Jumping Beans: Implications of Fat Tails in International Soybean Markets

by

Charles F. Mason†, Luca Taschini‡ and Neil Wilmot§

Abstract: Several recent policies have been promulgated to reduce reliance on fossil fuels in the the United States (US) transportation sector. To achieve these ambitious goals, it seems highly likely that US refiners will have to accommodate significant inflows of soybeans imported from Brazil; important large-scale (irreversible) investments will also be required. These investments are subject to substantial uncertainty, underscoring the importance of characterizing the stochastic nature of soybean prices. In this paper we investigate the potential presence of jumps in two key prices: the spot price for soybeans, in Brazil, and ethanol produced from soybeans, also in Brazil. We find compelling empirical evidence for the importance of jumps in both markets. The presence of jumps in these markets has important implications for large scale infrastructure investments, as would be necessary to produce ethanol-based motor vehicle fuels.

Keywords: soybean prices; ethanol prices; jump diffusion; GARCH
JEL codes: Q30, Q40, C58

this draft: March 20, 2023

†Corresponding author: H.A.True Chair in Petroleum and Natural Gas Economics, Department of Economics, University of Wyoming, 1000 E. University Ave., Laramie, WY 82071; phone: 1-307-766-5336; fax: 1-307-766-5090; e-mail: bambuzlr@uwyo.edu

‡Business School, University of Edinburgh

§Department of Economics, University of Minnesota Duluth and Associate Faculty, Institute on the Environment, University of Minnesota Twin Cities.

1. Introduction

In the past decade or so, a number of policies have been promulgated at the federal and state levels to move the United States (US) economy towards less reliance on fossil fuels. A number of these policies focus on the transportation sector. At the federal level, for example, the Renewable Fuel Standard (RFS) mandated that 36 billion gallons of renewable fuels (per year) are in use by 2022 (U.S. Energy Information Administration, 2013). At the state level, California has adopted the "Low Carbon Fuel Standard" (LCFS), which requires a 10% reduction in the carbon intensity of motor vehicle fuels by 2020. Both policies are likely to increase reliance on biofuels, both corn- and soybean-based. To facilitate the goals under the RFS, the US created "Renewable Identification Numbers" (RINs), which are essentially tradable certificates for producers of inputs into renewable fuels.

A variety of structural elements in the market for RINs complicate the expansion needed to meet the growing demand for ethanol associated with the LCFS and RFs. These elements include the relative immaturity of the RINs market, the presence of the "blend wall",[1] the large distances between refiners and major production basins for agricultural products such as corn that are used to create ethanol (LaRiviere et al., 2015); and diseconomies induced by the competition between fuel and food uses for products such as corn. In addition, there are concerns about the indirect carbon emissions that would arise from the requisite conversion of land into domestic corn production in the US (Searchinger et al., 2008).

One resolution of these difficulties would be to shift refiners' reliance from corn products as inputs in the ethanol production process to soybeans.[2] Accordingly, to achieve

[1] The blend wall refers to the point at which fuels contain 10% ethanol; it is believed that conventional internal combustion engines cannot function normally when fuels contain more than 10% ethanol. See Babcock (2013); Burkholder (2015); Knittel et al. (2017) and Meiselman (2016) for discussion.

[2] For example, the penalty assessed by California Air Resources Board on corn produced in the US

the ambitious goals of the LCFS, it seems highly likely that California refiners will have to accommodate significant increases in ethanol produced from soybeans; most likely, this will in turn require large inflows of soybeans and ethanol imported from Brazil. Indeed, Morrison and Chen (2011) argue that Brazilian ethanol could account for 25% of all transportation energy in California in the coming years.

An additional consideration is that the market for RINs has been shown to exhibit significant transitory shocks or jumps, and that RINs prices follow a more complex process than geometric Brownian motion (GBM) (Mason and Wilmot, 2016). As such, the distribution of the log-returns of RINs prices have substantially fatter tails than does a Normal distribution. The presence of fat tails has important implications for the incentives to invest in capital projects linked to renewables (Mason and Wilmot, 2022). To the extent that such investments are at least partially asset-specific to renewables, they reflect a sunk (or partially sunk) up-front cost; fat tails can delay investment in the presence of partially or fully irreversible investments (Martzoukos and Trigeorgis, 2002). This aspect of RINs prices, combined with the significant infrastructure that will have to be deployed to fully capitalize on the potential role of soybeans, raises questions regarding fatness of tails in soybean price returns as well.

Our goal in this paper is to analyze price returns for Brazilian soybeans, so as to determine the empirical importance of elements that might contribute to fat tails. To this end, we first describe an extension of the familiar model of a stochastic process that allows for unexpected changes, or jumps. This extension leads naturally to an econometric specification, which can be readily combined with time-varying volatility (also known as the generalized autoregressive conditional heteroscedasticity, or GARCH, framework). After incorporating these elements, we characterize the likelihood function that governs

implies that ethanol produced from Brazilian crops is less carbon-intensive than is ethanol produced from US corn. Likewise, using a life-cycle (well-to-wheel) analysis, Zhang et al. (2010) present results that suggest Brazilian ethanol could result in 18-33% lower emissions than US based corn ethanol.

the data generating process; this, in turn, leads directly to an estimation procedure and hypotheses tests regarding the appropriate specification of the stochastic process. We then apply this econometric methodology to times series for Brazilian spot prices for soybeans. Related to this time series are ethanol prices in Brazil, as this fuel is largely dependent upon Brazilian soybeans. Our data is based on daily observations, for both spot prices. We compare four stochastic data-generating processes: GBM (which we refer to as PD in the pursuant discussion), GBM allowing for a jump diffusion process (which we refer to as JD in the pursuant discussion), GBM allowing for GARCH (which we refer to as GPD in the pursuant discussion), and GBM allowing for both GARCH and a jump diffusion process (which we refer to as GJD in the pursuant discussion). Our findings generally point to the statistical importance of allowing for both GARCH and jumps, for both spot prices.

As both GARCH and jumps will induce fat tails, our empirical results may have important implications for motives to undertake large-scale investments such as import facilities where these products could be offloaded, facilities that would convert soybeans into ethanol once they have reached American shores, and refinery adaptations that are likely to be required so as to accommodate these new fuel sources.

2. Econometric Framework

In order to develop the maximum likelihood framework used to estimate the parameters of the different models, we begin with a brief examination of the stochastic processes under investigation. Let P_t denote price at time t; its time path is said to follow a geometric Brownian motion (*GBM*) process with trend α and variance parameter σ if[3]

$$dP_t = \alpha P_t dt + \sigma P_t dz. \tag{1}$$

In equation (1), dz represents an increment of a Wiener process $dz = \xi_t \sqrt{dt}$, where ξ_t has zero mean and a standard deviation equal to 1 (Dixit and Pindyck, 1993). Denote the log returns, *i.e.*, the natural logarithm of the ratio of price in period t to the price in period $t-1$, by $x_t \equiv \ln(P_t/P_{t-1})$. If P_t follows a *GBM* process then x_t is normally distributed with variance σ^2 and mean $\mu \equiv \alpha - \sigma^2/2$. This gives the pure diffusion (*PD*) model

$$x_t = \mu + \sigma z_t. \tag{2}$$

The term z_t in equation (2) is an identically and independently distributed (i.i.d.) random variable with mean zero and variance one.

We introduce jumps into the model in the style of Merton (1976), by assuming that two types of changes affect the log returns. The first type are 'normal' fluctuations, represented through the geometric Brownian motion process. The second type, 'abnormal' shocks, are modeled through a discontinuous process. These abnormal shocks can be thought of as occurring via the arrival of new information (Elder et al., 2013). We view these shocks as transitory, as opposed to quasi-permanent changes in the fundamental underlying structure of the market. This assumption makes it more natural to include a jump process, as opposed to a regime shifting framework. We assume the discontinuities are described by a Poisson distribution governing the number of discrete-valued events, $n_t \in \{0, 1, 2, ...\}$, that occur over the interval $(t-1, t)$; accordingly, the probability that j jumps

[3] Engel et al. (2015) use a similar approach to model soybean returns when studying how uncertainty in alternative land-use returns influences the decision of whether or not to deforest.

are observed during this interval equals

$$P\left(N_t = j\right) = \frac{\exp\left(-\lambda\right)\lambda^j}{j!}. \tag{3}$$

A key element in equation (3) is λ, which can be interpreted as the probability of observing a jump in any brief time interval of length dt. Thus, the arrival of jumps is a Poisson distribution,[4] from which we can describe the change in the number of jumps observed by

$$dn_t = \begin{cases} 0 & \text{with probability } 1-\lambda dt \\ 1 & \text{with probability } \lambda dt \end{cases} \tag{4}$$

As in Askari and Krichene (2008), when abnormal information arrives at time t, prices jump from P_{t-} (the limit as the time index tends towards t from left) to $P_t = \exp\left(J_t\right)P_{t-}$; accordingly, J_t measures the percentage change in price. The resultant stochastic process for the random variable P_t may then be written as

$$\frac{dP_t}{P_t} = \alpha dt + \sigma dz_t + \left(\exp\left(J_t\right) - 1\right)dn_t, \tag{5}$$

where dz_t has the same properties assumed in equation (1) and dn_t is the independent Poisson process described in equation (4). Together the terms dz_t and dn_t make up the instantaneous component of the unanticipated return. It is natural to assume these terms are independent, since the first component reflects ordinary movements in price while the

[4] One could of course use alternative specifications of the jump process, including Bernouilli or Levy. Our choice is motivated by the ability to combine the Poisson process – along with a GARCH process – into the basic PD econometric model. One advantage of our approach is that it leads to a relatively straightforward extension of the analytics associated with evaluating optimal investment; for example, Dixit and Pindyck (1993, p. 171) show that including a Poisson process into a conventional Brownian motion framework adds only one (non-linear) term to the key equation that defines the optimal value function associated with investing. Note too that we do not specify jump events *ex ante*, but rather let the econometric results pick out the key parameters. An alternative would be to use some criterion to decide when a jump has occurred, as in Chevallier and Sévi (2014).

second component reflects unusual changes in price. The size of the jump, $Y_{t,k}$, is itself a random variable; we assume it is normally distributed with mean θ and variance δ^2, and that it is independent of the distribution for the arrival of a jump. The jump component affecting returns between time t and time $t+1$ is then

$$J_t = \sum_{k=0}^{n_t} Y_{t,k}. \tag{6}$$

Thus, the mixed jump-diffusion (*JD*) process for the log-price returns can be described by

$$x_t = \mu + \sigma z_t + J_t. \tag{7}$$

An alternative explanation for the "fat tails" that are often observed in commodity price data is that P_t is subject to time-varying volatility. An example of such a phenomenon is the "generalized autoregressive conditional heteroskedastic" (GARCH) framework. Adapting the pure diffusion model to allow for this form of time-varying volatility gives the GARCH – diffusion (*GPD*) process:[5]

$$x_t = \mu + \sqrt{h_t} z_t, \tag{8}$$

where the conditional variance, h_t is described by the process

$$h_t \equiv E_{t-1}\left(\sigma^2\right) = \kappa + \alpha_1 (x_{t-1} - \mu)^2 + \beta_1 h_{t-1}. \tag{9}$$

Note that when $h_t = \sigma^2$ the GARCH diffusion model reduces to pure diffusion model. On the other hand, when $\kappa > 0$ and $\alpha_1 + \beta_1 < 1$, the unconditional variance of the volatility of

[5] The process described in equations (8)–(9) is characterized by four parameters, μ, κ, α_1 and β_1. There is a general consensus in the literature is that a GARCH model with a limited number of terms performs reasonably well, and so we restrict our focus to this more parsimonious representation.

the process exists and equals $\frac{\kappa}{1-\alpha_1-\beta_1}$.

Allowing for jump discontinuities would result in the GARCH(1,1) jump-diffusion (*GJD*) process:

$$x_t = \mu + \sqrt{h_t} z_t + J_t, \tag{10}$$

where h_t is described by equation (9). Duan (1997) shows that the diffusion limit of a large class of GARCH(1,1) models contain many diffusion processes allowing the approximation of stochastic volatility models by the GARCH process.

We evaluate the four models using maximum likelihood estimation methods.[6] To this end, we note that the parameters of our four candidate models – *PD, JD, GPD, GJD* – may be nested into the general log-likelihood function

$$L\left(\phi, x_t\right) = -T\lambda - \frac{T}{2}\ln(2\pi) + \sum_{t=1}^{T} \ln\left[\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \frac{1}{\sqrt{h_t + n\delta^2}} \exp\left(\frac{-(x_t - \mu - n\theta)}{2(h_t + n\delta^2)}\right)\right], \tag{11}$$

where n indexes the number of jumps, combined with the description of h_t given in equation (9).[7] In this framework, the *GPD* model corresponds to the parameter restriction $\lambda = \theta = \delta = 0$; the *JD* model corresponds to the restriction $\alpha_1 = \beta_1 = 0$; and the *PD* model corresponds to the restriction $\alpha_1 = \beta_1 = \lambda = \theta = \delta = 0$. Comparing any pair of potential models can thus be framed as a test of an appropriate parameter restriction. For example, the comparison of the *PD* and *GPD* models is conducted by testing the parameter restriction $\alpha_1 = \beta_1 = 0$; the comparison of the *PD* and *JD* models is conducted by testing the parameter restriction $\lambda = \theta = \delta = 0$. The empirical validity of the parameter restriction of interest can be evaluated by use of the likelihood ratio test (Johnston and DiNardo, 1997). This approach compares the likelihood function under a particular restriction, $L(\phi^R; x)$,

[6] Maximum likelihood estimates are known to be consistent and invariant with asymptotically normal distributions of the parameters.

[7] In the empirical results we report below, the number of jumps was truncated at 10 (Ball and Torous, 1985).

to that of the unrestricted or less restricted likelihood function, $L\left(\hat{\phi};x\right)$. Under the null hypothesis that the restriction is empirically valid, the decrease in the likelihood function associated with the restriction will be small. Such an approach can be used to make pairwise-comparisons between a more general model and a more restricted model. The test statistic is the log-likelihood ratio

$$LR = 2[L(\hat{\phi};x) - L(\phi^R;x)];$$

under the null hypothesis this statistic will be distributed as a Chi-square random variable with m degrees of freedom, where m is the number of parameter restrictions.

3. Data and data properties

The discussion in the Introduction motivates us to evaluate soybean and ethanol prices in Brazil; because the former might be thought of as a substitute to American corn (as an input into ethanol production), we also evaluate US corn prices. The data for this study consists of the daily closing prices of Brazilian soybeans and ethanol, and US corn.

Both Brazilian soybean prices and ethanol prices were obtained from The Centro de Estudos Avancadoes em Economia Aplicada (CEPEA). CEPEA Brazilian soybean prices are reported as daily present cash value equivalents in US dollars per 60-kilogram bag. Brazilian fuel ethanol prices are reported as daily present cash value equivalents in US dollars per cubic meter. These prices are retrieved from the CEPEA website. Both these prices are retrieved from the CEPEA website.[8] US corn prices were obtained from Bloomberg, and represent the front month corn futures prices based on the 5,000-bushel contract traded on the CME.

Summary statistics, including the first four moments (mean, variance, skewness

[8] The data are available at CEPEA soja and ethanol websites. We discuss the process used to construct the Brazilian soybean and ethanol data series in greater detail in the Appendix.

and kurtosis) for daily prices and log returns of each of the time series are given in Table 1. The price returns are calculated as

$$r_t = 100[ln(P_t/P_{t-1})].$$

In Figures 1–3, we plot the price returns for the three time series. Brazilian soybean returns are shown in Figure 1, corn returns are shown in Figure 2 and Brazilian ethanol returns are shown in Figure 3. The soybean series displays much lower variation, relative to the corn and ethanol series. Each series also displays evidence of asymmetry in the distribution, as displayed by the presence of skewness. Each series also displays evidence of leptokurtosis or "fat-tails" by the large value for kurtosis. The Anderson – Darling test, a quadratic empirical distribution function (EDF) test, is used to examine the normality of the data. The results of the test imply the null hypothesis of a normally distributed random variable is strongly rejected for each of our time series.

These results are corroborated by the "quantile–quantile" plots, which we present in Figures 4. Figure 4(a) shows the natural log of soybean returns, Figure 4(b) shows the natural log of corn returns and Figure 4(c) shows the natural log of natural log of ethanol returns. If soybean prices follow a geometric Brownian motion process, then the soybean prices would be log-normally distributed (*i.e.*, the natural log of the soybean returns would be Normally distributed). A quantile–quantile plot compares the values observed in the empirical distribution (measured on the y-axis) against the values from the inverse of a theoretical normal distribution whose mean and standard deviation correspond to the values associated with the empirical distribution (measured on the x-axis). If the empirical distribution of the natural log of soybean returns is close to a normal distribution, the quantile–quantile plot will be well described by a straight line. Alternatively, if there are significant departures from a linear relation, then the natural log of the soybean returns is not well-described by a normal distribution, arguing against the empirical validity of the geometric Brownian motion specification. Here, we see consistent departures from a linear relation, particularly in the tails. These departures indicate significant leptokurtosis, *i.e.*, fat tails.

4. Econometric Results

The results of the maximum likelihood estimate of the four stochastic processes (*PD, JD, GPD, GJD*) for each of the commodities are presented in Table 2. Incorporating a jump component into the model (*JD*) noticeably reduces the instantaneous rate of variance, σ, across all three commodities (soybeans, ethanol and corn). Such reductions are offset by the a large and significant value of the variance of the jumps, δ. The intensity of the jump process, λ, is significant across the three commodities. The coefficients in the soybean markets suggest that jumps occur, on average, quite frequently, while jumps occur less frequently in the ethanol market. Though insignificant, the mean jump size, θ, suggests that soybeans and ethanol returns tend to experience negative jumps. This is in contrast to the corn market, where a positive and significant θ indicates that the market tends to experience positive shocks on average.

The GARCH(1,1) model (*GPD*) provides variance parameter estimates that are significant and indicate a high degree of persistence ($\hat{\alpha} + \hat{\beta}$ is close to 1), a common feature of financial time series. The value of $\hat{\beta}$ suggests the effect of changes in volatility on future volatility will persist for a longer period of time, as the rate of decay is slower. In the mixed jump-diffusion model (*GJD*), the jump intensity λ remains significant though smaller in magnitude than in the *JD* model. This indicates that the *GJD* model predicts less frequent jumps relative to the *JD* model. Even so, while allowing for GARCH evidently captures some of the estimated effect of the jump in the *JD* model it does not render jumps irrelevant. Furthermore, the estimated frequency of jumps is economically meaningful: The results of the *JD* model suggest that the soybean (ethanol) price return series experiences a jump approximately every 2.7 (5.6) days, while the *GJD* suggests a jump occurs approximately every 8.1 (19.6) days.

The results of the pairwise Likelihood ratio (*LR*) tests are presented in Table 3. Each entry in the Table is a test statistic of a hypothesis X vs. Y, where the null hypothesis is that X is the appropriate stochastic process describing the data and the alternative hypothesis is that Y is the appropriate stochastic process describing the data. The parenthetical values below each test statistic give the associated p-value. For all three price returns, the results displayed in column two show that allowing for time-varying volatility improves model fit, relative to the pure-diffusion model. Likewise, the results displayed in column three indicate that allowing for jumps yields a statistically important increase in predictive power, relative to the pure-diffusion model, for each price return series. The results in the final two columns indicate that allowing for both jumps and time-varying volatility improves model performance. The results in column four indicate that incorporating time-varying volatility into a model that allows for jumps yields a statistically important improvement in model fit, for each commodity. Similarly, the results in column five show that incorporating jumps into a model that allows for time-varying volatility yields a statistically important improvement in model fit – again, for each commodity. The take-away message is that in every case, and for each of the three commodities, the more elaborate model is preferred to the less elaborate model. These conclusions hold with considerable confidence: the chance that the null hypothesis (of the simpler model) holding true is less than 1% in every case. As such, the test results point to a statistically important gain in predictive power associated with allowing for both jumps and time-varying volatility.

We conclude this section by providing some additional evidence on the desirability of the *GJD* model, based on a pseudo-forecasting analysis. To this end, we split our sample in half, and re-estimated the *PD* and *GJD* models, using the first half of the sample for each of the three commodity price returns. Then, employing the resulting point estimates, we stochastically simulated the associated price returns, using observations from the second half of the sample – again, for each of the three commodity price returns. We then calculated the total prediction errors and the summed squared prediction errors for both the *PD* and *GJD* models, for each of the three commodity price returns. The resultant values are collected in Table 4, along with the ratio of total prediction errors (in the top half of the table) as well as the ratio of summed squared errors (in the bottom half of the table) for the *GJD* model to the *PD* model.

While our intent here is to offer a descriptive analysis, the results are encouraging. For each commodity, the *GJD* model outperforms the *PD* model, in the sense that it delivers a lower magnitude of total prediction errors as well as lower summed squared prediction errors. The improvement in forecasting accuracy is particularly pronounced for the soybean and ethanol series, where the *GJD* model dramatically lowers the magnitude of prediction errors – by close to 89% for ethanol and close to 93% for soybeans. While less dramatic, the *GJD* model also reduces summed squared prediction errors for both commodities, by 5-10%. The improvement for the corn price returns is less dramatic: prediction errors fall by a bit more than 17% comparing *GJD* to *PD*, but summed squared errors are virtually unchanged. One potential explanation for the difference between these two sets of results is that prices of the former commodities are based on markets in Brazil, while the latter commodity is based on prices in the U.S. Perhaps the Brazilian commodity markets are more susceptible to factors that facilitate price jumps – conceivably because the U.S. market is thicker.

5. The influence of jumps on investment under uncertainty

In this section, we investigate the potential impact of including jumps in the stochastic specification of the price for a key commodity. To illustrate the basic ideas, we start with a conventional investment under uncertainty problem, under which the key underlying stochastic process is geometric Brownian motion Dixit and Pindyck (1993). In the present application this underlying variable would be the price of a key commodity input, such as soybeans or corn, or the price of an intermediate good such as ethanol. The investment problem involves a one-time sunk expenditure K; making this expenditure allows the decision-maker to obtain a new payoff flow. The investment could reflect expanding refinery capacity to process increased inflows of ethanol, or building a dedicated factory for biofuels. A key question here regards timing: when should the investment be taken? Answering this question requires a determination of the value associated with forestalling the investment – the "option value of waiting" – together with a determination of the value of investment.

We assume that the benefits associated with investing at a certain time t are proportional to the price of the key resource at that time.[9] This implies the benefits associated with investing at time t can be expressed by a stochastically evolving component, which we write as X_t. This construct could be the price of an important commodity (for example a fossil fuel whose role will change following adoption of the policy) or the value of some related financial instrument (for example, a carbon permit). Letting K denote the one-time investing cost, the net benefits of acting (investing) at t are equal to[10]

$$X - K.$$

[9] This implicitly assumes the quantity delivered is fixed, *i.e.* supply is perfectly inelastic. More generally, an upward-sloping supply curve would induce quantity as a function of price. Adapting the model to allow for such a structure is feasible, but at the cost of considerable extra complexity. See (Dixit and Pindyck, 1993, pp. 195-199) for discussion.

[10] In the pursuant discussion, we will often suppress the time subscript so as to reduce notational clutter.

These net benefits are compared against the value associated with the option value of waiting. Delaying investment can be beneficial, since the return from the investment is linked to the stochastic value X. At any time, there is a chance that X will evolve downwards, rendering the investment uneconomic; accordingly, choosing to invest at the precise moment when anticipated net benefits first become positive is ill-advised. By delaying, the decision-maker reduces the chance that s/he will regret making the investment; the increase in value associated with waiting to build at the optimal time in the future is the option value associated with waiting.

The option value is functionally related to the stochastically evolving component through the optimal value function $F(X)$. We start by working through the problem when X follows a geometric Brownian motion (GBM) process. Later, we discuss the determination of $F(X)$ when X is also subject to the potential for jumps.[11]

Under GBM, one can express the stochastic evolution of X as in eq. (1). At any moment where the decision to undertake the investment has yet to be made there are two possible decisions: either build now or wait. The decision to build now yields the immediate payoff $X-K$ (as noted above). The decision to wait earns a flow payoff of zero (since nothing has been done), while the option value, $F(X)$, is retained; delay will deliver anticipated change in $F(X)$ (which can be thought of as the anticipated capital gains) less the foregone capitalized option value (which can be thought of as the interest earned on the net returns). If delaying is optimal, the fundamental equation of optimality requires that these two effects balance out Dixit and Pindyck (1993), so that the optimal value function must satisfy:

$$\rho F(X) = \frac{1}{\mathrm{dt}} E\Big[\mathrm{d}(F)\Big], \tag{12}$$

[11] One aspect of the GBM process is that changes tend to exert an effect for a considerable length of time. An alternative approach would be to use a model in which the effect of changes in X tend to dissipate relatively more rapidly – for example, a mean-reverting process. Analysis such a process is more complicated, though the broad principles we describe in this section still apply Dixit and Pindyck (1993).

where ρ is the decision maker's discount rate and the expression on the right-hand side is the so-called Itô operator. The left-hand side of eq. (12) measures the capitalized option value, while the right-hand side is the anticipated capital gains. It can be shown that the solution to this equation takes the form (Dixit and Pindyck, 1993; Mason and Wilmot, 2022):

$$F(X) = aX^{\beta}, \tag{13}$$

where $\beta > 1$ depends positively on σ and negatively on α.

The value function $F(X)$ can be interpreted as the value of an option to invest in the future Dixit and Pindyck (1993). Accordingly, it is optimal to invest when this value equals the net benefit from acting now; this implies a cutoff value X^* for the underlying stochastic ingredient, which is implicitly defined by the "value-matching" condition

$$F(X^*) = X^* - K \tag{14}$$

along with the "smooth-pasting" condition

$$F'(X^*) = d(X^* - K)/dX^* = 1. \tag{15}$$

Applying the value-matching and smooth-pasting conditions to the functional form in eq. (13), it is easy to show that the cutoff value is:

$$X^* = \frac{\beta K}{\beta - 1}. \tag{16}$$

As noted above, β is increasing in σ and decreasing in α; it follows that X^* is also is increasing in σ and decreasing in α.

Since investment is delayed until X rises to this cutoff value, investment will tend to be undertaken sooner the larger is α or the smaller is σ. These features can also be characterized in terms of the option value. Because a larger option value raises the benefits from delay, it will tend to push back in time the moment at which the decision to invest is taken. Intuitively, an increase in the variance of the stochastic process raises the option value because of the potential for a more dramatic future increase in the underlying value X; delaying investment allows the decision maker to strategically take advantage of such future movements. This effect is more important the larger is the initial investment K.

Now suppose the value X evolves according to the mixed jump-diffusion process. Here, we assume changes in X are composed of two types of changes: 'typical' fluctuations, represented through the GBM process, and 'abnormal' fluctuations, due to the arrival of new information or some unusual event. We model the arrival of these abnormal fluctuations as following a Poisson process.[12] Letting n_t denote the number of such events that have occurred as of time t, the change in n_t during the interval $(t, t+\Delta t)$ is described by

$$dn_t = \begin{cases} 0, & \text{with probability } 1-\lambda dt \\ 1, & \text{with probability } \quad \lambda dt, \end{cases} \tag{17}$$

where $\lambda > 0$ is a parameter measuring the arrival frequency

We denote the size of a jump at time t, should one occur, is J_t. We assume the jump size is independently and identically distributed as a lognormal random variable, so that $\ln(J)$ is Normally distributed. As we discussed above, the resultant stochastic process for the random variable X is given in eq. (5). It can be shown that allowing for jumps changes the drift term in the expressions for the evolution of X to $\alpha + \lambda\theta$; an important related point is that incorporating jumps will increase the variability of X over time.[13]

[12] Some authors model price jumps using a Lévy process, an approach that requires an <u>ex ante</u> definition of a jump. For example, Benth et al. (2008) define a jump as an observation that falls outside of 2 standard deviations from the mean. Other authors assume jumps follow a Poisson process; one advantage of this approach is that there is no need to arbitrarily define a jump <u>ex ante</u>.

[13] See Wilmot (2010) for discussion. An alternative explanation for the "fat tails" that are often observed

In this setting, the solution is determined by the interaction between jump size, Y, and continuation value, V. Unlike the GBM variant, however, this problem cannot be solved analytically. Accordingly, we employ numerical simulations in the pursuant discussion. To facilitate numerical simulations, we must first specify the discount rate ρ; the mean α and standard deviation σ of the GBM formulation; and the jump intensity λ associated with the Poisson process. In our baseline simulations, we set these parameters as $\rho = 0.02, \alpha = 0.04, \sigma = 0.2$, and $\lambda = 0.10$. The distribution governing Y, the magnitude of a jump (should it occur), is assumed to be lognormal – *i.e.*, $\ln(Y)$ is Normally distributed – with mean $\theta = 0$ and standard deviation $\delta = 1$.

For a given parameterization, we solve for the critical value associated with investing; the interpretation is that when the expected value from investing meets or exceeds this critical value, the investment will be taken. This critical value will correspond to the sum of the investment cost itself and the option value of waiting. The difference between the critical value and the requisite up-front investment may then be interpreted as the option value of waiting. We also calculate the ratio of the critical value to up-front investment cost.

Our first set of simulations investigates the role played by the jump intensity. As we noted above an increase in λ raises the variance of the stochastic process, which should increase the option value of waiting; delaying investment allows the decision maker to capitalize on such future movements. In this set of simulations we vary λ between 0 and 0.2, by increments of 0.05; results from this set of simulations are summarized in Figure 5. This figure displays the option value associated with delaying investment, for various levels of up-front investment (*i.e.*, K) across the possible values of λ. The first

for many energy commodity is that those prices are subject to time-varying volatility. An example of such a phenomenon is the "generalized autoregressive conditional heteroskedastic" (GARCH) framework. Adapting the pure diffusion model to allow for this form of time-varying volatility gives a GARCH – diffusion process, under which the component σ in eq. (1) is replaced by a time-varying component h_t, as in eq. (9).

feature we observe is that the option value of delaying investment rises as the amount of money that must be invested increases. This is intuitive: because larger investments require risking more money, the decision-maker is more cautious about undertaking the investment. In these simulations, the tendency to delay investment tends to be more pronounced as the probability of a jump increases: while largely insensitive to λ when the required investment is small, the option value of waiting does respond to increased jump intensity at larger investment levels.[14] Moreover, we note that the impact of increasing λ is most pronounced at small value of λ. In particular, the largest effect appears to occur when the probability of a jump occurring is increased from 0 to a positive value – *i.e.*, when one allows for the possibility of jumps. Indeed, this effect become ever-more important as the up-front cost is increased.

In the second set of simulations, we vary θ – the expected value of the (natural log of) jump size – allowing for values ranging from -0.2 to 0.2, by increments of 0.1 (leaving other parameter values as in the preceding simulation). In this way we consider cases where abrupt movements in prices are negative on average as well as cases where jumps are positive on average. The results from this simulation are presented in Figure 6. As in the first set of simulations, we note that option value of delaying the investment rises as the amount of money that must be invested increases. As we noted above, an increase in θ will raise the drift in the stochastic process. This induces conflicting effects on the option value of waiting: on the one hand, larger drift depresses the option value of waiting; on the other, larger values of θ raise the variance of X, and larger variance increases the option value of waiting. In the simulations we report here the former effect appears to be somewhat larger, though the net effect is small. Evidently, the average jump value exerts a less significant influence on the value of delaying investment than does the potential for

[14] One should not make too much of the seeming equivalence of option values at the smallest level of K: The numerical grid we employ in the solution algorithm is not sufficiently granular to detect differences between option values at small levels of K.

a jump in the first instance.

The third set of simulations we consider varies δ, the standard deviation of the jump size; here we consider values ranging from 0.5 to 1.5, by increments of 0.25. Results from these simulations are presented in Figure 7. As we noted above, raising the variance of the jump size pushes up the variance of the stochastic variable X, which induces an increase in the option value of waiting. Interestingly, while this effect is small when $\delta < 1$, it becomes more pronounced when $\delta > 1$. That is, while the option value of waiting is not particularly responsive to changes in the variance of jump size at smaller levels of that variance, the impact upon option value becomes significantly more pronounced when the variance of the jump size increases above unity. Indeed, variations in the potential size of the jump play an ever-larger role as the amount of money that must be invested increases. Again, this seems intuitive: when prices are subject to possible jumps with particularly large variation, the impact on the value of waiting increases to an ever-larger degree – generating an increasing motive to delay. That is, greater variation in jump sizes make waiting more attractive, and hence raise the option value at the optimal investment time.

6. Conclusion

Our goal in this paper is to re-examine the assumption that the relative price returns of key energy prices, such as those for commodities related to biofuels, can be modeled using a continuous time process. In particular, a key goal was the development of a more accurate understanding of the stochastic forces driving these spot prices. We draw several important conclusions from our analysis. For all three prices under consideration – soybeans, corn and ethanol – the data strongly suggest that allowing for jumps or time-varying volatility in natural gas price returns generates improved fit, relative to the pure diffusion model. Moreover, combining a process that allows for jumps with a GARCH process (GJD) outperforms all alternative stochastic processes. Thus, our results indicate that incorporating both time-varying volatility and jumps into empirical models of these spot prices improves predictive power; the sharper predictions that result from this improvement should be of clear benefit to market traders.

There are many reasons why a better understanding of the stochastic process driving soybean and ethanol prices would be useful. These energy resources can have important microeconomic effects, with commodity price risk having a potentially significant impact on profits in a variety of lines of business. Knowledge of the underlying stochastic behavior of these assets could aid in forecasting spot prices, with attendant reductions in risk exposure. Moreover, decisions to invest in important infrastructure can be improved by an enhanced understanding of the stochastic processes driving the prices of related resource (Mason and Wilmot, 2022). For example, the accuracy of a decision to significantly expand the for a refinery to handle ethanol infrastructure, or to process imported soybeans, will almost surely be improved by such enhanced understanding.[15] This is particularly true when the prices of imported soybeans or ethanol are subject to infrequent jumps, as our results indicate. For in this case, the underlying distribution of oil prices is "fat-tailed" or leptokurtotic, and fat tails can be particularly important if prices exert a non-linear marginal impact on the agent's profit flow (Weitzman, 2009).

The potential for jumps in soybean and ethanol prices is of more than academic interest, as jumps in these prices have implications for investment in biofuel capacity and in the requisite infrastructure needed to accommodate a meaningful increase in the use of vehicles than can capitalize on expanded ethanol supplies (*i.e.*, E85 vehicles).[16] To

[15] This observation is independent of any qualitative assessment of the social desirability of using soybean as as feedstock for the production of biofuels. Fargione et al. (2008) argue that Brazilian soybean based ethanol is not socially desirable if its production is facilitated by clearing Amazonian rainforest. The case for Brazilian ethanol is far more compelling if its production is facilitated by converting Cerrado (grasslands).

[16] Babcock (2013) argues that more stringent future RFS standards will require new investment in E85 infrastructure, and "[w]hen the [RFS] mandate is set at a level that is not easily met with existing infrastructure, then the incentive to invest in infrastructure is large." As we noted, this incentive is reduced when there is value to waiting to build, as when RINs prices are influenced by the presence of jumps.

the extent that there are jumps in these prices, biofuels producers with excess capacity might be able to cash in on unexpectedly high price returns. But as our simulation results showed, it is also true that jumps in the underlying commodity price induce an option value associated with delaying investment in increased capacity (Mason and Wilmot, 2016). Similarly, the presence of jumps implies an option value to waiting to add E85 fueling stations.

Other benefits accrue from the ability to better frame the underlying stochastic model in an investment under uncertainty framework, which we believe has real potential for evaluating important large-scale infrastructure investments such as refinery expansions or import/export terminals. Because such enhancements to transportation infrastructure may have far-reaching benefits, for example by facilitating gas movements to regions with larger demand, the welfare consequences of these investments may be substantial. The potential for substantial welfare implications of these investments underscores the importance of developing a better understanding of the stochastic process underlying biofuels prices, which in turn highlights the value of developing a more accurate empirical model to describe these prices.

7. Appendix: Details on Brazilian Data Processes

In this Appendix, we provide additional detail on the Brazilian soybean and ethanol data.[17] Brazilian soybean prices are related only to soy delivered in Parana port, either Delivered at Place (DAP), or in silos or other delivery mechanisms which are accessible to ships' loading apparatus, known as Free at Shipside (FAS) delivery. All prices are converted to present values; specifically, futures contracts are converted to cash value based on the time in days between negotiation and payment. This is not related to the delivery term of futures contracts. CEPEA uses a conversion between Brazilian reals and US dollars based on the commercial market USD sale price as of 4:30 pm.

To build this data series, CEPEA contacted all possible industry members regardless of sophistication and size, ranging from soy producers to trading firms and brokers and soy consumers such as chicken and hog farmers. These organizations were each reviewed for their capacity to participate in the data provision in a reliable way, as well as with an eye toward selecting a representative sample of participants to capture a full picture of regional soy prices. Contributors are only retained if they participated regularly in meetings with CEPEA and if provided data regularly during those meetings.

Daily data is collected at random from qualifying contributors throughout the day from 0900-1700, to be aggregated and published by 1800. Once all data points are collected, which include unmet offers of sale and purchase, those offers which are outside the daily range of transacted prices are excluded. A simple mean of the remaining data points constitutes the initial average. Then, data points outside of the range of two standard deviations are excluded, and a new average calculated. Subsequently, the coefficient of variation is compared to a critical value (CV), defined as 25% above the average of the past 20 days' coefficients of variation. If the current day's CV is above this, the current day's

[17] An explanation of the methodology associated with the construction of this data is available in the file "Metodologia" (accessible at `http://www.cepea.esalq.usp.br/br/metodologia/metodologia-da-soja-esalq-bm-fbovespa-paranagua.aspx`.

average price is compared to the prior day's published price indicator, potentially resulting in exclusion of additional data (this process consists of removing the most "extreme" data points successively until the above critical value comparison is passed).

Since 5 April, 2015, any day with five or fewer qualifying data points, the prior day's published price indicator is added as a single data point and the above procedure is followed as usual. When there are two or fewer qualifying data points, all offers and bids are added in regardless of whether they are outside the range of transacted prices on that day. The remaining analysis on these dates follows the above process. We use data from March 2006 to April 2017; in total, there are 2,765 observations.

Brazilian ethanol prices are reported as daily present cash value equivalents in US dollars per cubic meter.[18] Prices are related only to fuel ethanol delivered in Paulínia or sent to other destinations such as Guarulhos, Barueri, Santo Andre, Sao Caetano do Sul, Sao Jose dos Campos, Cubatao, Ipiranga and Sao Paulo. The final prices are calculated taking the deliveries costs to Paulínia into account (*i.e.*, final price is the sum of ethanol price plus estimated freight between the mills and Paulínia).

[18] A discussion of the methodology used to construct this time series is available at `https://www.cepea.esalq.usp.br/en/methodology/methodology-12.aspx`.

References

Askari, H. and Krichene, N. (2008). Oil price dynamics (2002 – 2006), Energy Economics **30**: 2134–2153.

Babcock, B. A. (2013). RFS compliance costs and incentives to invest in ethanol infrastructure. CARD Policy Brief 13-PB 13, Iowa State University.

Ball, C. and Torous, W. N. (1985). On jumps in common stock prices and their impact on call option pricing, Journal of Finance **40**: 155–173.

Benth, E. F., Benth, J. and Koekbakker, S. (2008). Stochastic Modelling of Electricity and Related Markets, World Scientific, River Edge, NJ.

Burkholder, D. (2015). A Preliminary Assessment of RIN Market Dynamics, RIN Prices, and Their Effects. U.S. Environmental Protection Agency report EPA-HQ-OAR-2015-0111-0062.

Chevallier, J. and Sévi (2014). On the stochastic properties of carbon futures prices, Environmental and Resource Economics **58**: 127–153.

Dixit, A. K. and Pindyck, R. S. (1993). Investment under uncertainty, Princeton University Press, Princeton, NJ.

Duan, J. C. . (1997). Augmented GARCH(p,q) process and its diffusion limit, Journal of Econometrics **79**: 97–127.

Elder, J., Miao, H. and Ramchander, S. (2013). Jumps in oil prices: The role of economic news, The Energy Journal **34**: 217–237.

Engel, S., Palmer, C., Taschini, L. and Urech, S. (2015). Conservation payments under uncertainty, Land Economics **91**: 36–56.

Fargione, J., Hill, J., Tilman, D., Polasky, S. and Hawthorne, P. (2008). Land clearing and the biofuel carbon debt, Science **319**(5867): 1235–1238.

Johnston, J. and DiNardo, J. (1997). Econometric Methods, McGraw – Hill Companies Inc., New York, NY.

Knittel, C., Meiselman, B. S. and Stock, J. (2017). The pass-through of RIN prices to wholesale and retail fuels under the renewable fuel standard, The Journal of the Association of Environmental and Resource Economists **4**: 1081–1119.

LaRiviere, J., Lima, L. and Musinov, S. (2015). New markets and new market frictions: Evidence from ethanol and retail gasoline prices. University of Tennessee Working Paper.

Martzoukos, S. H. and Trigeorgis, L. (2002). Real (investment) options with multiple sources of rare events, European Journal of Operational Research **136**: 696 – 706.

Mason, C. F. and Wilmot, N. (2016). Price discontinuities in the market for RINs, Journal of Economic Behavior and Organization **132 (Part B)**: 79–97.

Mason, C. F. and Wilmot, N. A. (2022). Energy price jumps, fat tails and climate policy, Environmental Modeling & Assessment . forthcoming.

Meiselman, B. S. (2016). Breaching the blendwall: RINs and the market for renewable fuel. University of Michigan Working Paper.

Merton, R. (1976). Option pricing when the underlying stock returns are discontinuous, Journal of Financial Economics **3**: 125–144.

Morrison, G. M. and Chen, Y. (2011). Uncertain future for california's low-carbon fuel standard?, Transportation Research Record **2252**: 16–22.

Searchinger, T., Heimlich, R., Houghton, R. A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D. and Yu, T.-H. (2008). Use of u.s. croplands for biofuels increases greenhouse gases through emissions from land-use change, Science **319**(5867): 1238–1240.

U.S. Energy Information Administration (2013). RINs and RVOs are used to implement the Renewable Fuel Standard. Available at: http://www.eia.gov/todayinenergy/detail.cfm?id=11511.

Weitzman, M. (2009). On modeling and interpreting the economics of catastrophic climate change, Review of Economics and Statistics **91**: 1–19.

Wilmot, N. A. (2010). Investment under Uncertainty Utilizing Alternative Stochastic Processes, PhD thesis, University of Wyoming, Department of Economics & Finance.

Zhang, Y., Joshi, S. and MacLean, H. L. (2010). Can ethanol alone meet California's low carbon fuel standard? an evaluation of feedstock and conversion alternatives, Environmental Research Letters **5**: 014002.

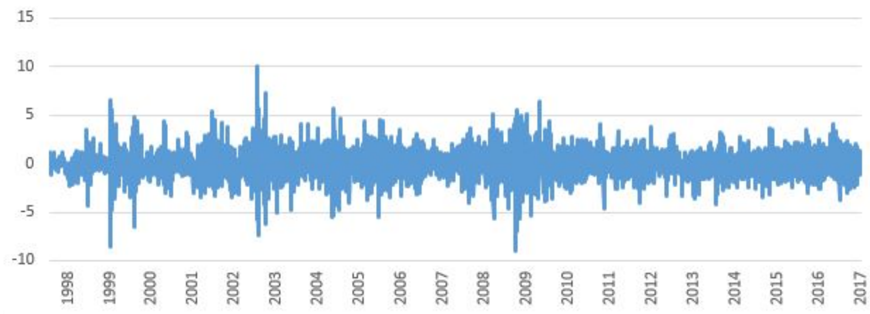


Figure 1: Soybean returns

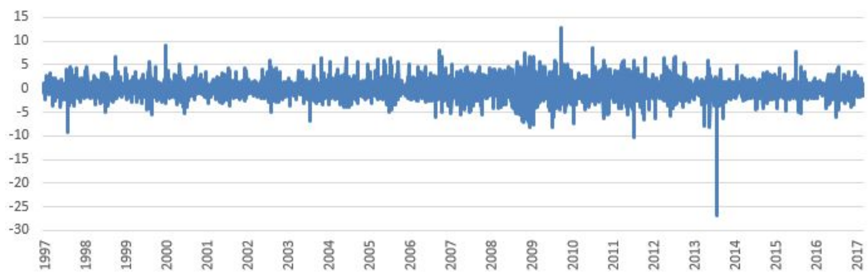


Figure 2: Corn returns

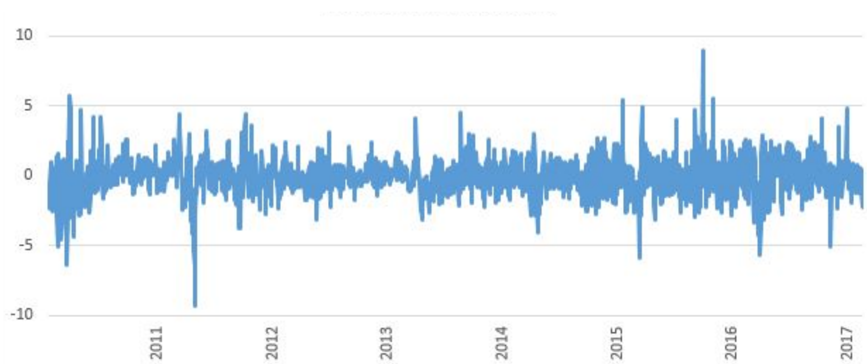
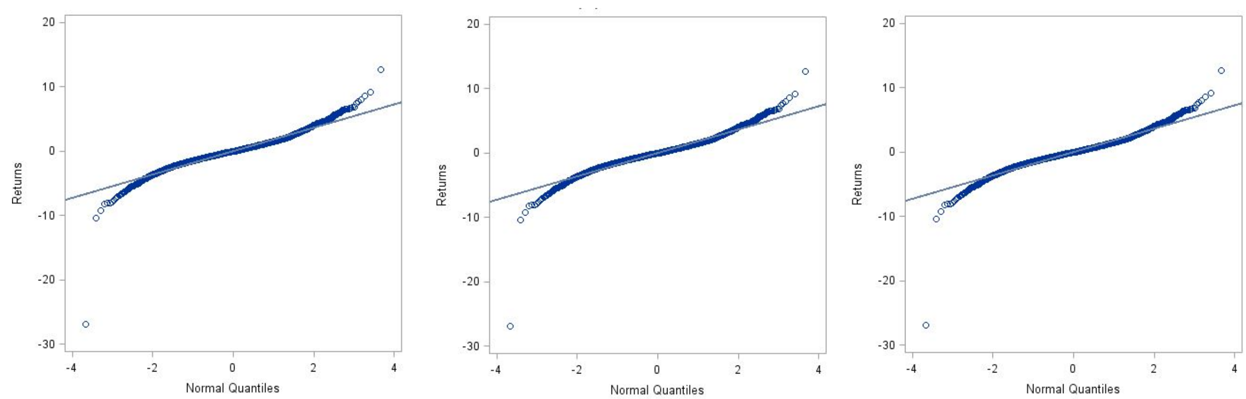


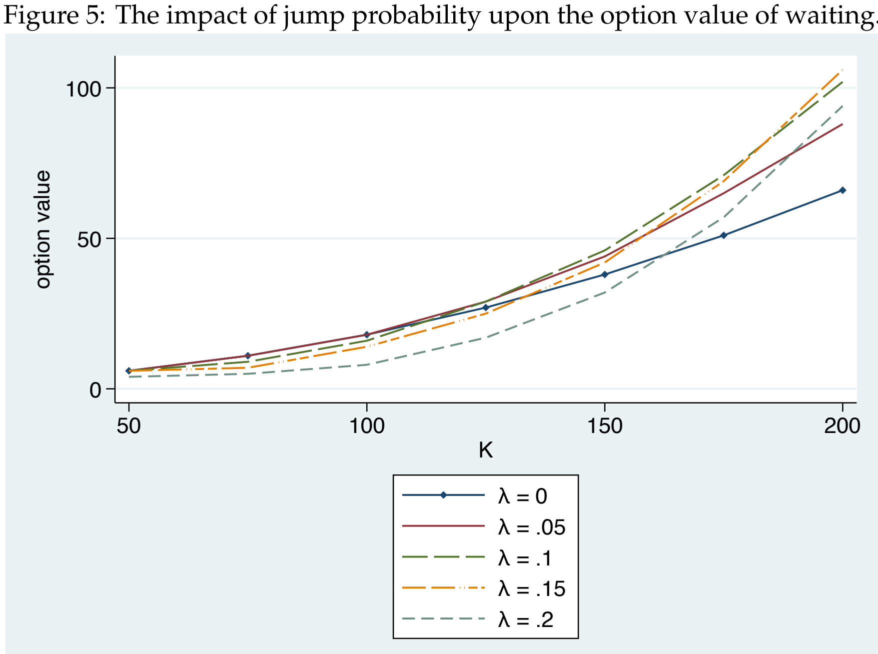
Figure 3: Ethanol returns



(a) (b) (c)

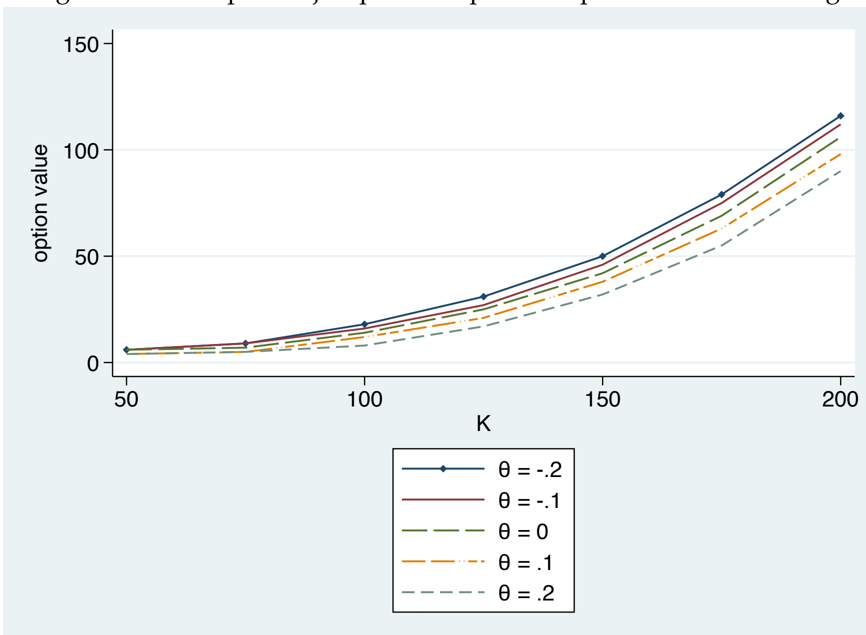
Figure 4: Quantile-quantile plots for soybeans (panel a), corn (panel b) and ethanol (panel c) price returns.

Figure 5: The impact of jump probability upon the option value of waiting.



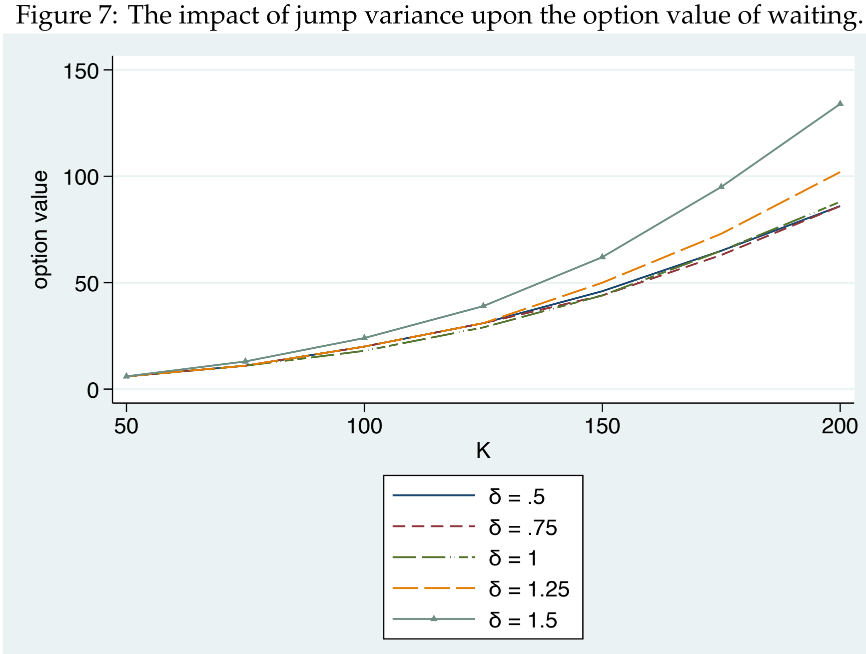
Source: Authors' calculations.

Figure 6: The impact of jump mean upon the option value of waiting.



Source: Authors' calculations.

Figure 7: The impact of jump variance upon the option value of waiting.



Source: Authors' calculations.

Table 1: Summary statistics

	Soybean		Corn		Ethanol	
	Prices	Returns	Prices	Returns	Prices	Returns
Start	7/29/97	7/30/97	12/10/96	12/11/96	01/25/2010	1/26/10
End	1/6/17	1/6/17	2/27/17	2/27/17	2/17/17	2/17/17
Mean	18.98	0.006	356.2	0.006	541.1	-0.01
Variance	59.54	1.805	25809	3.280	15873	1.91
Std Dev	7.716	1.344	160.7	1.811	126.0	1.38
Coeff. of Variation	40.7	24190	45.1	30582	23.3	-10387
Skewness	0.5253	-0.1767	1.0988	-0.5394	0.6915	-0.0655
Kurtosis	-0.7700	3.932	0.2150	11.86	0.5409	4.039
Anderson – Darling Normality test	129.9*	31.30*	248.2*	36.66*	14.10*	10.21*
N	4842	4841	5091	5090	1753	1750

*: statistically significant at better than .01 level

Soybean returns measured by CEPEA / ESALQ Soybean Price Index

Corn returns measured by Corn Futures Price - Front Month Contracts

Ethanol returns measured by CEPEA/ESALQ hydrous ethanol Index

Kurtosis is measured as “excess” kurtosis, so that normal distributed variables should have values close to 0.

Table 2: Estimation of the Model Parameters for Daily price returns

	μ	σ	κ	α_1	β_1	λ	θ	δ
				A. Soybeans				
PD	0.0056	1.34***						
	(0.016)	(0.014)						
JD	0.0448*	0.883***				0.376***	-0.105	1.63***
	(0.025)	(0.047)				(0.091)	(0.069)	(0.145)
GPD	0.0202		0.0487***	0.0939***	0.880***			
	(0.017)		(0.008)	(0.009)	(0.011)			
GJD	0.0241		0.0137***	0.0792***	0.891***	0.124***	-0.178	1.69***
	(0.017)		(0.004)	(0.008)	(0.011)	(0.024)	(0.118)	(0.143)
				B. Corn				
PD	0.0067	1.82***						
	(0.012)	(0.018)						
JD	-0.0482*	1.16***				0.382***	0.144*	2.21***
	(0.025)	(0.040)				(0.057)	(0.076)	(0.131)
GPD	-0.0007		0.0552***	0.0622***	0.925***			
	(0.023)		(0.014)	(0.011)	(0.013)			
GJD	-0.0321		0.0239***	0.0496***	0.922***	0.103***	0.266	2.86***
	(0.022)		(0.007)	(0.006)	(0.010)	(0.022)	(0.187)	(0.248)
				C. Ethanol				
PD	-0.0133	1.38***						
	(0.029)	(0.023)						
JD	-0.0082	1.03***				0.177***	-0.0289	2.17***
	(0.046)	(0.048)				(0.061)	(0.092)	(0.297)
GPD	0.0406		0.0482***	0.117***	0.862***			
	(0.029)		(0.015)	(0.019)	(0.021)			
GJD	0.0004		0.0604**	0.124***	0.819***	0.0510**	0.626	2.28***
	(0.029)		(0.026)	(0.027)	(0.041)	(0.026)	(0.426)	(0.459)

Standard errors in parentheses. Asterisks signify statistical significance: *: better than 10% level; **: better than 5% level; ***: better than 1% level.

Table 3: Likelihood ratio test statistics

	PD vs. JD	PD vs. GPD	JD vs. GJD	GPD vs. GJD
Soybean Spot Returns	539.4	870.2	547.5	216.7
	(0.000)	(0.000)	(0.000)	(0.000)
Corn Futures Returns	644.4	538.9	414.8	520.4
	(0.000)	(0.000)	(0.000)	(0.000)
Ethanol Returns	197.6	315.9	60.3	178.6
	(0.000)	(0.000)	(0.000)	(0.000)

p-values presented (in parentheses) below test statistics.

Table 4: Forecasting results: PD vs. GJD

	Summed prediction error		
	Soybeans [2,420]	Ethanol [875]	Corn [2,544]
PD	-113.40	-75.72	129.2
GJD	-8.143	8.447	106.740
$\|GJD/PD\|$	0.0718	0.1116	0.8260
	Summed squared prediction error		
	Soybeans [2,420]	Ethanol [875]	Corn [2,544]
PD	8674.7	3601.2	17007.0
GJD	8246.2	3282.2	17005.6
$\|GJD/PD\|$	0.9506	0.9114	0.9999

Sample represents second half of full sample; parameter estimates used to form predictions are based on first half of sample. Values in square brackets are number of observations used for each half.