984 through May 1991. These tests are given in Table 1. We offer tests on three alternative versions of the error correction model. Version (i) is the test sequence for logged data with $k = 2$ lags and no predetermined variables ($\psi$ in Equation 1 is set to zero). Version (ii) is the test sequence for logged data with $k = 2$ and a set of eleven seasonal dummy variables (to capture seasonal affects associated with the yearly crop calendar). Finally version (iii) is the test sequence associated with seasonal dummy variables and an ‘intervention’ dummy variable associated with the 1987–88 trading of the HFCS futures contract on the Minneapolis Grain Exchange. Notice for all three versions of the sequential testing that we reject the null hypothesis that $r = 0$, without trend, we reject the null hypothesis that $r = 0$, with trend, and we fail to reject the null hypothesis that $r \leq 1$, without trend.

The possibility that this one cointegrating vector might arise because one of the underlying series is itself stationary was investigated through the likelihood ratio tests in Table 2. These tests provide evidence on the stationarity of each of the series, only here the null is not non-stationary, as with the Dickey–Fuller tests and Durbin–Watson tests, but rather stationary. We reject, at very low significance levels, the hypothesis that each series enters the cointegrating vector with weight zero (equivalently that the one cointegrating vector found in Table 1 was due to the other series being itself stationary). This finding appears to hold under all three versions of the error correction representation (models i, ii, and iii of Table 2).

In Table 3 we investigate the weak exogeneity of each series. Here we test each element of the $\alpha$ vector against zero. If we fail to reject one such test it implies that the changes in the associated variable do not respond to perturbations in the long-run equilibrium (deviations in the cointegrating vector). Results are consistent over all three versions of the error correction model. We strongly reject weak exogeneity of HFCS price. HFCS price does respond to perturbations in the long-run equilibrium (we reject the hypothesis that $\alpha_{21} = 0$ at a 0.00 significance level). On the other hand, we do not reject the hypothesis of weak exogeneity for sugar

Table 1. Trace tests on alternative representations (models i, ii, and iii) of cointegration between sugar prices and high fructose corn syrup prices, 1984:1–1991:5

<table>
<thead>
<tr>
<th>Model</th>
<th>HO $^2$</th>
<th>T $^*$ $^3$</th>
<th>C(5%)</th>
<th>HO $^2$</th>
<th>T $^4$</th>
<th>C(5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>$r = 0$</td>
<td>24.11</td>
<td>20.17</td>
<td>$r = 0$</td>
<td>21.43</td>
<td>15.20</td>
</tr>
<tr>
<td>ii</td>
<td></td>
<td>21.79</td>
<td></td>
<td></td>
<td>17.57</td>
<td></td>
</tr>
<tr>
<td>iii</td>
<td></td>
<td>32.04</td>
<td></td>
<td></td>
<td>22.98</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>$r \leq 1$</td>
<td>2.63**</td>
<td>9.09</td>
<td>$r \leq 1$</td>
<td>0.23</td>
<td>3.96</td>
</tr>
<tr>
<td>ii</td>
<td></td>
<td>4.70**</td>
<td></td>
<td></td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>iii</td>
<td></td>
<td>4.13**</td>
<td></td>
<td></td>
<td>0.70</td>
<td></td>
</tr>
</tbody>
</table>

$^1$ Model refers to alternative forms of the predetermined variables in the error correction representation. Model i has no predetermined variables. Model ii has a set of eleven seasonal dummy variables. Model iii has a set of eleven seasonal dummy variables and a zero–one dummy for the period (1987:5–1988:11) of trading of the HFCS futures contract on the Minneapolis Grain Exchange. All models have $k$ from Equation 1 set equal to 2.

$^2$ $r$ is the number of cointegrating vectors.

$^3$ $T^*$ indicates no trend in levels data. Critical value, at the 5% significance level, is given by Johansen and Juselius (1990), table A3.

$^4$ $T$ indicates trend in levels data. Critical value, at the 5% significance level, is given by Johansen and Juselius, (1990), table A1.

** indicates first non-rejection in Johansen sequential testing (Johansen, 1992).
Table 2. Test on exclusion of $\beta_0$ from the cointegrating vector in alternative error correction representations (models i, ii, and iii), 1984:01–1991:05$^1$

<table>
<thead>
<tr>
<th>Model</th>
<th>HO:</th>
<th>Chi-squared</th>
<th>Critical $\alpha$ level</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>$\beta_{11} = 0, \beta_{21} \neq 0, \beta_{31} \neq 0$</td>
<td>5.55</td>
<td>0.02</td>
</tr>
<tr>
<td>ii</td>
<td>$\beta_{11} = 0, \beta_{21} \neq 0, \beta_{31} \neq 0$</td>
<td>4.57</td>
<td>0.03</td>
</tr>
<tr>
<td>iii</td>
<td>$\beta_{11} = 0, \beta_{21} \neq 0, \beta_{31} \neq 0$</td>
<td>5.11</td>
<td>0.02</td>
</tr>
<tr>
<td>i</td>
<td>$\beta_{11} \neq 0, \beta_{21} = 0, \beta_{31} \neq 0$</td>
<td>18.45</td>
<td>0.00</td>
</tr>
<tr>
<td>ii</td>
<td>$\beta_{11} \neq 0, \beta_{21} = 0, \beta_{31} \neq 0$</td>
<td>12.37</td>
<td>0.00</td>
</tr>
<tr>
<td>iii</td>
<td>$\beta_{11} \neq 0, \beta_{21} = 0, \beta_{31} \neq 0$</td>
<td>23.77</td>
<td>0.00</td>
</tr>
</tbody>
</table>

$^1$The test is given as $T = N(\ln(1 - \lambda_{i1} - \ln(1 - \lambda_{i2})), j = 2, 3$, are the eigenvalues calculated under the restricted and the unrestricted hypothesis. Here the hypothesis is considered under the hypothesis of one cointegrating vector ($r = 1$) without ($j = 2$) and with ($j = 3$) the restriction that the parameters on raw sugar price and HFCS price enter the cointegrating vector with zero and non-zero values. The $\alpha$ column gives the marginal significance level at which the null hypothesis is rejected refers to alternative forms of the predetermined variables in the error correction representation. Model i has no predetermined variables. Model ii has a set of eleven seasonal dummy variables. Model iii has a set of eleven seasonal dummy variables and a zero-one dummy for the period (1987:5–1988:11) of trading of the HFCS futures contract on the Minneapolis Grain Exchange. All models have $k$ from Equation 1 set equal to 2.

price, at usual levels (we have to go to a 0.73 significance level to reject the null hypothesis that $\alpha_{1,1} = 0$).

The choice of particular versions of the error correction model (i, ii, or iii of Tables 1–3) can be aided by the diagnostic tests on residuals given in Table 4. Here we list normality. The $Q$-statistics are distributed chi-squared with 19 degrees of freedom under the null hypothesis of white noise residuals. The ARCH test is distributed chi-squared with two degrees of freedom under the null of no second-order ARCH effects (see Harvey, 1991) and the Jarque–Bera test is distributed chi-squared with two degrees of freedom under the null of normality.

From Table 4, we see little to object to with respect to the residuals from the $\Delta$ SUGAR equation. There appears to be no autocorrelation, no second-order ARCH effects and no serious departure from normality, under all three versions of the error correction model. On the other hand models (i) and (ii) from the $\Delta$ HFCS equation show problems with either autocorrelation or non-normality or both. The $Q$-statistic on residuals from the $\Delta$ HFCS equation of model (i) exceeds the critical value at a 1% significance level, as does the Jarque–Bera statistic on normality of the residuals.

While there is no strong evidence to indicate autocorrelated residuals in model (ii), the Jarque–Bera statistic indicates problems with the normality assumption. Model (iii) passes the autocorrelation and ARCH effects tests at levels lower than 5% and passes the Jarque–Bera normality test at approximately the 5% level.

The difference between model (ii) and model (iii) is that the latter has a zero–one dummy variable, indicating trading of the HFCS contract on the Minneapolis Grain Exchange in 1987–88 (both have seasonal dummy variables included). Evidently, such trading (or phenomena correlated with that trading) did have an effect on the residuals in the error correction representation.

Model (iii), which generates our prices over the 1984:1–1991:5 sub-period, is given as:

$$
\begin{bmatrix}
\Delta \text{SUGAR} (t)
\end{bmatrix}
= 
\begin{bmatrix}
0.056 & 0.024 \\
0.20 & 1.8 \\
1.053 & 0.257 \\
1.30 & 2.6
\end{bmatrix}
\begin{bmatrix}
\Delta \text{SUGAR} (t - 1)
\end{bmatrix} 
+ 
\begin{bmatrix}
0.000 & 0.000 & 0.000 \\
0.206 & -0.460 & 0.596 \\
0.26 & 0.08 & -0.05 \\
0.26 & 3.4 & 1.7
\end{bmatrix}
\begin{bmatrix}
\text{SUGAR} (t - 1) \\
\text{HFCS} (t - 1)
\end{bmatrix}
+ 
\begin{bmatrix}
0.002 \\
-0.064 \\
0.03 & 0.00 & 0.01 \\
1.8 & 0.1 & 2.8
\end{bmatrix}
[DUM(t)]
+ 
\begin{bmatrix}
0.00 \\
0.00 \\
0.00 \\
0.00
\end{bmatrix}
[D_D(t)]
+ 
\begin{bmatrix}
0.00 \\
0.00 \\
0.00 \\
0.00
\end{bmatrix}
[D_N(t)]
+ 
\begin{bmatrix}
0.26 & 0.05 & -0.4 \\
2.6 & 3.4 & 1.7
\end{bmatrix}
[D_F(t)]
$$

Here $\Delta$ SUGAR$(t)$ and $\Delta$ HFCS$(t)$ refer to first differences of the logarithms of refined sugar prices and the logarithms of...
high fructose corn syrup prices in period \( t \), respectively. \( \text{SUGAR}(t) \) and \( \text{HFCS}(t) \) refer to levels of the same variables, \( \text{DUM}(t) \) refers to a zero–one indicator variable, set equal to one for the period of trading of the HFCS contract on the Minneapolis Grain Exchange, and the variables \( D_{D}(t), D_{N}(t), \ldots, D_{F}(t) \) refer to eleven monthly zero–one variables, set equal to one if \( t \) corresponds to the month December, November, …, February, respectively and zero otherwise; e.g. \( D_{D}(t) = 1 \) if observation \( t \) occurs in December, and zero if not. The numbers in parentheses are \( t \)-statistics.

Our \( \pi \) matrix, the second parameter matrix on the right-hand side of Equation 3, is factored as:

\[
\alpha \beta' = \begin{bmatrix} 0.000 \\ 0.206 \end{bmatrix} \begin{bmatrix} 1.00 & -2.233 & 2.892 \end{bmatrix}
\]

The long-run equilibrium between refined sugar prices and high fructose corn syrup prices is given by \( \beta' = (1.00 - 2.233 + 2.892) \). Expressing perturbations in this equilibrium as \( z_t \), we have \( z_t = \ln(\text{sugar}_t) - 2.233 \ln(\text{hfcs}_t) + 2.892 \).

The variable \( z_t \) represents stationary deviations in the long-run equilibrium between these two prices, to which the high fructose corn syrup price responds through the \( \alpha \) vector (0.206); when the logarithm of sugar price is high relative to the long-run equilibrium in period \( t \), the logarithm of high fructose corn syrup price will increase in period \( t + 1 \). The logarithm of the price of sugar does not respond to shocks in the long run relationship.

The indicator variable \( \text{DUM}(t) \) suggests that trading of the HFCS futures contract resulted in significant negative changes in the logarithm of HFCS prices (\( t \)-statistic of 2.0) and no significant changes in the logarithm of the price of refined sugar (\( t \)-statistic of 0.3). The seasonal dummy variables show more pronounced seasonal effects in the logarithm of HFCS price than in the logarithm of refined sugar price – large negative effects being noted in the fall of the year and a large positive effect in April. Presumably the

### Table 3. Tests of weak exogeneity of sugar price and HFCS price in alternative error correction representations (models i, ii, and iii), 1984:01–1991:05

<table>
<thead>
<tr>
<th>Model</th>
<th>HO: ( \alpha_1 = 0, \alpha_2 \neq 0 )</th>
<th>Chi-squared</th>
<th>Critical ( \alpha ) level</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>( \alpha_1 = 0, \alpha_2 \neq 0 )</td>
<td>0.00</td>
<td>0.97</td>
</tr>
<tr>
<td>ii</td>
<td>( \alpha_1 = 0, \alpha_2 \neq 0 )</td>
<td>0.53</td>
<td>0.47</td>
</tr>
<tr>
<td>iii</td>
<td>( \alpha_1 = 0, \alpha_2 \neq 0 )</td>
<td>0.45</td>
<td>0.50</td>
</tr>
<tr>
<td>i</td>
<td>( \alpha_1 \neq 0, \alpha_2 = 0 )</td>
<td>21.66</td>
<td>0.00</td>
</tr>
<tr>
<td>ii</td>
<td>( \alpha_1 \neq 0, \alpha_2 = 0 )</td>
<td>11.83</td>
<td>0.00</td>
</tr>
<tr>
<td>iii</td>
<td>( \alpha_1 \neq 0, \alpha_2 = 0 )</td>
<td>23.24</td>
<td>0.00</td>
</tr>
</tbody>
</table>

1 The test is given as \( T = N(\text{ln}(1-\lambda_{i,t}) - \text{ln}(1-\lambda_{i,z})) \), where \( \lambda_{i,t}, i = 1,2,3 \) are the eigenvalues calculated under the restricted and the unrestricted hypotheses. Here both the restricted and the unrestricted hypotheses are considered under one cointegrating vector \( r = 1 \) without \( i = 2 \) and with \( i = 3 \) the restriction of ‘weak exogeneity’ of the series listed in the left-hand column. The marginal significance level at which the null hypothesis is rejected. Model refers to alternative forms of the predetermined variables in the error correction representation. Model i has no predetermined variables. Model ii has a set of eleven seasonal dummy variables. Model iii has a set of eleven seasonal dummy variables and a zero–one dummy for the period (1987:5–1988:11) of trading of the HFCS futures contract on the Minneapolis Grain Exchange. All models have \( k \) from Equation 1 set equal to 2.

### Table 4. Diagnostics on residuals from alternative error correction representations (models i, ii, and iii) fit to 1984:1–1991:5 monthly sugar and HFCS prices

<table>
<thead>
<tr>
<th>Model</th>
<th>( \Delta \text{SUGAR} )</th>
<th>( \Delta \text{HFCS} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BP(19)(^2)</td>
<td>ARCH(2)(^3)</td>
</tr>
<tr>
<td>i</td>
<td>8.28</td>
<td>2.34</td>
</tr>
<tr>
<td>ii</td>
<td>10.56</td>
<td>5.56</td>
</tr>
<tr>
<td>iii</td>
<td>10.83</td>
<td>6.21</td>
</tr>
</tbody>
</table>

1 Model refers to alternative forms of the predetermined variables in the error correction representation. Model i has no predetermined variables, model ii has a set of eleven seasonal dummy variables. Model iii has a set of eleven seasonal dummy variables and a zero–one dummy for the period (1987:5–1988:11) of trading of the HFCS futures contract on the Minneapolis Grain Exchange. All models have \( k \) from Equation 1 set equal to 2.
2 BP(19) refers to the Box–Pierce \( Q \)-statistic. It is distributed chi-squared with 19 degrees of freedom. Critical values at the 0.05 and 0.01 significance levels are 30.1 and 36.2, respectively.
3 ARCH(2) is the Lagrangian multiplier test for second-order autoregressive conditional heteroscedastic errors, which is distributed chi-squared with two degrees of freedom under the null hypothesis of no second-order ARCH effects. Critical values at the 0.05 and 0.01 levels are 5.99 and 9.21, respectively.
4 JBNOR(2) is the Jarque–Bera test for normality of the residuals, which is based on transformations of the skewness and kurtosis of the observed residuals; see Jarque and Bera (1980). Under the null hypothesis of normality the test statistic is distributed chi-squared with two degrees of freedom. Critical values at the 0.05 and 0.01 levels are 5.99 and 9.21, respectively.
negative effects are related to corn harvest and the positive effect reflects preparation for the increasing seasonal demand for soft drinks.

As additional evidence that these price series are tied together by cointegration, we consider out-of-sample forecasting results from an error correction representation versus forecasts from a vector autoregression in first differences, which is equivalent to repeating the above procedure with \( r = 0 \) (the number of cointegrating vectors set equal to zero). Following Engle and Granger (1987), we expect that, if HFCS and refined sugar are cointegrated, forecast performance from the \( r = 1 \) models should be superior to that from the \( r = 0 \) models (root mean squared error should be lower) for at least one of the series, at long forecast horizons.

Both models were estimated over data points 1984:1–1988:12. Each model was then used to forecast the subsequent six (out-of-sample) data points – 1989:1, 1989:2, 1989:3, 1989:4, 1989:5, and 1989:6. The models were then moved ahead one data point (to 1989:1), re-estimated and the next six data points were forecasted. This procedure was repeated through 1991:4, giving us 28 one-step-ahead forecasts, 27 two-step-ahead forecasts, ... , and 23 six-step-ahead forecasts. Here we do not impose the condition that \( \alpha_{11} = 0 \) and \( \alpha_{12} \neq 0 \), but allow \( \alpha \) for both refined sugar and HFCS to be non-zero in the \( r = 1 \) forecasting models.

Forecast performance was evaluated using the root mean square error (RMSE) criterion for the levels-converted error correction model and the levels-converted difference VAR. Figures 4 and 5 give the root mean squared errors. The error correction forecasts of HFCS prices show considerable improvement, relative to the difference VAR. On the other hand refined sugar prices are forecasted better using a difference VAR \( (r = 0) \). These results are consistent with those found above. Refined sugar prices do not respond to perturbations in the long-run equilibrium (to perturbations in \( \beta \), that is \( \alpha_{11} = 0 \)) and forecasts of refined sugar prices which allow non-zero values of \( \alpha_{11} \) are not better than forecasts of refined sugar prices from a VAR in first differences (which sets \( \alpha_{11} = 0 \)). Just the opposite holds for forecasts of HFCS prices – the error correction forecasts are superior to the difference VAR forecasts (which is what cointegration would suggest).

V. CONCLUSION

This article applied the literature of cointegration to analyse refined sugar prices and HFCS prices. The results from within-sample tests of fit and out-of-sample forecasting indicate that refined sugar prices and HFCS prices have been cointegrated over much of the period 1984–91. Before 1984, the within-sample tests of fit indicate that these prices were not consistently cointegrated. Further, there was an additional ‘window’ of time, in 1987 and 1988, when these prices were not cointegrated. This window corresponds, approximately, with the trading on the Minneapolis Grain Exchange of the HFCS futures contract. The overall results suggest, quite clearly, that the price of HFCS was being determined by the refined sugar price over much of the 1984–91 time period. Participants in the HFCS market appear to be looking adaptively at the refined sugar price and setting prices in a ‘lock–step’ fashion. There is no long-run feedback from HFCS prices to refined sugar prices. It is interesting that the brief window of trading of the HFCS futures contract on the Minneapolis Grain Exchange apparently broke the ‘lock–step’ relation between refined sugar prices and HFCS prices. This is what one would expect in ‘efficient’ markets. That is, the trading on the Minneapolis Grain Exchange apparently squeezed out any predictable component in the HFCS–refined sugar market. The sign on our dummy variable corresponding to trading over that period was negative, indicating that HFCS price would be
lower if such futures trading was re-instituted. It is perhaps interesting as well that the cointegration relation did not begin until early 1984, when major soft drink manufacturers adopted a 100% allowance of HFCS in their caloric drinks.

Of course, the analysis presented here is based on observational data. Alternative explanations for the empirical regularities may exist and should be discussed in subsequent literature.

ACKNOWLEDGEMENTS

David W. Bullock of the Minneapolis Grain Exchange and Fred Linse of the Commodity Futures Trading Commission provided helpful information on the trading duration of the Minneapolis Grain Exchange’s futures contract on HFCS. Thanapat Chaisautikulawat wrote the RATS computer program used to forecast the error correction representations used in this paper. A referee provided helpful comments on an earlier version of the paper.

REFERENCES


US Department of Agriculture (various dates) *Sugar and Sweetener Situation and Outlook*, United States Department of Agriculture, Washington, DC.