The Australian Journal of
Agricultural and Resource Economics

Journal of the Australian Agricultural and Resource Economics Society

The Australian Journal of Agricultural and Resource Economics, 56, pp. 542–557

Agricultural commodities pricing model applied to the Brazilian sugar market

Leonel M. Pereira, Celma de Oliveira Ribeiro and José R. Securato[†]

This article suggests a pricing model for commodities used to produce biofuel. The model is based on the concept that the deterministic component of the Wiener process is not constant and depends on time and exogenous variables. The model, which incorporates theory of storage, the convenience yield and the seasonality of harvests, was applied in the Brazilian sugar market. After predictions were made with the Kalman filter, the model produced results that were statistically more accurate than those returned by the two-factor model available in the literature.

Key words: biofuels, commodity, ethanol, pricing, sugar.

1. Introduction

The problem addressed in this article is the price formation of agricultural commodities involved in biofuel production. The option to substitute renewable alternatives such as ethanol produced from sugar cane in place of fossil fuels derived from petroleum has thrust the price formation of certain commodity goods into a new light. The proposed model was applied and tested in the Brazilian sugar market because of the relationship between the price of sugar and the price of ethanol derived from sugarcane and because sugar is the main Brazilian commodity that is both related to the concept of renewable energy and traded in futures contracts.

To build the model, certain foundational premises were set forth based on the interdependence of commodity prices and petroleum prices, the theory of storage, the seasonality of harvests and market volatility. The proposed model uses a generalised form of the deterministic component of the stochastic process of commodity returns because it considers the influence of time and exogenous variables. Previous models in the literature are based on the assumption that returns follow a constant deterministic trend. This premise effectively makes the previous models particular cases of the model herein proposed.

[†] Leonel Molero Pereira (email: leonelmlp@gmail.com, leonelmp@insper.edu.br, leonel@usp.br) and José Roberto Securato are at the Department of Economics, Administration, and Accounting (FEA), University of São Paulo, São Paulo – SP, Brasil. Celma de Oliveira Ribeiro is at the Department of Production Engineering (Escola Politécnica), University of São Paulo, São Paulo – SP, Brasil.

doi: 10.1111/j.1467-8489.2012.00594.x

2. Biofuels

Brazil and the United States currently are the world's leading countries in the use of alternative fuels, primarily as ethanol either added to or substituted for gasoline. In Brazil, PROALCOOL, the National Ethanol Program (Programa Nacional do Álcool), was created in the 1970s to incentivise the use of ethanol in place of gasoline in light vehicles. Brazil's current annual consumption of ethanol, at 18 billion litres, is equivalent to the nation's annual consumption of gasoline. This is due largely to Brazil's fleet of flex-fuel vehicles, which use gasoline and ethanol interchangeably or in combination as a mixture of the two fuels. The difference between the Brazilian and American ethanol markets is the source of the raw material used to produce the fuel: sugar cane in Brazil and corn in the United States.

Recently, the use of biofuels derived from foodstuffs has turned problematic. Apologists point out that plants sequester carbon from the atmosphere and help reduce the effects of global warming. On the other hand, critics of biofuels assert that the use of agricultural products to produce energy can lead to food shortages. Moreover, critics claim that increasing the planted area of agricultural commodities destined for biofuel production can accelerate tropical deforestation.

According to UNICA, the Sugar Cane Industry Association (União da Indústria de Cana-de-açúcar), which represents sugar cane advocates in Brazil, sugar cane grown on only 1 per cent of Brazil's arable land would supply enough energy to substitute for half of the fuel used to power light vehicles in Brazil. EPE, the Energy Research Company (Empresa de Pesquisa Energética) identified sugar cane as the second-largest energy source in Brazil, as measured in equivalent tonnes of petroleum, surpassing even hydroelectric power.

Given this situation, the necessity has arisen for a commodity price–formation model which recognises that certain agricultural products have become part of the energy matrix and therefore are likely to be priced in a way that is interdependent with the price of oil.

3. Theoretical foundations

Previous commodity pricing models have been rooted in the theory of storage. In their seminal papers, Kaldor (1939), Working (1948, 1949), Brennan (1958), Telser (1958) and Johnson (1960) use the theory of storage to explain price differences in spot and futures markets for commodities.

According to the theory of storage, holding the physical commodity confers a nonpecuniary economic benefit, measured in terms of the convenience yield, a concept developed by Fama and French (1987), Benirschka and Binkley (1995), Litzenberger and Rabinowitz (1995) and Frechette and Fackler (1999).

The convenience yield, as defined in the literature, is an intangible economic concept that is a value, expressed as an interest rate, beyond the spot price of the asset. It can be understood as the economic advantage of physical possession of the commodity, in that possession allows the owner to wait for the best moment to sell his or her stock to maximise the value of the sale or to keep the production process running. The convenience yield is highest when global stocks of the commodity are at their lowest.

With their foundations in the theory of storage, commodity pricing models evolved over time to include additional variables to explain price behaviour more effectively. The principal-added variables have been the convenience yield, interest rates and seasonal factors.

Modelling commodity prices began with Brennan and Schwartz (1985) to evaluate the viability of investments in copper mining. The authors suggested a single-factor model that used as its only independent variable the spot price of the commodity as described by standard Brownian motion.

Gibson and Schwartz (1990) built a two-factor model, which they applied to the price of a barrel of oil. They assumed that the spot price and the convenience yield move in a joint process in which the spot price of the commodity undergoes Brownian motion, while the convenience yield follows a mean-reversion process.

In later work, Schwartz (1997) compared the one- and two-factor models and suggested a three-factor version in which the interest rate, following the Ohrnstein–Uhlenbeck process of mean-reversion, joined the spot price and the convenience yield as an independent variable. Schwartz concluded that this three-factor model was superior to the single-factor model but was not significantly different from the two-factor model in terms of the results produced.

More recently, Schwartz and Smith (2000) developed a model that assumes that the spot price undergoes a process of mean-reversion and that futures contracts with longer-term maturities tend towards an equilibrium price. This short-term/long-term model considers that short-term shocks, defined as the difference between the spot price and the long-term equilibrium price, revert to zero. For Schwartz and Smith (2000), random short-term shocks are the results of changes in short-term demand caused by weather changes and intermittent supply disruptions. The ability of market participants to adjust their stock levels makes up for these supply irregularities. This model in empirical use requires the prices of short- and long-term contracts in futures markets.

Later, Sorensen (2002), adding a seasonal component to the model of Schwartz and Smith (2000), suggesting that the prices of agricultural commodities are cyclical. This occurs because supply drops between harvest seasons and rises during harvest seasons.

The following section presents the premises of a new model, applied in the Brazilian sugar market, that is based on more up-to-date factors that influence the formation of prices of commodities linked to the production of biofuels.

4. The proposed model

Previous models in the literature assumed that the deterministic trend of the returns of a commodity is a constant estimated in the stochastic process. This restrictive premise limits the practical application of the models in price formation of agricultural commodities. It is probable that observable exogenous variables lead to persistent changes in the trend of a commodity's returns and that, therefore, a constant is of questionable value in estimating this trend. It is for this reason that the model adopts a generalised form of Brownian motion for the returns of the price of a commodity whose deterministic trend depends on time and exogenous variables.

The model assumes that there is an interdependent relationship in the formation of prices of commodities. Price movements in one market can predict the path that prices will follow in an interrelated market. The link between oil and commodities used in bioenergy production is an example of this kind of relationship.

In the proposed model, we assume the following:

- *Interdependence between the prices of sugar and ethanol:* the principal subproducts of sugar cane are ethanol and sugar. Factories that reduce sugar cane to its component products can alter the proportion in which they produce ethanol and sugar. Given this fact, the model assumes that a rise in demand that alters the price of ethanol can in turn affect the price of sugar and vice versa, that is, movements in the price of sugar also can cause changes in the price of ethanol.
- *Impact of the price of oil:* the model takes as given that the price of oil can influence the price of ethanol and, consequently, the price of sugar. The return on futures contracts for WTI oil on the NYMEX is used as an observable exogenous variable in the model and is considered as a component part of the process of sugar spot-price formation.
- *Spot price of sugar:* the spot price of sugar is considered to be a nonobservable variable and a state variable. Observations of sugar spot-price series can be problematic because the spot price of sugar is the average of prices struck in different places, and a shortage in one or more particular places can distort the average price. The return of the spot price of sugar therefore is described by generalised Brownian motion, with a deterministic component that depends on time and exogenous variables.
- *The price of sugar–futures contracts:* the prices of futures contracts are taken into account in the model as an observable signal variable, due to the transparency of bargaining and the possibility of arbitrage between different sugar futures markets. In the absence of arbitrage, the difference between the spot and future price is considered the equivalent of the difference between the risk-free rate and the convenience yield.
- *Convenience yield:* in line with the theory of storage, the model assumes that producers have the ability to retain stocks in hopes of receiving better prices later. The likelihood of market shortfalls caused by climatic factors

or supply disruptions increases the convenience yield. Market participants must be able to adjust their stock levels to accommodate the possibility of this random shock occurring, thereby bringing about a reversion to the long-run mean. In the model described in this article, the convenience yield is considered to be a state variable, measured in terms of the interest rate, which follows a mean-reversion process.

- *The influence of volatility in the markets:* another relevant premise of the model is the fact that increased demand for commodities can be the result of heightened volatility in financial markets. When financial markets pass through the periods of turbulence caused by threatening economic scenarios or bouts of inflation, capital flows turn to more secure investment alternatives for the purposes of capital preservation. Commodities maintain their value regardless of inflation. Demand for commodities in such a situation can drive up prices in the short term to levels that do not reflect the actual degree of physical demand for the goods. In the proposed model, conditional volatility is treated as an observable exogenous variable and is calculated using an ARMA/GARCH process.
- *Seasonality:* the model accounts for seasonality (i.e. harvest and non-harvest periods) in sugar cane spot-price formation using a sine function.

The proposed model is composed of a system of three equations in state-space form: the first describes the equilibrium relationship between spot prices and futures contracts, the second describes the stochastic process of the convenience yield and the third and most important equation of the model describes the trajectory of spot prices and includes three exogenous factors.

In conditions of no arbitrage, the difference between the future price and the spot price of the commodity is formed by the risk-free rate and the convenience yield of the period, as shown in the following equation:

$$F_t = S_t e^{(r-\delta_t)(T-t)} \tag{1}$$

where: S_t: spot price of the asset at time t; F_t: future-contract price at time t, with expiration at time T, which comes later than t; r: risk-free rate; δ_t: convenience yield at time t.

Taking the log of this equation describing a condition of no arbitrage between the future and spot price yields the following:

$$\ln(F_t) = r(T-t) + \ln(S_t) - \delta_t(T-t) \tag{2}$$

The model's second equation defines the process of mean-reversion for the convenience yield. In the case of a momentary supply shortage caused by an unforeseeable random phenomenon, the holder of the physical commodity benefits from the short-term price distortion, and the convenience yield reaches a maximum. As stock levels return to normal, the convenience yield reverts to its long-term mean. For this reason, the stochastic process of the convenience yield is written as follows:

$$\mathrm{d}\delta_t = \phi(\mu_\delta - \delta_t)\mathrm{d}t + \sigma_\delta \sqrt{\mathrm{d}t} z \tag{3}$$

where: δ_t: convenience yield at time t; ϕ: rate of reversion to the mean for the convenience yield; μ_δ: long-term mean of the convenience yield; σ_δ: standard deviation of the convenience yield.

The preceding equation, when converted into discrete changes in time with $\Delta t = 1$, gives the second equation used in the model:

$$\delta_t = \phi\mu_\delta + (1 - \phi)\delta_{t-1} + \sigma_\delta z \tag{4}$$

The third equation describes the path of the spot price of the commodity. This equation is the most important one because it treats the variable whose behaviour the model aims to describe. Also, it is the equation that is used for variable forecasting in empirical testing of the model. This third equation represents the Brownian motion of the returns of the spot price of the commodity in the generalised form. In standard Brownian motion, the parameters for the mean (μ) and the variance (σ^2) are constant. The generalised form permits these parameters to vary over time (Dixit and Pindyck 1994) as follows:

$$\mathrm{d}y = a(y, x, t)\mathrm{d}t + b(y, x, t)\sqrt{\mathrm{d}t} z \tag{5}$$

Here, $\sqrt{\mathrm{d}t} z$ is an increment in the Wiener process, $a(y, x, t)$ is the instantaneous drift and $b(y, x, t)$ is the instantaneous standard deviation, all of which have become dependent on time and current state. Generalised Brownian motion allows the drift of the spot price to be modelled in terms of exogenous variables cited in the proposed theory of biofuel price formation. The equation that represents the process adopted for the spot price of the commodity has a variable deterministic component and a constant standard deviation, as shown below:

$$R_t^S = \mu_{S,t}\mathrm{d}t + \sigma_S \sqrt{\mathrm{d}t} z \tag{6}$$

where: R_t^S: continuous return of the spot price of the commodity at time t, with $R_t^S = \ln(S_t) - \ln(S_{t-1})$; $\mu_{S,t}$: mean of the returns of the spot price on date t; σ_S: standard deviation of the returns of the spot price.

It is therefore possible to represent the process that describes the return of the spot price R_t^S for a unit of time in the following form:

$$R_t^S = \mu_{S,t} + \sigma_S z \tag{7}$$

In discrete time, for $\Delta t = 1$, the logarithm of the spot price can be solved for:

$$\ln(S_t) = \mu_{S,t} + \ln(S_{t-1}) + \sigma_S z \tag{8}$$

The deterministic trend of the spot price $\mu_{S,t}$, in generalised form, can undergo persistent changes over time, caused by observable exogenous variables, and can be expressed through a linear relation where $\mu_{S,\,t} = \beta'\mathbf{x}_t$:

$$R_t^S = \beta'\mathbf{x}_t + \sigma_S z \tag{9}$$

where: $\mathbf{x}_t$: $n \times t$ matrix of observable exogenous variables; β: $n \times 1$ vector of linear coefficients.

The following sections present the variables, or exogenous factors, that are part of the model used to describe the commodity-price trajectory.

4.1. Relation between petroleum and the commodity price

The deterministic component $\mu_{S,t}$ that describes the path of the returns of the spot price cannot be a constant parameter estimated in the model. There are other variables that can help anticipate the tendency of the commodity price.

In the case of agricultural commodities linked to biofuel production, the price of petroleum can be one of these variables. A biofuel commodity can help cover an energy shortfall caused by an oil shortage, and such a situation interferes with the trajectory of its price. Additionally, an increase in oil prices can translate into commodity price increases through production costs, according to Harri *et al.* (2009).

The mean of the returns of the commodity's prices can depend on the lagged returns of the price of petroleum R_{t-1}^P. The equation that represents this linear relation will be presented in the sections that follow.

4.2. Relation between market volatility and the commodity price

Investors can use commodities, being real goods with economic value in the supply chain, as secure means of capital preservation. Commodities, by definition, are assets that maintain their value regardless of currency denomination and therefore can be used as value benchmarks. Turbulent economic periods, set off by bouts of inflation or economic crises, can provoke momentary distortions in the relative value of currencies. In such circumstances, investors seek shelter in commodities markets to protect their capital. This rush into capital preservation causes a spike in commodity prices that is not directly related to physical demand for the goods themselves.

Therefore, the deterministic component of the commodity price, $\mu_{S,t}$, undergoes persistent changes over time due to the market's conditional volatility, $\hat{\sigma}_t^M$. In the model, market volatility is the standard deviation of the continuous returns of a portfolio of all assets in the market. As it is not possible to have an actual portfolio with these characteristics, a representative index from the stock market is used as a proxy.

4.3. Seasonal component

Agricultural commodities undergo price oscillations due to the timing of the planting and harvest seasons. This happens because, according to the theory of storage, just before the harvest commodity stocks are low, allowing holders of the physical goods to obtain better prices; prices will be lower just after the harvest. This seasonal effect on agricultural commodity prices is not related to market volatility or the price of oil and therefore must be included in the model as an additional component exclusively used to describe this cyclical effect on price formation. On the basis of this theory, a deterministic seasonal component, λ_t, was added to the equation which describes the stochastic process of the commodity's spot price.

In the literature, there are many ways to model seasonality. Hannan *et al.* (1970) suggest a sine function to describe the cyclical effect in time series. The sine function allows one to identify the cycles, amplitude and the frequency of seasonality over a period of 1 year. Other approaches in the literature suggest additional monthly dummy variables, either multiplicative or additive, to model seasonality.

However, according to Sorensen (2002), the sine function is preferred over dummy variables because it permits greater flexibility in the timescale used in the empirical analysis. The sample data can be drawn on either a monthly or a daily basis; the model will be basically identical for any time interval.

The model's seasonal component, λ_t, can be described by the following sine function:

$$\lambda_t = \sum_{k=1}^{K} \left[\beta_{c,k}\cos(2\pi k\theta) + \beta_{s,k}\sin(2\pi k\theta)\right] \tag{10}$$

where θ: time, in years; K: frequency of annual seasonal cycles; k: 1, ..., K number of parameters estimated in the sum; $\beta_{c,k}$, $\beta_{s,k}$: coefficients estimated for each of the K cycles.

Lastly, the equation for the process that describes the return of the spot price, $R_t^S = \beta'\mathbf{x}_t + \sigma_S z$, includes three exogenous factors and can be represented as follows:

$$R_t^S = \beta_0 + \beta_1 R_{t-1}^P + \beta_2 \Delta\hat{\sigma}_t^M + \lambda_t + \sigma_S z \tag{11}$$

where: R_t^S: continuous return of the spot price of the commodity at time t; R_{t-1}^P: continuous return of petroleum in a discrete unit of time at t−1; $\Delta\hat{\sigma}_t^M$: first difference of the conditional market volatility at time t; λ_t: deterministic seasonal component at time t; β_0, β_1, β_2: linear coefficients; σ_S: standard deviation of the returns of the spot price.

On the basis of the preceding equation, the commodity's returns can be described in the form of a constant, plus the lagged return of petroleum, by the first difference of the conditional volatility of the market and a seasonal component.

For $\Delta t = 1$ in the equation above, the logarithm of the spot price can be represented in the following form:

$$\ln(S_t) = \beta_0 + \beta_1 R_{t-1}^P + \beta_2 \Delta\hat{\sigma}_t^M + \lambda_t + \ln(S_{t-1}) + \sigma_S z \tag{12}$$

4.4. Model in the state-space form

The concepts elaborated above allow the construction of a commodity pricing model in state-space form using a system of three equations in discrete time that define the variables' stochastic processes:

$$\begin{cases} \ln(F_t) = r(T-t) - (T-t)\delta_t + \ln(S_t) \\ \delta_t = \phi\mu_\delta + (1-\phi)\delta_{t-1} + \sigma_\delta z \\ \ln(S_t) = \beta_0 + \beta_1 R_{t-1}^P + \beta_2 \Delta\hat{\sigma}_t^M + \lambda_t + \ln(S_{t-1}) + \sigma_S z \end{cases} \tag{13}$$

The first equation defines the no-arbitrage condition for the logarithm of the commodity's future price. The logarithm of the spot price and the convenience yield are the state variables in current form. There is no random component.

The second is a state equation that defines the mean-reversion process of the convenience yield. It has a random component whose variance is represented by σ_δ^2. The equation does not include exogenous variables.

The model's third equation is a state equation that represents the Brownian motion of the spot price in a generalised form with mean $\mu_{S,t}$ dependent on time and exogenous variables. The lagged return of petroleum, the primary difference in the market volatility and the seasonal component are the equation's exogenous factors that describe the mean of the spot-price return. This equation has a random component whose variance is represented as σ_S^2.

4.5. Estimation process

A model in standard state-space form is represented by:

$$y_t = Z_t\alpha_t + d_t + \varepsilon_t \tag{14}$$

$$\alpha_t = T_t\alpha_{t-1} + c_t + R_t\eta_t \tag{15}$$

The parameters of the proposed model were estimated using standard iterative techniques, to maximise the likelihood with respect to the unknown parameters under the assumption that ε_t and η_t are Gaussian, see Harvey (1989, p. 126).

The deterministic component c_t in equation (15) can be described by a linear function of unknown parameters, where $c_t = B_t u_t$, B_t is a matrix of estimated parameters and u_t is the vector of observable exogenous variables. Thus, we can represent the state-space equations in the following matrix form:

$$\ln(F_t) - r(T-t) = \begin{bmatrix} -(T-t) & 1 \end{bmatrix} \begin{bmatrix} \delta_t \\ \ln(S_t) \end{bmatrix} \tag{16}$$

$$\begin{bmatrix} \delta_t \\ \ln(S_t) \end{bmatrix} = \begin{bmatrix} (1-\phi) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \delta_{t-1} \\ \ln(S_{t-1}) \end{bmatrix} + B_t u_t + \begin{bmatrix} \eta_{\delta_t} \\ \eta_{s_t} \end{bmatrix} \tag{17}$$

The equations (16) and (17) correspond to equations (14) and (15), respectively. The matrix $B_t \in \mathrm{M}_{2\times(3+2k)}$ and vector u_t can be represented in the following way:

$$B_t = \begin{bmatrix} \phi\mu_\delta & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \beta_0 & \beta_1 & \beta_2 & \beta_{c,1} & \beta_{c,2} & \cdots & \beta_{c,k} & \beta_{s,1} & \beta_{s,2} & \cdots & \beta_{s,k} \end{bmatrix} \tag{18}$$

$$u_t' = \begin{bmatrix} 1 & R_{t-1}^P & \Delta\hat{\sigma}_t^M & \cos(2\pi\theta) & 0 & \cdots & 0 & \sin(2\pi\theta) & 0 & \cdots & 0 \end{bmatrix} \tag{19}$$

The vector c_t has a nonlinear relationship between parameters, ϕ and μ_δ are multiplied to each other. However, we use a relatively simple resource to make the relationship of the linear parameters. Referring to $\Gamma = \phi\mu_\delta$ in the matrix B_t, we get:

$$B_t = \begin{bmatrix} \Gamma & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \beta_0 & \beta_1 & \beta_2 & \beta_{c,1} & \beta_{c,2} & \cdots & \beta_{c,k} & \beta_{s,1} & \beta_{s,2} & \cdots & \beta_{s,k} \end{bmatrix} \tag{20}$$

Given the optimal estimators B_t^* and T_t^*, we have $\Gamma^* = \phi^*\mu_\delta$, from where we obtained μ_δ. Note that the two-factor model of Schwartz (1997) also presents the model in state-space form and possesses similar nonlinearity properties to the parameters of the convenience yield.

5. Application of the model

The proposed model was applied in the Brazilian sugar market, with parameters estimated and predictions made with the Kalman filter by employing data on the sugar futures market available from the BM&FBOVESPA – Brazilian Stock and Futures Exchange. The objective of the empirical analysis was to compare the results obtained from the proposed model with the results produced by the two-factor model of Gibson and Schwartz (1990) and to determine which model is superior in terms of forecasting capacity.

The two-factor model was chosen for comparison because, according to Schwartz (1997), this model is superior to one- and three-factor models. The short-term/long-term model can be used as an alternative for comparison, but it depends on the futures market having liquidity in both short- and long-term contracts, that is, a curve amply populated with commodity futures contracts of various expiration periods. This is not what one finds in practice in the market for sugar futures on the BM&FBOVESPA in São Paulo.

The sample data collected included daily and monthly data taken from the period from 2 January 2002 to 30 June 2008 and totalled 1631 daily and 78 monthly observations. Backwardation, the condition of spot prices being higher than prices in the futures market, appeared in 87.49 per cent of the data.

The spot price of sugar was obtained in Brazilian reais from the Center for Advanced Studies on Applied Economics (CEPEA) and was then converted to dollars using the average daily exchange rate to standardise all variables in the model in terms of currency. Futures contracts executed on the BM&FBOVESPA are quoted in dollars per 50 kg bag. The sample of daily settlement prices of the first maturity of the sugar futures contracts was collected. As the contract approached its expiration date, it was rolled over to the following expiration date at least 10 days before expiring. This process was adopted to avoid short-term distortions in the final settlement days before the contract's expiration. For NYMEX oil futures, only the first maturity month was used as a reference. The federal funds rate announced daily by the Federal Reserve (in dollars) was used as the risk-free rate.

The graphs in Figure 1 show the daily series of values for the future price of sugar (FUT), the spot price of sugar (ACUCARBR), the price per barrel of WTI crude oil (PETR) and the Dow Jones index (DJONES).

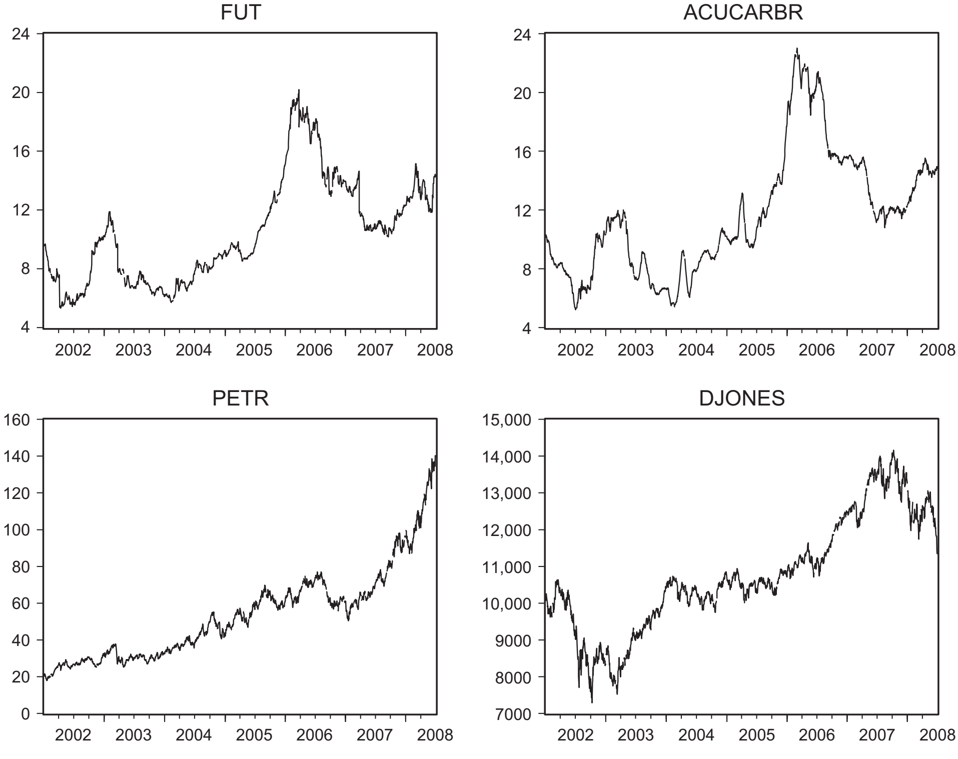

Figure 1 Graphs of variables – daily data.

The conditional volatility of the DJONES was calculated as the square root of the conditional variance generated through the GARCH model. Figure 2 shows the returns of the Dow Jones index (RETDJONES) and the conditional volatility calculated through the GARCH process (VOL) using both daily and monthly values. In the graphs, it is possible to identify the moments of greater oscillation in the index's returns and, therefore, higher volatility.

Before estimating the model's parameters, unit-root and co-integration tests were performed by following Engle and Granger's (1987) methods. The variables used to estimate the model's parameters must be stationary, and the regression's residuals must have a mean of zero and a finite variance and must not show serial correlation or trend. The unit-root test used was augmented Dickey and Fuller (1979, ADF), while the test to verify whether the model's variables co-integrate was based on Johansen's (1995) approach. The results of the tests demonstrated that the model's variables – sugar price, oil price and market volatility – are first-order integrated $I(1)$ but are first- and second-difference stationary. As the variables are first-order integrated, co-integration was tested. The null hypothesis was not rejected at the 5 per

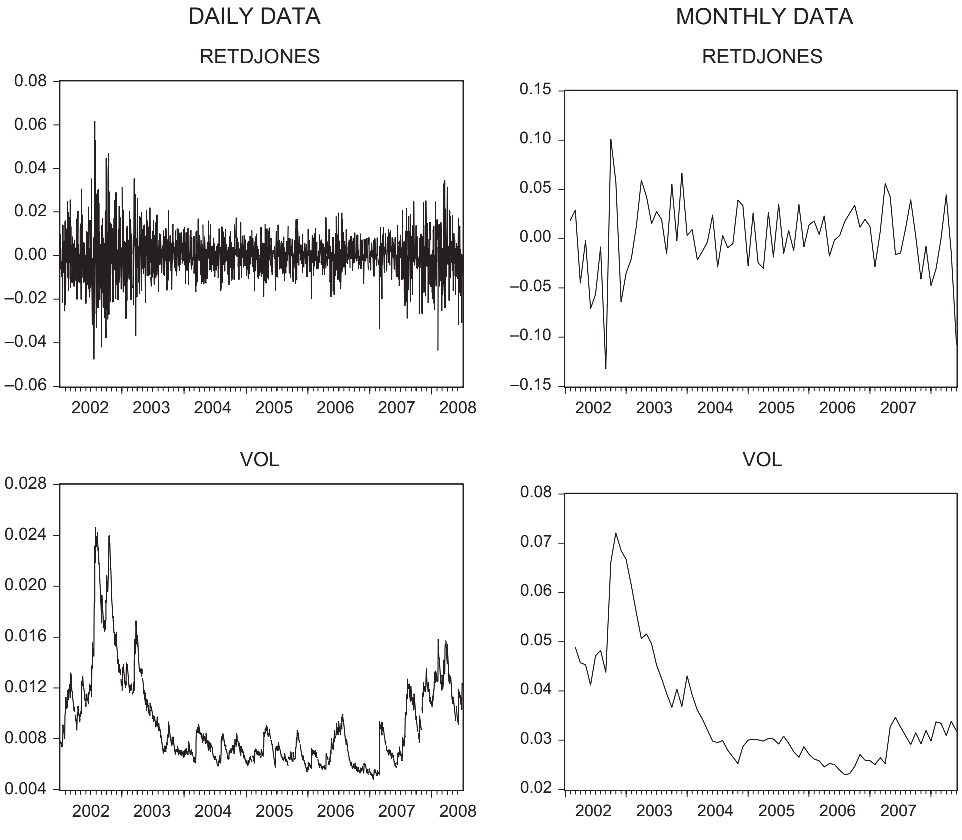

Figure 2 Return and volatility of the Dow Jones index.

cent level of significance, that is, the variables are first-difference stationary and do not co-integrate.

The parameters of the proposed model and the two-factor model were estimated with Kalman filter prediction using the equations in state-space form, and the results are shown in Table 1.

The results in Table 1 show that a positive relation exists between the lagged return of petroleum and the logarithm of the spot price of sugar using both daily and monthly data. Volatility also exhibits a positive relation, which can be attributed to the fact that sugar prices rise when volatility increases.

In terms of likelihood, the proposed model shows greater likelihood than the two-factor model. The estimated parameters show that the estimated mean of the annual long-term rate of convenience yield is 11.65 and 14.31 per cent for daily and monthly data, respectively. The estimated mean rate of convenience yield above the average federal funds rate, at 2.81 per cent, is consistent with the presence of backwardation in the sample. The results derived from the monthly data also verify that the mean-reversion coefficient ϕ is substantially greater in the proposed model than in the two-factor model.

Series of forecasts n steps ahead were generated and were then compared using the following error-prediction methods between the observed values S_o and the predicted values S_f:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{N} \left(S_{ot} - S_{ft}\right)^2}{N}} \tag{21}$$

$$\mathrm{MAE} = \sum_{j=1}^{N} \frac{\left|S_{ot} - S_{ft}\right|}{N} \tag{22}$$

$$\mathrm{MAPE} = \frac{1}{N} \sum_{j=1}^{N} \left|\frac{S_{ot} - S_{ft}}{S_{ot}}\right| \tag{23}$$

The first metric, RMSE, is the root mean square error; the second metric, MAE, is the mean absolute error; and the third metric, MAPE, is the mean absolute percentage error. In this step, the objective was to compare the two models' forecast error based on the difference between the actual sugar price published by CEPEA and the price predicted by the two models both 1 and 3 months in advance, as shown in Table 2.

The data in Table 2 demonstrate that the proposed model is superior in terms of predictive ability. Using all three metrics, the proposed model produced smaller forecast error relative to the two-factor model when using both daily and monthly data. Monthly data resulted in less error than daily data. The data also show that the longer the forecast time horizon, the more effective the proposed model is relative to the two-factor model. The gains from the proposed model in error reduction for a 3-month time horizon in terms of the decrease in the MAPE were 7.59 and 10.79 per cent for daily and monthly data, respectively.

Table 1 Parameter estimates from the filtering process

Parameter	Daily data		Monthly data	
	Proposed model	Two factors model	Proposed model	Two factors model
ϕ	0.008398	0.008339	1.112879	0.124878
μ_δ	0.116495	0.111897	0.143086	0.083884
$\exp(\sigma_\delta)$	−6.556159	−6.588269	−39.29111	−39.29111
β_0	0.000311	0.000254	0.011527	0.006749
β_1	0.022868	–	0.049383	–
β_2	1.069168	–	1.187362	–
$\beta_{c,1}$	0.001367	–	0.002386	–
$\beta_{s,1}$	−0.001836	–	−0.035554	–
$\exp(\sigma_S)$	−8.686838	−8.661802	−4.934495	−4.793600
Log likelihood	4555.53	4548.11	64.31	60.32

Table 2 Results of forecast error

Model	Sample periodicity	Forecast (months)	RMSE	MAE	MAPE
Proposed model	Daily	1	1.5877	1.2202	0.1169
	Daily	3	2.2243	1.6887	0.1656
Two-factor model	Daily	1	1.6198	1.2553	0.1191
	Daily	3	2.4285	1.8093	0.1792
Proposed model	Monthly	1	1.6127	1.2045	0.1125
	Monthly	3	2.1638	1.5659	0.1477
Two-factor model	Monthly	1	1.6476	1.2431	0.1131
	Monthly	3	2.2769	1.6996	0.1655

MAE, mean absolute error; RMSE, root mean square error; MAPE, mean absolute percentage error.

Beyond error analysis, the predictive ability of the two models was statistically compared using the Granger and Newbold (1976) method, with the null hypothesis being that the two models are equal in terms of predictive error. For this to be true, their errors must not be correlated. If the proposed model has greater predictive error, the correlation will be positive; if the two-factor model has greater predictive error, the correlation will be negative. Table 3 shows the results of a two-tailed t-test for the two models using both daily and monthly data and 1- and 3-month time horizons (P-values are in parentheses):

The test results indicate that the values of the correlation are negative or that in all the samples evaluated the proposed model is superior, in terms of predictive ability, to the two-factor model. In addition, the null hypothesis that the two models are equal in terms of predictive ability can be rejected with 99.9 per cent certainty based on daily data with a prediction horizon of 1 and 3 months. When using monthly data, the null hypothesis can be rejected for 3-month predictions but not for 1-month predictions.

Table 3 Results of the Granger–Newbold test

Data	Forecast (months)	Degrees of freedom	Error correlation	t-statistic
Daily	1	1606	−0.1024	−4.1261 (0.0000)
	3	1564	−0.9504	−120.8194 (0.0000)
Monthly	1	74	−0.1138	−0.9853 (0.3277)
	3	71	−0.5022	−4.8941 (0.0000)

6. Conclusion

The proposed pricing model for agricultural commodities related to the production of biofuel is based on the interdependence of commodity prices and petroleum, theory of storage, the seasonality of harvests and the influence of market volatility. The premises of the model suggest that prices may undergo persistent path changes that can be explained by observable exogenous variables.

The proposed model is built upon a system of equations using space-state analysis. The first equation defines the no-arbitrage condition between the spot price and the future price of sugar. The model's second equation represents the mean-reversion process of the convenience yield and the third equation defines the trajectory of the spot price of sugar as described by generalised Brownian motion. The process that describes the spot price has a nonconstant deterministic trend and therefore is less restrictive than the other models in the literature.

The model was applied in the Brazilian sugar market, with predictions made using the Kalman filter. The results were then compared with those generated by the two-factor model to verify the proposed model's predictive capacity. One- and 3-month predictions were compared with the actual sugar prices published by CEPEA. The results indicate that the proposed model generates predictions that are statistically superior to those produced by the two-factor model regardless of whether daily or monthly data are used. In addition, the results demonstrate that the longer the predictive time horizon, the more accurate the proposed model is relative to the two-factor model. The gains from the proposed model in error reduction for a 3-month time horizon were >10 per cent when using monthly data.

The results suggest that the proposed model of commodity-price formation with a deterministic trend described by exogenous variables is superior in terms of predictive ability when applied in the Brazilian sugar market.

References

Benirschka, M. and Binkley, J.K. (1995). Optimal storage and marketing over space and time, *American Journal of Agricultural Economics* 77, 512–524.

Brennan, M.J. (1958). The supply of storage, *The American Economic Review* 48, 50–72.

Brennan, M.J. and Schwartz, E.S. (1985). Evaluating natural resource investments, *The Journal of Business* 58, 135–157.

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74, 427–431.

Dixit, A.K. and Pindyck, R.S. (1994). *Investment Under Uncertainty*. Princeton University Press, Princeton.

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

Fama, E.F. and French, K.R. (1987). Commodity futures prices: some evidence on forecast power, premiums, and the theory of storage, *The Journal of Business* 60, 55–73.

Frechette, D.L. and Fackler, P.L. (1999). What causes commodity price backwardation?, *American Journal of Agricultural Economics* 81, 761–771.

Gibson, R. and Schwartz, E.S. (1990). Stochastic convenience yield and the pricing of oil contingent claims, *The Journal of Finance* 45, 959–976.

Granger, C.W.J. and Newbold, P. (1976). Forecasting transformed series, *Journal of the Royal Statistical Society* 38, 189–203.

Hannan, E.J., Terrell, R.D. and Tuckwell, N.E. (1970). The seasonal adjustment of economic time series, *International Economic Review* 11, 24–52.

Harri, A., Nalley, L. and Hudson, D. (2009). The relationship between oil, exchange rates, and commodity prices, *Journal of Agricultural and Applied Economics* 41, 501–510.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University, Cambridge.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.

Johnson, L.L. (1960). The theory of hedging and speculation in commodity futures, *The Review of Economic Studies* 27, 139–151.

Kaldor, N. (1939). Speculation and economic stability, *The Review of Economic Studies* 7, 1–27.

Litzenberger, R.H. and Rabinowitz, N. (1995). Backwardation in oil futures markets: theory and empirical evidence, *The Journal of Finance* 50, 1517–1545.

Schwartz, E.S. (1997). The stochastic behavior of commodity prices: implications for valuation and hedging, *The Journal of Finance* 52, 923–973.

Schwartz, E.S. and Smith, J.E. (2000). Short-term variations and long-term dynamics in commodity prices, *Management Science* 46, 893–911.

Sorensen, C. (2002). Modeling seasonality in agricultural commodity futures, *The Journal of Futures Markets* 22, 393–426.

Telser, L.G. (1958). Futures trading and the storage of cotton and wheat, *The Journal of Political Economy* 66, 233–255.

Working, H. (1948). Theory of the inverse carrying charge in futures markets, *Journal of Farm Economics* 30, 1–28.

Working, H. (1949). The theory of price of storage, *The American Economic Review* 39, 1254–1262.