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 refineries 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, as well as ecological implications associated with deforestation that is likely to accompany any increases in Brazilian soybean production.

Keywords: soybean prices; biofuels; jump diffusion; GARCH; investment under uncertainty
JEL codes: Q30, Q40, C58

*We thank participants at the 9^{th} International Symposium on Environment and Energy Finance Issues for their constructive input on an earlier version of this paper. The usual disclaimer applies.

†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) be in use by 2022 (U.S. Energy Information Administration, 2013). At the state level, California has adopted the "Low Carbon Fuel Standard" (LCFS), which required 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 the 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).

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 pro-

[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.

cess 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. As we discuss below, the presence of fat tails exerts effects qualitatively similar to increases in both mean and variance of a GBM stochastic prices, both of which commonly induce a delay in investment (Dixit and Pindyck, 1993). To the extent that such capital projects 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).[2] These effects are also consistent with behavior by risk averse decision-makers, in that the objective function governing investment decisions under uncertainty is commonly concave. In such a framework, a change that increases the expected utility of delaying investment will make that action more attractive; as we discuss below, this sort of effect arises when the stochastic variable of interest is influenced by a jump process. In addition, changes to the stochastic environment that mimic inclusion of a mean-preserving spread, as when one allows for fat tails (either via jumps or time-varying volatility) lowers the appeal of actions tied to the variable, thereby requiring an increase in expected payoff to induce activity. Either way, the presence of fat tails seems likely to delay investment.

One resolution of these difficulties would be to shift refiners' reliance from corn products as inputs in the ethanol production process to soybeans.[3] Accordingly, to achieve 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. As

[2] This aspect, 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.

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

Figure **??** illustrates, Brazil and the US have been the two largest sources of soybean globally for some time. But while Brazilian production has steadily expanded during this period, US production stagnated during the past decade. Figure **??** shows this pattern. The combined implication is that Brazil will likely play an ever-expanding role, particularly as an input into the production of biofuels. Indeed, Morrison and Chen (2011) argue that Brazilian ethanol could account for 25% of all transportation energy in California in the coming years.

Our goal in this paper is to analyze price returns for soybeans, from both Brazil and the US, so as to determine the empirical importance of elements that might contribute to fat tails. As the two largest sources of soybean production in the world, any shifts in information regarding production and yield within Brazil or the US are likely to exert important effects on global soybean markets. Moreover, any changes in expectations regarding soybean exports from competing nations, like China, can swiftly reverberate onto soybean prices, with fluctuations influenced by changes in export volumes from major producers; the potential for such a market environment to lead to jumps in prices has been recognized for some time (Koekebakker and Lien, 2004). Dramatic weather phenomena, such as hot and dry conditions in critical growing regions, could be a significant contributors to market dynamics, prompting revisions in crop estimates and production forecasts (Braun, 2024). To allow for such possibilities, 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 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 are 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.

The consequences of increased reliance on soybeans, particularly production in Brazil, is that such increased production could be associated with deforestation. This raises important concerns related to the potential impact on land allocation. Indeed, there is a historical pattern of deforestation in Brazil – particularly in the Amazon basin – related to increased soy production.[4] Potentially working in the other direction, the potential for fat tails in Brazilian soybean prices might serve to deter large-scale investment in production processes that utilize Brazilian soy inputs. Related to this element, deforestation is an example of significant up-front and largely irreversible costs; we argue below that the presence of fat tails, say due to jumps, can increase the option value of waiting to undertake such investment, particularly when the up-front costs are significant. Policy interventions that serve to raise these costs might thereby benefit from the presence of jumps. As both GARCH and jumps will induce fat tails, our empirical results may have important

[4] See Gasparri et al. (2013); Gollnow et al. (2018). Much of this past pattern of behavior was related to the use of soy as a feedstock for the production of beef; these concerns lead to the promulgation of a voluntary agreement, known as the "Soy Moratorium," to reduce impacts on forests. Under this agreement, soy producers pledged to produce without contributing to deforestation. The original agreement had a finite life; negotiations to extend the agreement extended the terms indefinitely. While there is evidence that the agreement initially lead to a sharp reduction in deforestation (Kastens et al., 2017), there is some concern that deforestation rates have been increasing over the past decade. For example, Gollnow et al. (2018, p. 377) noted in 2018 that deforestation rates in Brazil increased after 2013, a pattern that World Resources Institute (2022) recently corroborated.

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[5]

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

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,

[5] 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.

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 are observed during this interval equals

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

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,[6] 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 $P_t = \exp(J_t)\,P_{t-}$; accordingly, J_t measures the percentage change in price. The resultant stochastic process

[6] 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).

for the random variable P_t may then be written as

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

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 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:[7]

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

where the conditional variance, h_t is described by the process

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

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 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.[8] 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

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

[7] 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.

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

where n indexes the number of jumps, combined with the description of h_t given in equation (9).[9] 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, $L(\phi^R;x)$, 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

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

study consist 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 values, converted into US dollars; these reflect the value of current soybean trades or bids reported by CEPEA collaborators, per 60-kilo bag, for soybean delivered at the unit that loads ships at Paranaguá port, in Paraná State. Brazilian fuel ethanol prices are reported as daily present cash value equivalents in US dollars per cubic meter. Both these prices are retrieved from the CEPEA website.[10] 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 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 3(a)–3(c), we plot the price returns for the three time series. Brazilian soybean returns are shown in Figure 3(a), corn returns are shown in Figure 3(b) and Brazilian ethanol returns are shown in Figure 3(c). 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, as well as evidence of leptokurtosis or "fat-tails," as evidenced by the large value for kurtosis. The Anderson – Darling test, a quadratic empirical distribution function test, is used to examine the normality of the data. The results of the test imply the null hypothesis of a normally distributed random

[10] 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.

variable is strongly rejected for each of our time series.

These results are corroborated by the "quantile–quantile" (QQ) 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 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 QQ 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 QQ 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

4.1. Main 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 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 estimated values for λ suggest that jumps occur, on average, quite frequently (once every 8 days) in the soybean markets, but much less frequently (once every 20 days) in the ethanol market.[11] Though insignificant, the mean jump size, θ, suggests that soybeans returns tend to experience negative jumps. This is in contrast to the corn and ethanol markets, where a positive (though insignificant) 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 suggests a jump occurs on average approximately once every 3 days for soybeans and corn and once every 4 days for ethanol; the *GJD* suggests a jump occurs on average approximately once every 6 days for soybeans, and every 10 days for corn and ethanol.

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

[11] The expected arrival time for a jump is the inverse of the jump intensity, *i.e.* $1/\lambda$. Thus, the expected arrival times are approximately 8 (= 1/.124) days for soybeans and 20 (= 1/.051) days for ethanol; the expected arrival time for corn markets, 10 (= 1/.103) lies in between these values.

displayed in column two show that allowing for jumps yields a statistically important increase in predictive power, relative to the pure-diffusion model, for each price return series. Likewise, the results displayed in column three indicate that allowing for time-varying volatility improves model fit, relative to the pure-diffusion model. 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.

4.2. Extensions

We conclude this section by providing material related to a potential break. In December 2013, the U.S. Environmental Protection Agency (EPA) proposed new rules for biofuels; this lead to promulgation of several new initiatives in the US Congress. There is some indication that this heightened legislative activity triggered increased potential for jumps (Mason and Wilmot, 2016). Indeed, Lade et al. (2018, p. 708) argue that the 2014 proposed rule exerted important effects on corn and soybean price returns and impacted advanced biofuel and biodiesel producers. These are sources of production that might well have been viewed as of potentially increasing importance. The adverse impact from the proposed adjustments to the RFS could potentially have connoted a structural break in the three time series.[12]

We explore the potential impact on the appropriate characterization of the stochastic process for these price returns in this subsection. Anticipating the potential lag between the legislative activity and market response, we allow for a structural break in the middle of 2014. To this end we split the data into two subsamples: the period up to June 2014 and the period after July 2014. We then repeat the analysis documented in subsection 4.1 for each of these sub-samples. We report results for the period through June 2014 in Table 4, while results for the period from July 2014 onward are presented in Table 5. All these results are qualitatively similar to those in Table 2; in particular, there is substantial evidence for importance of jumps both before and after break, with the combined GJD model providing multiple statistically important variables. In light of that result, it is perhaps not surprising that the GJD model statistically outperforms the other models.[13] Comparing the GJD estimates in the two tables, we see that the estimated jump probability decreases for soybeans but is largely unchanged for corn and ethanol.

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 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

[12] Formal analysis of the times series via the Zivot-Andrews Unit Root Test (allowing for a structural break) identifies the structural break date for Brazilian soybeans as 30 June, 2014, and the structural break date for US corn as 12 July 2013. In light of this evidence, and the information in the above paragraph, we use a break date of June 2014 in the following analysis.

[13] Results available upon request.

the decision-maker to obtain a new payoff flow. The investment could reflect expanding refinery capacity to process increased inflows of ethanol, building a dedicated factory for biofuels or undertaking a large-scale change in land use via deforestation. 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.[14] This implies the benefits associated with investing at time t can be expressed by a stochastically evolving component, which we write as X_t. Letting K denote the one-time investing cost, the net benefits of acting (investing) at t are equal to[15]

$$X - K.$$

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

[14] 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 (Dixit and Pindyck, 1993, pp. 195-199).

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

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.[16]

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:

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

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. This component can be expressed as (Dixit and Pindyck, 1993):

$$\frac{1}{\mathrm{dt}} E\Big[\mathrm{d}(F)\Big] = \alpha X F'(X) + \frac{\sigma^2}{2} X^2 F''(X). \tag{13}$$

[16] 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).

It can then be shown that the solution eq. (12) is a power function:

$$F(X) = aX^b, \tag{14}$$

where the parameter $\beta > 1$ is the positive solution to the quadratic expression

$$Q(b) = \frac{\sigma^2}{2} b(b-1) + \alpha b - \rho = 0. \tag{15}$$

It is easy to see that b depends positively on σ and negatively on α.[17]

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

$$F(X^*) = X^* - K \tag{16}$$

along with the "smooth-pasting" condition

$$F'(X^*) = d(X^* - K)/dX^* = 1. \tag{17}$$

Applying the value-matching and smooth-pasting conditions to the functional form in

[17] It is easy to see that Q is convex, with $Q(0) < 0$. It follows that one of the two roots is negative; the boundary condition that requires the value F to tend to zero as X becomes small then forces the scale coefficient on the term with the negative root to be zero, thereby picking out the positive root in eq. (15). The requirement that $\alpha < \rho$ then forces the positive root to exceed 1. Larger values of α shift Q up, rendering a smaller value of the positive root, while larger values of σ shift Q down, rendering a larger value of the positive root.

eq. (14), it is easy to show that the cutoff value is:

$$X^{*} = \frac{bK}{b-1}. \tag{18}$$

As noted above, b 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 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.

An alternative way to think about this problem emerges from consideration of the concavity in the value function $F(X)$. This concavity evokes the concept of risk aversion, which in turn connects to analyses of decision-making that are based on the expected utility representation in the presence of key stochastic components. Elements that increase the expected utility of a particular action, such as delaying investment, will make that action more attractive. That would be the case with an increase in the drift term of the stochastic variable X, *i.e.* α; or decreases in the riskiness of that variable, *i.e.* decreases in σ. It is also true that adjusting the stochastic variable via a mean-preserving spread lowers the appeal of actions tied to the variable, thereby requiring an increase in expected payoff to induce activity. We return to this point below.

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, which we denote by J.[18] The probability a jump will occur during a brief time interval of length dt is then λdt, where $\lambda > 0$ is a parameter measuring the arrival frequency. Should a jump occur, we assume it generates a value Y that we interpret as creating a proportional change to X; this proportion is independently and identically distributed as a lognormal random variable – so that $\ln(Y)$ is Normally distributed. We denote the mean and variance of $\ln(Y)$ by θ and δ^2, repsectively. Recalling the characterization of a GBM process from eq. (5), we may describe the combined jump - GBM process by

$$dX_t = \alpha X_t dt + \sigma X_t dz_t + Y X_t dJ_t. \tag{19}$$

Including jumps in this manner 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.[19]

If a jump occurs it make move the price directly into that region where investment is undertaken, or it may leave price in the continuation region. In the latter event the function characterizing the continuation value is still given by eq. (12), but now the term

[18] Some authors model price jumps using a Lévy process, an approach that requires an *ex ante* 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. As we noted above, a number of other authors assume jumps follow a Poisson process.

[19] See Dixit and Pindyck (1993) for discussion. Another explanation for "fat tails" often observed for many energy commodities 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).

on the right-hand side is impacted by the jump component:

$$\frac{1}{\mathrm{dt}}E\Big[\mathrm{d}(F)\Big] = \alpha XF'(X) + \frac{\sigma^2}{2}X^2F''(X) - \lambda F(X) + \lambda\mathcal{E}_Y\Big[F\Big(YX\Big)\Big] \tag{20}$$

where the expectation in the final term on the right-hand side is take with respect to the random variable Y. The solution to the differential equation describing the continuation value remains a power function, where the exponent β in this function is now the solution to the equation (Dixit and Pindyck, 1993)

$$0 = \widetilde{Q}(\beta) = \frac{\sigma^2}{2}\beta(\beta-1) + \alpha\beta - r + \lambda\Big(\mathcal{E}_Y\Big[Y^\beta\Big] - 1\Big) \tag{21}$$

$$= \frac{\sigma^2}{2}\beta(\beta-1) + \alpha\beta - r + \lambda\left(e^{\beta\left(\theta+\frac{\delta^2}{2}\right)} - 1\right). \tag{22}$$

The solution for β must be obtained numerically, though it is clear that this value is smaller than the solution to eq. (15); this implies that the problem that includes jumps renders a more concave continuation value function, and therefore increases the cutoff value of X at which investment will occur.

Because a positive value of $\lambda\theta$ will raise the drift term, it will induce a delay in investment. That is, when the jump component is on average positive, a greater arrival frequency should be associated with delayed investment. In addition, interpreting fat tails as comparable to a mean-preserving spread, and bearing in mind the analogy we drew above to decision-making by a risk-averse individual, we anticipate that anything which renders "fatter" tails – be that in increase in the variance of the value taken by a jump, should it occur, or anything that increases the magnitude of the GARCH component h_t, seems likely to motivate increased delay in investment. We provide corroborating evidence of these conjectures in the simulation analysis, discussed below.

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. We then present the relation between this critical investment value and the requisite level of investment, at various levels of three key parameters: the jump intensity λ, the mean jump size, θ, and the standar deviation in the jump size, δ.

Our first set of simulations investigates the role played by the jump intensity, λ. As we noted above an increase in λ increases the drift term and raises the variance of the stochastic process, each of which should in principle incentivize a delay in investment. In this set of simulations we vary λ between 0 (which corresponds to geometric Brownian motino, GBM) and 0.2, by increments of 0.1; results from this set of simulations are summarized in Figure 5. The critical investment value under GBM is described by the solid curve, while the cutoff investment values for $\lambda = .1$ are represented by the dashed curve and the cutoff investment values when $\lambda = .2$ is given by the long-dashed curve. The first feature we observe is that for both positive values of λ, the critical investment value exceeds that level under GBM, as we conjectured above. The second feature we observe is that the gap between the critical investment value for positive λ and GBM increases 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. Finally, we observe that the cutoff investment value is larger for $\lambda = .2$ than for $\lambda = .1$. The impication is that the tendency to delay the investment becomes more pronounced as the probability of a jump increases: while less sensitive to λ when the required investment is small, the option value of waiting does respond to increased jump intensity at larger investment levels.

In the second set of simulations, we identify the cutoff investment for three values of θ (the expected value of the natural log of the jump size): 0, -0.1 and 0.1. 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. Here we observe that increases in θ are associated with increases in the cutoff value, for a given required level of investment. As we noted above, an increase in θ will raise the drift in the stochastic process (so long as $\lambda > 0$), and so the results embodied in this figure confirm the intuitions we developed above.

The third set of simulations we consider varies δ, the standard deviation of the jump size; here we consider values equaling 1 plus or minus .25 (*i.e.*, .75 and 1.25). 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 reduces the appeal of the investment for any particular expected gain that obtains at the moment of investment. In turn, this motivates the agent to delay investment until the anticipated gain is larger – which is associated with a larger triggering value associated with the activity. In this way we expect the critical investment value to increase with δ, as Figure 7confirms. We also expect that the cutoff level will exceed that which arises under GBM, because of the heightened uncertainty. This too is confirmed by the figure. Moreover, heightened variation in the potential size of the jump playa 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. For example, the accuracy of a decision to significantly expand a refinery to handle ethanol infrastructure, or to process imported soybeans, will almost surely be improved by such enhanced understanding.[20] 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 distributions of these prices are "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).

On the other hand, fat tails in soybean prices can increase the option value of delaying the conversion of a plot of land, so as to facilitate expanded soybean production. This effect could amplify the impact of policies designed to inhibit deforestation, such as requiring the purchase of a permit prior to clearing a forest, or some form of punishment (such as a substantial fine or jail time) for undertaking such land conversion without prior government approval.

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).[21] To 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

[20] 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).

[21] 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.

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, or delaying land conversion.

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.

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 Koekebakker, S. (2008). Stochastic Modelling of Electricity and Related Markets, World Scientific, River Edge, NJ.

Braun, K. (2024). Have Brazilian Estimates Slipped too far From Reality? Reuters. Published 18 January, 2024; accessed online 10 February, 2024 at `https://www.reuters.com/markets/commodities/have-brazilian-soybean-estimates-slipped-too-far-reality-2024-01-18/`.

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.

Gasparri, N. I., Grau, H. R. and Gutiérrez Angonese, J. (2013). Linkages Between Soybean and Neotropical Deforestation: Coupling and Transient Decoupling Dynamics in a Multi-Decadal Analysis, Global Environmental Change **23**(6): 1605–1614.

Gollnow, F., de Barros Viana Hissa, L., Rufin, P. and Lakes, T. (2018). Property-Level Direct and Indirect Deforestation for Soybean Production in the Amazon Region of Mato Grosso, Brazil, Land Use Policy **78**: 377–385.

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

Kastens, J. H., Brown, J. C., Coutinho, A. C., Bishop, C. R. and Esquerdo, J. C. D. M. (2017). Soy Moratorium Impacts on Soybean and Deforestation Dynamics in Mato Grosso, Brazil, PLoS ONE **12**: e0176168.

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.

Koekebakker, S. and Lien, G. (2004). Volatility and Price Jumps in Agricultural Futures Prices-Evidence from Wheat Options, American Journal of Agricultural Economics **86**(4): 1018–1031.

Lade, G. E., Lin Lawell, C.-Y. C. and Smith, A. (2018). Policy Shocks and Market-Based Regulations: Evidence from the Renewable Fuel Standard, American Journal of Agricultural Economics **100**(3): 707–731.

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.

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.

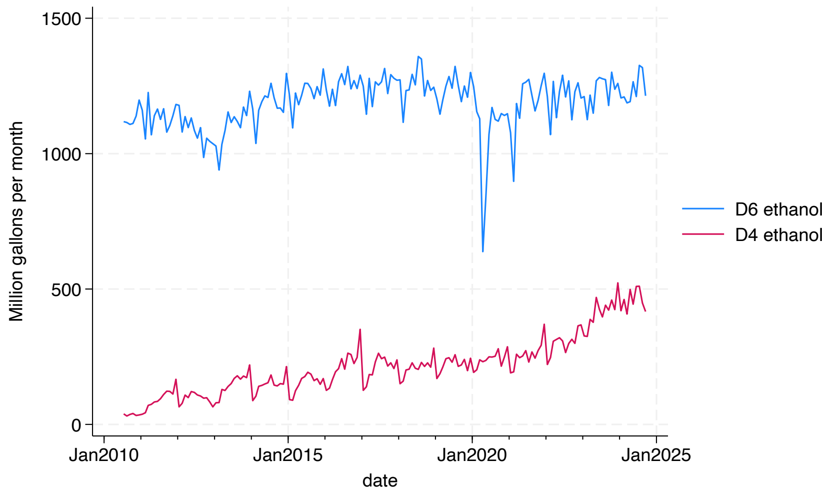
U.S. Energy Information Administration (2013). RINs and RVOs are used to implement the Renewable Fuel Standard. Available at: http://www.eia.gov/todayinenergy/detail.cfm?id=11511.

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

World Resources Institute (2022). Global forest review, 2022, update 3. Washington, DC. Available online at https://research.wri.org/gfr/global-forest-review.

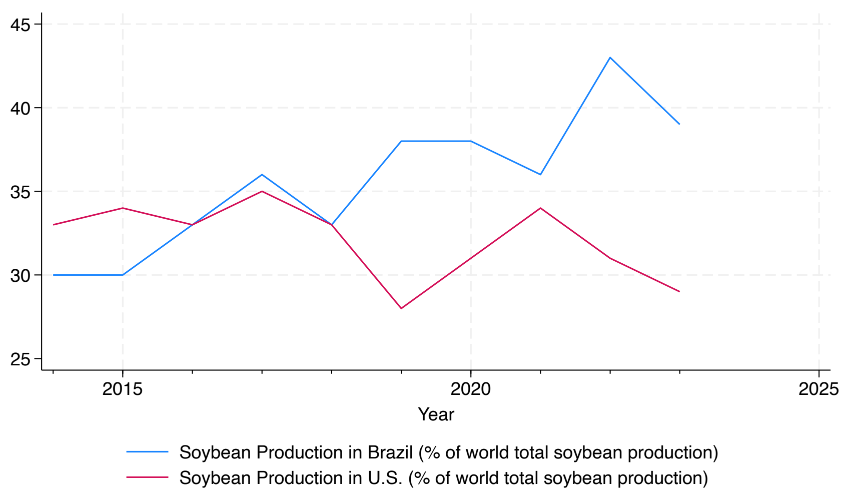
Zhang, Y., Joshi, S. and MacLean, H. L. (2010). Can ethanol alone meet California's low carbon fuel standard? an evaluation of feedstock and conversion alternatives, Environmental Research Letters **5**: 014002.

Figure 1: US Ethanol production, by type: D4 vs. D6.



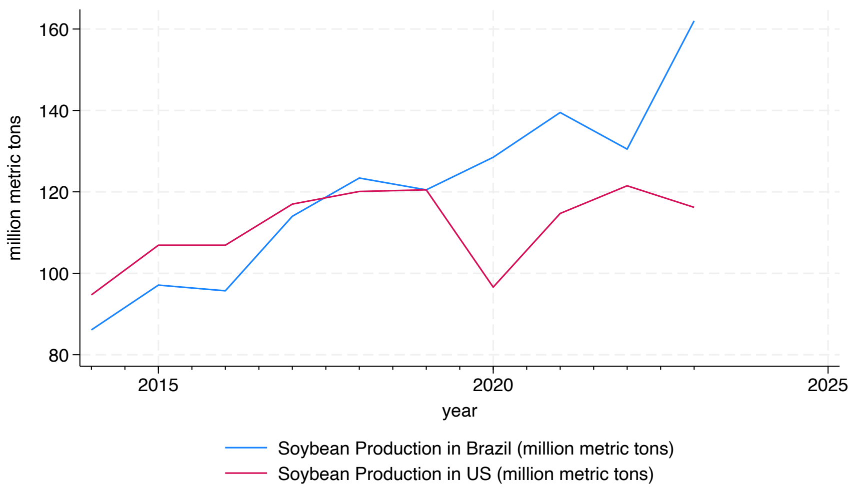
Source: .

Figure 2: The share of global soybean production: Brazil and the US.

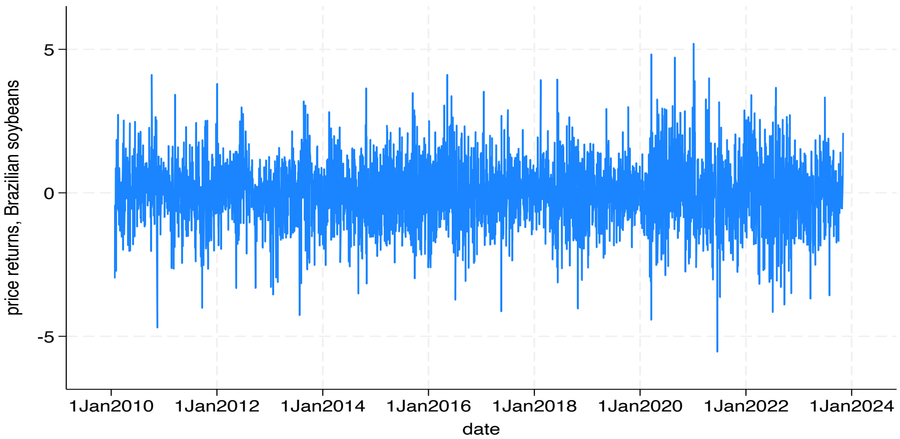


Source: .

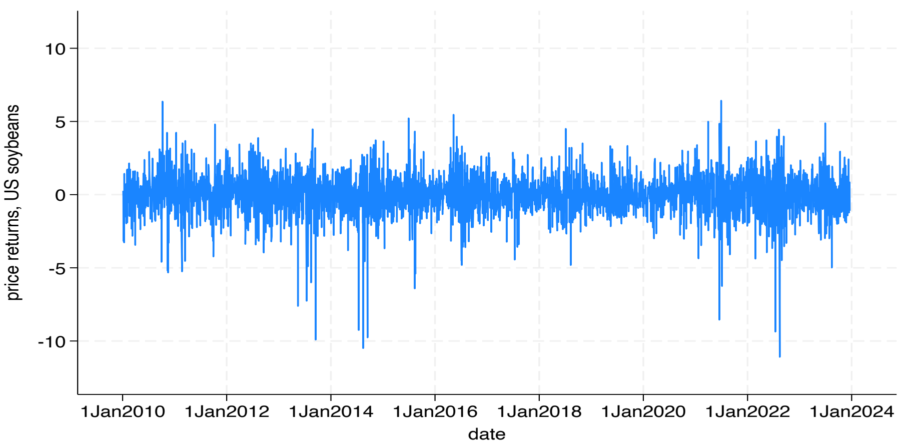
Figure 3: Soybean production from Brazil and the US.



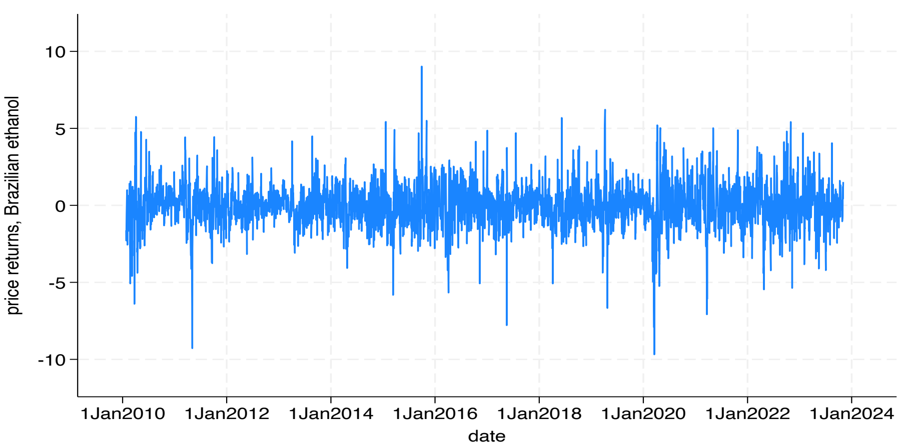
Source: .



(a)



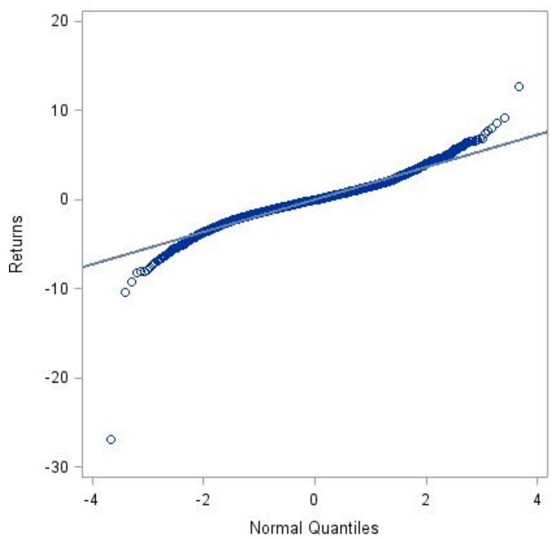
(b)



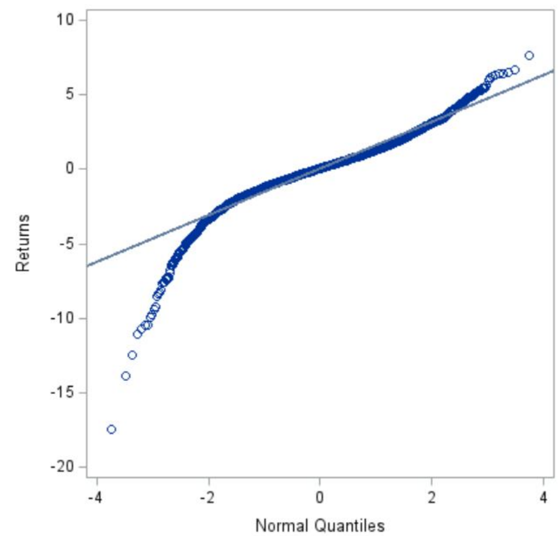
(c)

Figure 4: Price returns plots for Brazilian soybeans (panel a), US soybeans (panel b) and ethanol (panel c) price returns.

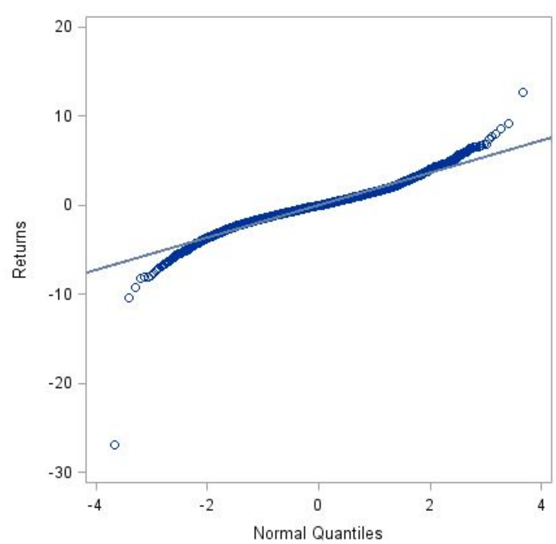
Source: Authors' calculations.



(a)



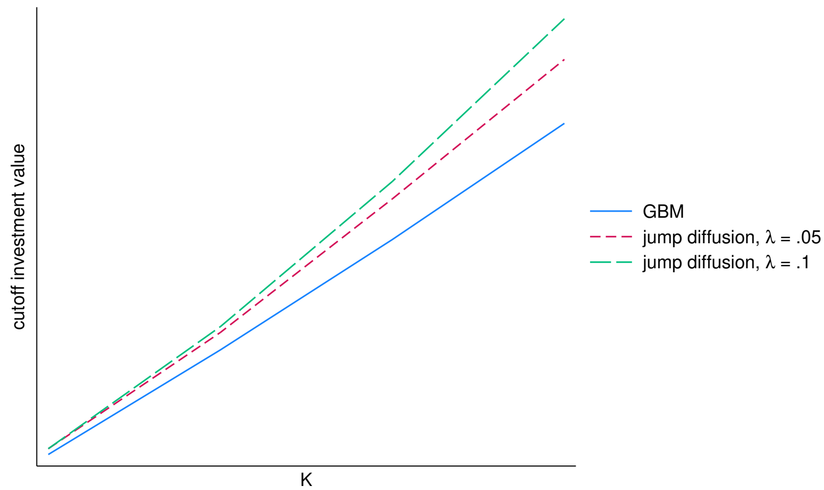
(b)



(c)

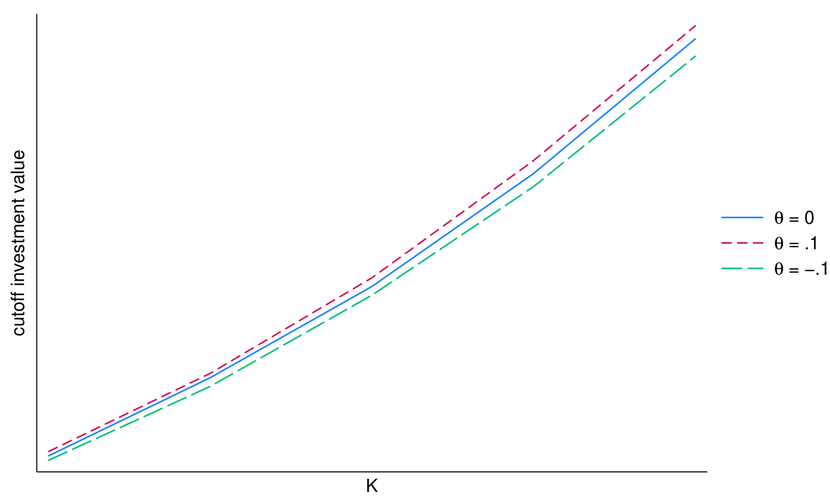
Figure 5: Quantile-quantile plots for soybeans (panel a), corn (panel b) and ethanol (panel c) price returns.
Source: Authors' calculations.

Figure 6: The impact of jump probability upon the critical investment value.



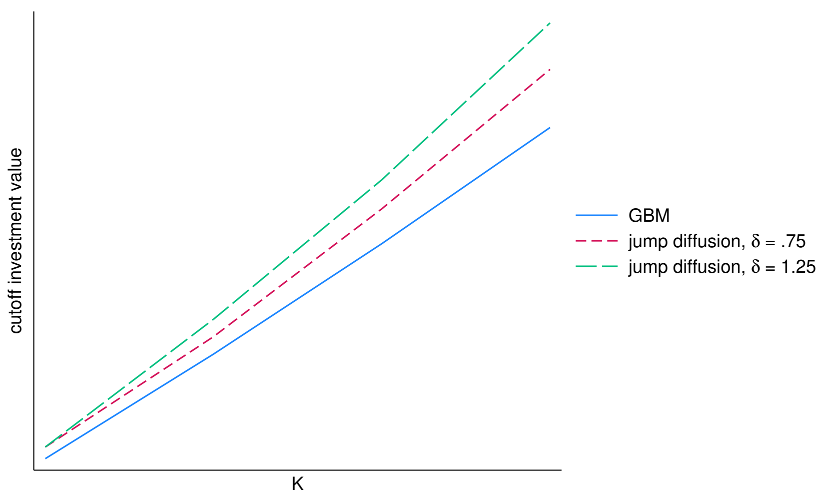
Source: Authors' calculations.

Figure 7: The impact of jump mean upon the the critical investment value.



Source: Authors' calculations.

Figure 8: The impact of jump variance upon the the critical investment value.



Source: Authors' calculations.

Table 1: Summary statistics

	Brazilian Soybean		Brazilian Ethanol		US Soybeans	
	Prices	Returns	Prices	Returns	Prices	Returns
Start	7/29/1997	7/30/1997	7/29/1997	7/30/1997	7/29/1997	7/30/1997
End	12/19/2023	12/19/2023	12/15/2023	12/15/2023	12/19/2023	12/19/2023
Mean	20.72	0.01	388.93	194.48	948.53	0.0081
Variance	62.26	1.70	27101.11	51372.81	122,533.13	2.46
Std. Dev.	7.89	1.30	164.62	226.66	350.05	1.57
Skewness	0.31	-0.14	0.78	0.83	0.28	-1.00
Kurtosis	-0.85	3.53	-0.40	-0.42	-1.02	7.87
Anderson – Darling stat.	82.18	33.88	225.12	56.90	101.66	61.84
N	6574	6573	6648	6647	6652	6651

Note: Brazilian Soybean returns measured by CEPEA Soybean Price Index, as reported by the Department of Luiz de Queiroz College of Agriculture (ESALQ); US Soybean returns measured by Futures Price - Front Month Contracts; Brazilian Ethanol returns measured by CEPEA/ESALQ hydrous ethanol Index. Kurtosis is measured as "excess" kurtosis (*i.e.*, above 3), so that normal distributed variables should have values close to 0. All values of the Anderson-Darlington Normality test are statistically significant at better than the 1% level, strongly rejecting null hypothesis of Normality.

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

	μ	σ	κ	α_1	β_1	λ	θ	δ
				A. Brazilian Soybeans				
PD	0.008	1.303***						
	(0.008)	(0.011)						
JD	0.033*	0.901***				0.343***	-0.073	1.586***
	(0.019)	(0.039)				(0.078)	(0.047)	(0.130)
GPD	0.014		0.043***	0.076***	0.900***			
	(0.014)		(0.006)	(0.007)	(0.009)			
GJD	0.019		0.010***	0.067***	0.904***	0.171***	-0.113	1.476***
	(0.016)		(0.004)	(0.006)	(0.009)	(0.042)	(0.081)	(0.136)
				B. Brazilian Ethanol				
PD	-1.468**	1.448						
	(0.671)	(1.549)						
JD	-0.022	1.021***				0.253***	0.045	2.021***
	(0.021)	(0.038)				(0.056)	(0.074)	(0.175)
GPD	0.050**		0.073***	0.108***	0.859***			
	(0.022)		(0.020)	(0.015)	(0.021)			
GJD	0.010		0.050***	0.097***	0.847***	0.097***	0.209	2.036***
	(0.023)		(0.018)	(0.018)	(0.030)	(0.037)	(0.188)	(0.296)
				C. US Soybean				
PD	0.008	1.570***						
	(0.008)	(0.014)						
JD	0.089***	1.092***				0.201***	-0.404***	2.417***
	(0.019)	(0.028)				(0.032)	(0.106)	(0.153)
GPD	0.012		0.033***	0.076***	0.913***			
	(0.016)		(0.006)	(0.006)	(0.007)			
GJD	0.041**		0.029***	0.044***	0.930***	0.042***	-1.106***	3.258***
	(0.002)		(0.005)	(0.004)	(0.006)	(0.010)	(0.353)	(0.337)

Note: Standard errors in parentheses. Sample period: start of data reported in Table 1. Number of observations: 6,573 for soybeans; 6,647 for corn; 3,412 for ethanol; 6810 for US soybeans. 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
Brazilian Soybean Spot Returns	614.6	976.7	636.4	274.2
	(0.000)	(0.000)	(0.000)	(0.000)
Brazilian Ethanol Returns	3,835.8	4,006.0	314.9	144.7
	(0.000)	(0.000)	(0.000)	(0.000)
US Soybean Returns	1,116.4	1,340.2	532.6	308.9
	(0.000)	(0.000)	(0.000)	(0.000)

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

Table 4: Estimation of the Model Parameters for Daily price returns: Pre - July, 2014

	μ	σ	κ	α_1	β_1	λ	θ	δ
				A. Brazilian Soybeans				
PD	0.014	1.364***						
	(0.018)	(0.023)						
JD	0.071***	0.777***				0.527***	-0.108***	1.515***
	(0.023)	(0.048)				(0.100)	(0.067)	(0.172)
GPD	0.033*		0.048***	0.103***	0.874***			
	(0.018)		(0.008)	(0.011)	(0.012)			
GJD	0.045**		0.012**	0.083***	0.880***	0.181***	-0.192*	1.556***
	(0.019)		(0.005)	(0.009)	(0.013)	(0.042)	(0.099)	(0.148)
				B. Brazilian Ethanol				
PD	-0.024	1.310***						
	(0.039)	(0.028)						
JD	0.010	0.856***				0.263***	0.132	1.911***
	(0.030)	(0.045)				(0.069)	(0.138)	(0.221)
GPD	0.058*		0.050***	0.134***	0.839***			
	(0.033)		(0.016)	(0.023)	(0.025)			
GJD	-0.001		0.121**	0.231***	0.618***	0.119	0.254	1.621***
	(0.034)		(0.053)	(0.046)	(0.080)	(0.083)	(0.280)	(0.409)
				C. US Soybean				
PD	0.014	1.671***						
	(0.028)	(0.018)						
JD	0.123***	1.128***				0.237***	-0.458***	2.426***
	(0.024)	(0.037)				(0.042)	(0.122)	(0.171)
GPD	0.009		0.045***	0.075***	0.910***			
	(0.021)		(0.009)	(0.008)	(0.010)			
GJD	0.062***		0.034***	0.043***	0.930***	0.055***	-1.133***	3.014***
	(0.022)		(0.008)	(0.005)	(0.008)	(0.017)	(0.399)	(0.377)

Note: Standard errors in parentheses. Sample period: start of data reported in Table 1 through 12/31/2004. Number of observations: 1,842 for Soybeans, 2,028 for corn. Asterisks signify statistical significance: *: better than 10% level; **: better than 5% level; ***: better than 1% level.

Table 5: Estimation of the Model Parameters for Daily price returns: Post - June, 2014

	μ	σ	κ	α_1	β_1	λ	θ	δ
				A. Brazilian Soybeans				
PD	-1.460*	1.186						
	(0.843)	(2.154)						
JD	-0.027	0.985***				0.293	0.087	1.216***
	(0.033)	(0.065)				(0.197)	(0.113)	(0.256)
GPD	-0.010		0.019***	0.031***	0.955***			
	(0.023)		(0.008)	(0.007)	(0.010)			
GJD	-0.028		0.006	0.030***	0.958***	0.101*	1.158	1.525***
	(0.026)		(0.005)	(0.008)	(0.011)	(0.053)	(0.194)	(0.281)
				C. Brazilian Ethanol				
PD	-0.005	1.506***						
	(0.032)	(0.022)						
JD	-0.027	1.124***				0.210***	0.103	2.174***
	(0.036)	(0.046)				(0.062)	(0.193)	(0.250)
GPD	0.021		0.337***	0.180***	0.839***			
	(0.030)		(0.106)	(0.039)	(0.025)			
GJD	0.007		0.050**	0.063***	0.618***	0.101***	0.142	2.284***
	(0.030)		(0.021)	(0.016)	(0.080)	(0.037)	(0.233)	(0.346)
				D. US Soybean				
PD	-0.003	1.369***						
	(0.018)	(0.020)						
JD	0.0387	1.056***				0.114***	-0.364	2.501***
	(0.027)	(0.055)				(0.058)	(0.268)	(0.509)
GPD	0.015		0.026***	0.072***	0.916***			
	(0.023)		(0.007)	(0.010)	(0.010)			
GJD	0.018		0.029***	0.043***	0.930***	0.027***	-1.011	3.657***
	(0.024)		(0.008)	(0.007)	(0.011)	(0.011)	(0.689)	(0.650)

Note: Standard errors in parentheses. Sample period: 1/1/2005 through end of data reported in Table 1. Number of observations: 2,998 for Soybeans, 3,061 for corn. Asterisks signify statistical significance: *: better than 10% level; **: better than 5% level; ***: better than 1% level.

7. Appendix: Details on Brazilian Data Processes

In the first part of the Appendix, we provide additional detail on the Brazilian soybean and ethanol data.[22] 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

[22] 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`.

20 days' coefficients of variation. If the current day's CV is above this, the current day's 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.[23] 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).

[23] 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`.