Jump Processes in the Market for Crude Oil

Neil Wilmot[1] and Charles F. Mason[2]

April 4, 2011

Abstract

In many commodity markets, the arrival of new information leads to unexpectedly rapid changes – or jumps – in commodity prices. Such arrivals suggest the assumption that log-return relatives are normally distributed may not hold. This article investigates the potential presence of such jumps in the price of crude oil, both in terms of spot prices and the futures prices. The investigation is carried out over three data frequencies (*Monthly, Weekly, Daily*), which allows for an investigation of temporal properties. Based on the methodology employed, likelihood ratio tests are used to compare among four stochastic data-generating processes. Maximum-likelihood estimation results suggest that jumps are important when examining high frequency (*Daily*) data, but the significance of jumps is 'washed out' at lower frequencies (*Monthly*). However, allowing for time-varying volatility calls into question the empirical relevance of jumps, even for higher frequency data.

Keywords: Crude Oil prices; Structural breaks; jump diffusion; GARCH

JEL Classification: Q40, G12, G13

[1] Corresponding Author; Department of Economics, University of Minnesota Duluth, 1318 Kirby Drive, Duluth MN 55812; phone: 1-218-726-7439; fax: 1-218-726-6509; email: nwilmot@d.umn.edu

[2] H.A.True Chair in Petroleum and Natural Gas Economics, Department of Economics & Finance, University of Wyoming, 1000 E. University Ave., Laramie, WY 82071; e-mail: bambuzlr@uwyo.edu

1. Introduction

A central characteristic of many natural resources prices is the intrinsic stochastic element driving the pricing process. Understanding this underlying stochastic process is of clear importance, particularly for crude oil due to its essential role in the world economy (Hamilton, 2008; Huntington, 2007; Jimenez-Rodriguez and Sanchez, 2005). Yet changes in oil prices continue to catch both experts and consumers by surprise (Wirl, 2008). From a firm's perspective, since investment behavior is directly tied to the market price of output, the stochastic element of natural resource prices can have important implications on the decision to undertake investment (Postali and Picchetti, 2006). As well, policymakers must be aware of the economic impacts to a domestic economy from a dynamic and recently volatile crude oil price.

Much of the early literature on modeling petroleum prices assumes the market price of crude oil follows a continuous stochastic process that assumes smooth changes, either in a single factor or multi-factor Gaussian framework (Schwartz, 1997; Pindyck, 1999; Sadorsky, 1999; Schwartz and Smith 2000; Casassus and Collin-Dufresne, 2005; Cortazar and Naranjo, 2006). The assumption of continuity has provided researchers with tractable models that typically facilitate closed form solutions, especially in the real options literature, where interest is focused on determining the optimal timing of an irreversible investment.[3] However, as described by Wirl (2008), such continuous time models do not pass the modest test of econo-physics because the processes fail to reproduce the relatively fat tails observed in a distribution of returns – a common feature of financial data. Furthermore, Cont and Tankov (2004) argue continuous stochastic processes are insufficient because they fail to reproduce the most important feature observed in markets: *discontinuous moves in price*. This feature has particular potential importance for crude oil markets as they are frequently hit with unexpected news such as natural disasters (hurricanes, earthquakes), geopolitical developments (nationalization, strikes) strategic actions (OPEC), and other unforeseen events (spills, pipeline disruptions). For example, the effects of hurricane Katrina forced the closure of nine oil refiners and a significant movement was observed in both the crude oil spot and futures prices. These sorts of effects can lead to unexpectedly large changes in oil prices, either upwards or downwards. Because they cannot

[3] Postali and Picchetti, 2006; Fackler, 2006; Conrad and Kotani, 2005; Baker *et al*, 1998; Paddock *et al*, 1988.

generate these "jumps", the empirical relevance of continuously evolving diffusion models is suspect.

Recent research has examined the relevance of discontinuities for modeling oil prices, with conflicting results. Askari and Krichene (2008) use daily observations to investigate several jump-diffusion processes over the relatively short and volatile period of 2002 – 2006. The findings indicate oil prices are highly sensitive to unexpected news and other shocks. Lee *et al* (2010) develop what is referred to as a component – ARJI model, which allows for jumps in price as well as decomposing the conditional variance into two components; a transitory component and a permanent component. The analysis utilizes daily crude prices, and confirms the presence of the two components and suggests that the transitory component is the main factor responsible for influencing jumps in price returns. On the other hand, Postali and Pichetti (2006) present numerous stochastic processes for the price path of crude oil, including both continuous and jump processes. Using annual crude oil prices, the authors argue that the choice of a geometric Brownian motion process is a sufficient proxy for the data generating process. It would appears that the frequency of the observations plays a crucial role in determining the appropriate data generating process for crude oil prices.

In this study we investigate the price path of oil across several data frequencies, while allowing for the potential presence of jumps. To this end, we model the price processes for crude oil spot and futures prices as alternatively continuous and discontinuous processes. The discontinuous process is assumed to follow the mixed jump-diffusion process, similar to that of Merton (1976). Because the fat-tail result we alluded to above may be driven by the sort of volatility clustering associated with the time-varying volatility in commodity prices (Pindyck, 2004), the continuous model we investigate uses the GARCH framework (Bollerslev, 1986). While previous studies have applied generalized autoregressive models to oil prices exist (Li and Thompson, 2010; Narayan and Narayan, 2007; Moshiri and Foroutan, 2006), such studies do not incorporate discontinuities in the pricing process. Allowing for the potential interaction of the GARCH and jump diffusion models, we then have four potential processes: continuous stochastic diffusion, jump diffusion, continuous diffusion with GARCH, and jump diffusion with GARCH, which we estimate using maximum likelihood methods. The nested feature of the ML framework allows for direct tests among the models utilizing the Likelihood ratio (*LR*) test. Of particular interest is the empirical performance of the continuous model relative to the discontinuous model. We observe that the explicit modeling of discontinuities significantly improves the models fit. Moreover, accounting for jumps appears more important for higher-frequency data (*daily, weekly*) than with low frequency data (*monthly*). This implies the rate of temporal aggregation of a particular variable under study would significantly influence the modeler's conclusion with regards to the 'best-fitting' model. However, we find that the inclusion of time-varying volatility significantly improves the models fit, across all data frequencies, as well as calling into questioning the empirical relevance of jumps.

The structure of this paper is as follows. The next section presents the econometric framework used to estimate the parameters describing the price path for the various stochastic processes. Section 3 describes the oil price data that are to be investigated. In this paper we take investigate three data frequencies, while the presentation focuses on a long horizon of daily oil prices from three crude oil price series. As well, several important statistical properties of the data are examined. In particular, several unit root tests are presented and the results examined and contrasted with the results of several alternative studies. The empirical results of the maximum likelihood estimation, applied to the various crude oil spot and futures series are presented in Section 4. Finally, Section 5 presents some concluding remarks.

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*, which is said to follow a geometric Brownian motion (*GBM*) process with trend α and variance parameter σ if

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

In equation (1), *dz* represent an increment of a Wiener process. Following Dixit and Pindyck (1994), the increment of a Wiener process in continuous time is given by

$$dz = \xi_t \sqrt{dt}, \tag{2}$$

where ξ_t has zero mean and a standard deviation equal to 1. Let x_t denote the natural logarithm of the ratio of price in period *t* to the price in period *t*-1, $x_t \equiv \ln(P_t) - \ln(P_{t-1})$. 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

$$x_t = \mu + \sigma z_t. \tag{3}$$

The term z_t in equation (3) is an identically and independently distributed (i.i.d.) random variable with mean zero and variance one. Jumps are introduced into the model in the style of Merton (1976), so that when an 'event' occurs, the jump component is modeled as a Poisson-driven process *q*, where

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

If a jump does occur at time *t*, then its size is $dq_t = Y_t - 1$ where $(Y_t - 1)$ is a random variable, which describes the percentage change in the asset return.[4] The jumps may take either sign, and their magnitude is itself a random variable. The resultant stochastic process for the random variable P_t may then be written as

$$\frac{dP_t}{P_t} = \alpha dt + \sigma dz_t + dq_t, \tag{5}$$

where dz_t has the same properties assumed in equation (1) and dq_t is the independent Poisson process described in equation (4). Together the terms dz_t and dq_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.

Jumps are assumed to arrive at any moment *t* with probability λ, so that $1 - \lambda$ is the probability no jump occurs at time *t*. Jump sizes are assumed to be lognormally distributed with

[4] Merton (1976) describes *(Y-1)* as an impulse function that produces a finite jump in *S* to *SY*.

θ as the mean of the logarithm of the jump size and δ as the standard deviation of the logarithm of the jump size; hence $\ln(Y_t) \sim N(\theta, \delta^2)$. If no jump occurs at time t then $dq_t = 0$. Using these notational conventions, we may express the mixed jump-diffusion (*JD*) process as

$$x_t = \mu + \sigma z_t + \ln(Y_t) J_t, \tag{6}$$

where again $z_t \sim N(0,1)$. In equation (6), J_t is a Bernoulli random variable which takes on the value 1 when a jump occurs and 0 otherwise. Referring back to equation (4), we note that if $\lambda = 0$ then jumps never occur, and so equation (6) reduces to the pure diffusion model of equation (3).

An alternative explanation for the commonly observed "fat tails" is that P_t follows a time-varying error process. Maximum likelihood estimation can be used to estimate models that incorporate the generalized autoregressive conditional heteroskedastic (GARCH) framework. A general consensus in the literature is that a GARCH model, with a limited number of terms, performs reasonably well, and so we limit our investigation to the GARCH(1,1) process. Adopting this convention, we have the GARCH – diffusion (*GPD*) process, which can be described by:

$$x_t = \mu + \sqrt{h_t} z_t, \tag{7}$$

where the conditional variance, h_t is described by the process

$$h_t \equiv E_{t-1}(\sigma^2) = \kappa + \alpha_1 (x_{t-1} - \mu)^2 + \beta_1 h_{t-1}. \tag{8}$$

Note that when $h_t = \sigma^2$ the GARCH diffusion model reduces to pure diffusion model. When $\kappa > 0$ and $\alpha_1 + \beta_1 < 1$, the unconditional variance of the volatility of the process exists and equals $\kappa/(1-\alpha_1-\beta_1)$. Allowing for jump discontinuities would result in the GARCH(1,1) jump-diffusion (*GJD*) process:

$$x_t = \mu + \sqrt{h_t} z_t + (y_t) J_t, \tag{9}$$

where h_t is described by equation (8). 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.

The parameters of these four models (*PD, JD, GPD* and *GJD*) are estimated using the numeric maximum likelihood estimation methods, based on the observations x_t, $t = 1,\ldots,T$. Estimation involves maximizing the general log-likelihood function

$$L(\phi;x_t) = \sum_{t=1}^{T} \ln\left\{ \frac{1-\lambda}{\sqrt{2\pi h_t}} \exp\left[-\frac{1}{2}\frac{(x_t-\mu)^2}{h_t}\right] + \frac{\lambda}{\sqrt{2\pi\left(h_t+\delta^2\right)}} \exp\left[-\frac{1}{2}\frac{(x_t-\mu-\theta)^2}{\left(h_t+\delta^2\right)}\right]\right\}, \qquad (10)$$

with respect to the parameter space $\phi \in \mathbf{\Phi}$, $\phi = (\mu,\kappa,\alpha,\beta,\lambda,\theta,\delta)$. The pure diffusion model is estimated from equation (10) with $\lambda = 0$ and $h_t = \sigma^2$, while the mixed jump-diffusion can be estimated with the restriction $h_t = \sigma^2$. Maximum likelihood estimation provides a number of advantages in this context.

Maximum likelihood estimates are known to be consistent and invariant with asymptotically normal distributions of the parameters. To allow for comparisons among the different models, we use a likelihood ratio test (Johnston and DiNardo, 1997). This approach compares the likelihood function under a particular restriction, $L(\tilde{\phi};x)$, to that of the unrestricted likelihood function, $L(\hat{\phi};x)$. 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, such as the jump diffusion model, and a more restricted model, e.g. the pure diffusion model.[5] The test statistic is the log-likelihood ratio

---

[5] The parameter restriction in this case would be $(\lambda, \theta, \delta^2) = (0, 0, 0)$.

$$LR = 2\left[\ln L\left(\hat{\phi}; x\right) - \ln L\left(\tilde{\phi}; x\right)\right]; \tag{11}$$

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

3. <u>Data and Its Properties</u>

3.1. *Data*

The data for this study consists of the daily closing spot price of West Texas Intermediate (*WTI*) crude oil, Brent crude oil and the 1-month futures price of West Texas Intermediate crude. The data is available from the Energy Information Administration's website (Energy Information Administration, 2010a, 2010b). The futures data is based on the NYMEX Division light, sweet crude futures contract. According to the Chicago Mercantile Exchange, crude oil is the most actively traded commodity in the world. The international pricing benchmark for crude oil has become the New York Mercantile Exchange's (NYMEX) Division light, sweet crude oil futures contract. As the world's largest-volume futures contract of all commodities, it is the world's forum for trading crude oil. These futures contracts trade in units of 1,000 barrels, with a delivery point in Cushing, Oklahoma. The futures contract price obtained is for the earliest delivery date, or what is referred to as the front month. This series is the longest, with price information available from early 1983 through to August 2010. The spot price of *WTI* covers a period of almost 22 years, beginning in January of 1986. The final series, the spot price of Brent oil, begins in late May of 1987 and runs to August 2010. The period of study presents a number of events which, a priori, would be good candidates for arrivals of "unusual" information, such as Black Monday (1987), the Gulf war (1990), Y2K concerns (1999), the New York terrorist attacks on the World Trade Center (2001) and the devastating effects of hurricane Katrina (2005). This period also includes relatively recent increases in the price which have seen crude reach nominal record prices and a period in which oil prices have more than tripled, as well as the recent downturn in the global economy. In this context, the price returns of the individual series are calculated as

$$r_t = 100\left[\ln\left(P_t / P_{t-1}\right)\right], \tag{12}$$

where P_t is the closing price on day *t*.

Summary statistics for the relative price returns of each of the three crude oil series is given in Table 1, for three different frequencies: daily, weekly and monthly. Each series displays a significant amount of variation and some evidence of asymmetry in the distribution, as displayed by the presence of the negative skewness. Each of the series displays evidence of leptokurtosis or “fat tails” by the large value for kurtosis. These results are consistent with the results of the Jarque-Bera (*JB*) test, which measures the departure from normality. In eight out of the nine cases, the null hypothesis of a normally distributed random variable is rejected at the standard levels of significance.

INSERT TABLE 1 ABOUT HERE

A histogram of realized log price changes is shown in Figure 1 for the spot price of (*WTI*) crude oil. The figure includes a normal probability distribution (superimposed) with an identical mean and variance to the log changes in the spot price. If the relative prices follow a geometric Brownian motion process, the histogram should fit the probability distribution well. Noticeably, the figure points to the existence of a significant number of large changes, especially apparent in the extremes of both tails, which has the potential to influence the trigger value at which the firm holding a real option should optimally invest. The perceived leptokurtosis reinforces the inadequacy of the continuous process to fit the data, reaffirming the notion of a discontinuous process in the data-generating process. An analogous depiction of the two remaining series demonstrates the common feature of large changes beyond those expected from a normal distribution.

INSERT FIGURE 1 ABOUT HERE

3.2. *Stationarity*

A key concern with time series data is the potential existence of a unit root (i.e., non-stationarity). Time series data that have a unit root contain a stochastic trend, which has the potential to cause spurious results (Maddala and Kim, 1998). While the presence of a unit root in the price of crude oil has been investigated by numerous authors[6], a potentially significant limitation of these previous studies is the possibility that a structural change has been neglected. Perron (1989) describes how the presence of a structural change may reduce the power of the test to reject the null hypothesis of a unit root, when in fact it should be rejected. To allow for endogenous structural breaks, Lee and Strazicich (2003, 2004) extend the Lagrangian multiplier (*LM*) unit root test of Schmidt and Phillips (1992). The advantage of the *LM* test is that the rejection of the null hypothesis is unaffected by the existence of breaks. In applying the LM unit root test, the null hypothesis cannot be rejected when examining weekly crude oil prices (Maslyuk and Smyth, 2008) while at the annual level, the results is not as robust. Postali and Picchetti (2006) are able to reject the null hypothesis while Lee *et al* (2006) find mixed evidence of a unit.

In light of the mixed evidence related to the temporal resolution of price data in the extant literature, we examine three data frequencies (monthly, weekly and daily) for crude oil spot and futures prices to investigate the presence of unit root.[7] The analysis of the unit root test with endogenous structural breaks follows the methodology of Lee *et al* (2006), which includes a quadratic trend, with the belief that a quadratic trend may exist in some natural resource prices. Define y_t as the commodity price in period *t*, where $t = 1,...,T$, the data-generating process can be described as

$$y_t = \delta' Z_t + e_t, \qquad (13)$$

where $e_t = be_{t-1} + u_t$, and Z_t is a matrix of exogenous variables which include a constant term, linear time trend, *t*, and two structural breaks in level and trend, such that $Z_t = \left[1, t, D_{1t}, D_{2t}, DT_{1t}, DT_{2t}\right]$. The terms D_{jt}, for $j = 1,2$ allow for two shifts in the intercept term, and take on the value of 1 for $t \geq T_{Bj} + 1$, $j = 1,2$ and 0 otherwise, where T_B is the date of the structural break. The model also allows for two changes in the slope of the trend, described by DT_{jt}, $j = 1,2$, where $DT_{jt} = t - T_{Bj}$ for $t \geq T_{Bj} + 1$, $j = 1,2$ and 0 otherwise. This leads to the following set of hypotheses:

[6] Sivapulle and Moosa, 1999; Serletis and Rangel-Ruiz, 2004; Sadorsky, 1999; Coimbra and Esteves, 2004
[7] Traditional unit root tests with no structural breaks indicate that oil prices are best modeled as non-stationary, but that first-differences in oil prices are stationary; results are available from the authors upon request.

$$H_0 : y_t = \mu_0 + d_1 B_{1t} + d_2 B_{2t} + d_3 D_{1t} + d_4 D_{2t} + y_{t-1} + v_{1t}$$
$$H_A : y_t = \mu_1 + \gamma t + d_1 D_{1t} + d_2 D_{2t} + d_3 DT_{1t} + d_4 DT_{2t} + v_{2t}. \quad (14)$$

To investigate the presence of a quadratic trend and endogenous structural breaks, we include the quadratic term t^2, which gives $Z_t = \left[1, t, t^2, D_{1t}, D_{2t}, DT_{1t}, DT_{2t}\right]$.

The test statistics for the *LM* unit root test is obtained from the following regression:

$$\Delta y_t = \delta' \Delta Z_t + \phi \tilde{S}_{t-1} + u_t, \quad (15)$$

where $\tilde{S}_{t-1} = y_t - \tilde{\psi}_x - Z_t \tilde{\delta},\ t = 2, \ldots, T$. The *LM* unit root test compares the null hypothesis $\phi = 0$ against the alternative hypothesis $\phi < 0$ using a t-statistic. To determine the location of the breaks, $\left(\varpi_{ij} = T_{Bj} / T, j = 1, 2\right)$, a grid search is undertaken to determine where the value of $\tilde{\tau}$ is at its minimum:

$$\text{LM}_t = \inf_{\varpi} \tilde{\tau}\left(\varpi\right). \quad (16)$$

The search is carried out over a subset $\left(0.15T, 0.85T\right)$ of the entire sample, *T*.

Table 2 reports results from the LM unit root tests when one allows for structural breaks, along with the dates of the breaks. We consider four combinations, reflecting the potential for either one or two breaks, and the either a linear or quadratic trend. The null hypothesis of a unit root in the presence of one endogenous structural break cannot be rejected, in either the linear or quadratic trend cases. The indicated time of the break is approximately the middle of 2004 through early 2005, with the exception of weekly futures prices with a linear trend. The dates of the structural breaks in WTI spot prices, determined in the one-break case are presented in panel a of Figure 2 .[8] Noticeably, this time also corresponds to the 'beginning' of a strong upward trend in oil prices which would eventually lead to a nominal record high. This period corresponds to a time of rising global demand for crude oil products, and follows closely on the start of the Iraq war (Energy Information Administration, 2009). When two endogenous breaks

[8] In looking at Figure 2, one might expect a break to occur at the peak in prices around mid 2008, prior to the steep decline, associated with the global decline in economic activity. However, this date falls outside the search area, which was defined as (0.15T, 0.85T).

are considered, with either a linear or quadratic trend, the unit root hypothesis cannot be rejected. The dates of the breaks, presented in Table 2, experience more variation relative to those observed in the one break case. For spot prices, the inclusion of a linear trend, results in the first break occurring in late 2002 or 2003, while the second break typically occurs in 2006. The timing of the first break appears to coincide with a period of geopolitical instability resulting from concerns regarding the Iraq war and a strike in Venezuela. The second break coincides with a number of geopolitical events as well as natural disasters.[9] The futures prices consistently experience the first break in October of 1997, a period which corresponds to a period of uncertainty in the supply of crude[10], while the second break occurs during a period of geopolitical instability resulting from concerns regarding the Iraq war, among other concerns. The inclusion of a quadratic trend results in breaks that are relatively less dispersed. The first break occurs consistently at the end of 2003 or early 2004, which correspond to those of the linear one break case, and would therefore have similar causes. The second break is reported as the spring or summer of 2006, which corresponds to the closure of an Alaskan pipeline by British Petroleum in March.

As Maslyuk and Smyth (2008) suggest, the fact that crude oil prices are nonstationary means that the assumption of a geometric Brownian motion process maybe sufficient for a firm looking to avoid large errors in the optimal investment decision. However, the finding of nonstationary does not rule out the potential for large, discontinuous moves, which can also affect the optimal timing of investment. As such, investigation the potential presence of such jumps would seem to be prudent; a task we now turn to.

INSERT TABLE 2 ABOUT HERE

INSERT FIGURE 2 ABOUT HERE

4. Empirical Results

4.1. *Maximum Likelihood Estimation*

[9] Such geopolitical events would include an attack on refineries in Saudi Arabia and conflict in the Niger Delta Region. Also, in early 2006 the Department of Energy released oil from the strategy reserve following hurricanes Katrina and Rita. Also, in March a leak forced BP to shutdown its Prudhoe Bay pipeline.

[10] During this period, attacks to a pipeline in Columbia appeared likely to reduce supply, and the UN's oil for food program in Iraq was starting to take shape. In addition, later in 1997 OPEC agreed to increase its production quotas for the first time in four years.

Using Maximum likelihood methods, we estimate the four stochastic processes (*PD, JD, GPD, GJD*) for time series based on three levels of temporal aggregation (monthly, weekly and daily).[11] Monthly results are reported in Table 3, weekly results are reported in Table 4, and daily results are reported in Table 5. The monthly *PD* model results indicate the drift parameter, μ is not significantly different from zero, while the instantaneous rate of variance, σ, is statistically significant. The results for μ and σ in the mixed jump-diffusion model (*JD*), are similar, however, the inclusion of the jump component into the model noticeably reduces the instantaneous rate of variance. This reduction in σ is offset by a much larger and significant value of δ, the variance of the jumps. While statistically insignificant, the mean size of the jump, θ, is consistently negative across the three series. However, the probability of a jump, λ is not significant for any of the data series. The GARCH(1,1) model (*GPD*) provides variance parameter estimates $\left(\hat{\alpha}_1, \hat{\beta}_1\right)$ that are less than one in magnitude, and generally statistically smaller than one.[12] Finally, the combination of jumps and time varying volatility is examined with the results of the individual models maintained. Significance is observed among the time-varying volatility terms, while the probability of a jump is remains insignificant.

INSERT TABLE 3 ABOUT HERE (Monthly)

The *PD* estimates based on weekly data, presented in Table 4, provide similar results to those observed for monthly data. Prominently, the *JD* terms are consistently and highly significant, with similar signs as those previously observed. Our results suggest that the *WTI* spot price experiences a jump approximately every 6 months, while *Brent* spot and *WTI* futures prices experience jumps approximately every 9 and 7.5 months, respectively. The *GPD* model provides results that approximate the monthly results, with variance parameter estimates that are statistically significant, and whose sum is close to one both numerically and statistically. However, the *GJD* model provides perhaps the most noteworthy results. The estimated volatility parameters are significant, with $\hat{\alpha}_1 + \hat{\beta}_1$ is strictly less than one, and the jump terms are

[11] Estimation was undertaken in GAUSS, utilizing the constrained maximum likelihood module. The code to estimate the GARCH model was obtained from Schoenberg (2006).

[12] Wang (2003) indicates that even if the difference is not statistically significant, the standard asympototically based test are generally valid.

significant, with the probability of a jump occurring indicating that the expected time between 'jumps' has increased, relative to the *JD* results. This suggests that when one allows for time-varying volatility, some of the observations that were classified as jumps in the *JD* model are reclassified, with the apparently large variation incorporated into the variance component (as suggested by the larger jump variance terms). Moreover, the variance of the jump component, δ, is larger in the *GJD* model than in the *JD* model.

INSERT TABLE 4 ABOUT HERE (Weekly)

An important difference is apparent when comparing the results of weekly and monthly data. As the time between observations is increased, the variance of the returns increases, as shown previously in Table 1. This is further reflected in the larger instantaneous variance estimates as well as the estimated variance of the jumps. In turn, this leads to a reduction in the magnitude and significance of the parameter estimates in the models containing jumps, recognized by the insignificance of the probability of a jump occurring, λ. One plausible explanation for this set of results is that aggregation of data to the monthly level induces a loss of information, relative to the more frequent daily observations. The increased level of volatility would then make it hard to 'pick out' a jump relative to large continuous movement. The results of the highest observation frequency (daily), presented in Table 5, substantially reinforce this idea. The significance of the parameter estimates mirror the weekly estimates previous reported. We note several important results, in terms of the daily observations. The jump probability, λ, in the mixed jump-diffusion model, is consistently significant at the 1% level, and indicate a jump is expected to occur approximately every 8 days for *WTI* spot and futures returns, and every 11 days for *Brent* returns. In the both the *GPD* and *GJD* model, the variance terms are significantly different from zero, and sum to less than one. The jump terms, again shows significance and indicate a longer time between jump arrivals relative to the *JD* results.

INSERT TABLE 5 ABOUT HERE (Daily)

4.2 *Likelihood Ratio Tests*

Since the models are nested, the Likelihood ratio (*LR*) test can be used to test whether the hypothesized restrictions are valid. In comparing the geometric Brownian motion model with that of the mixed jump-diffusion process, the restriction (null hypothesis) is that the parameters λ, θ, δ of the *JD* model are equal to zero. Alternatively, the mixed jump-diffusion model restricts the volatility term relative to both the unrestricted GARCH models, *GJD* and *GPD*. The results of the *LR* test, support the presence of jumps over the geometric Brownian motion (*PD*) process, since the null hypothesis is rejected. Similarly, the GARCH process with (*GJD*) and without jumps (*GPD*) fits significantly better than the *PD* process. However, the *LR* tests support the time-varying volatility model over a model that imposes a constant variance rate.[13] Finally, the *LR* test results for the remaining data frequencies, weekly and monthly observations, are generally consistent with those of the daily observations. Again, one exception arises in the monthly results, where the restrictions imposed by the *JD*, relative to the unrestricted *GJD* case cannot be rejected for both the *Brent* and *Futures* return series. While the mixed jump-diffusion process appears to be a better fit than the geometric Brownian motion process, the greater flexibility in the error structure of a GARCH(1,1) process seems to be outperforming both the *PD* and *JD* models.[14]

4.3 *Effect of a Structural Break on the Parameter Estimates*

We next investigate how well the models perform over two subsamples that are suggested by the structural break analysis presented above. For this purpose, we use the dates obtained from the scenario with one structural break and a quadratic trend, and focus on daily observations.[15] The pre-break and post-break results are presented in Tables 6 and 7, respectively. The results of the pre-break data are very similar to those presented for the entire

---

[13] The *LR* test fails to reject the null hypothesis that $\lambda = 0$, indicating that the unrestricted GARCH model (GJD), which allows for jumps in a time-varying volatility framework, does not improve the fit over the restricted GARCH model (GPD), which does not include jumps.

[14] As an additional means of comparing the models, the Schwartz criterion was calculated for each model / series / frequency combination. The results, which are available on request, replicate those of the Likelihood ratio tests

[15] The presentation is limited to daily observations since both subsamples were large enough to enable the models to solve. For the remaining frequencies, the sizes of the post-break samples are inadequate to allow some models to convergence. For example, the *WTI* futures data has 64 monthly and 288 weekly observations in the post-break periods.

sample. Note here that the trend is generally smaller, which is not unexpected since the pre-break data does not include the period of rapidly increasing oil prices, from 2005 through 2008.

INSERT TABLE 6 ABOUT HERE (prebreak)

In examining the post-break estimates, a number of interesting results emerge. The trend observed in geometric Brownian motion process is positive and larger in magnitude than that observed previously, while remaining insignificant. The instantaneous rate of variance is significant and larger than the full sample results, indicating a larger amount of variation in the post-break results. In the mixed jump-diffusion case, the probability of a jump, and the variance of the jump sizes are approximately the same, while the average jump size appears much smaller and, in some cases, has changed sign. However, in each case the term is not significantly different from zero. Finally, in examining the *GJD* model, we note the probability of a jump is noticeably lower in the post-break period while remaining significant. As well, the variance of the jump size is larger and significant, as are the estimated values of both α and β. The results of the *LR* tests for both subsamples mirror those obtained for the full sample. Such results indicate the necessity of including time-varying volatility, as well as the superiority of the mixed jump-diffusion process over the pure diffusion process.

INSERT TABLE 7 ABOUT HERE (postbreak)

5. Conclusion

There are many reasons why a better understanding of the stochastic process driving the price of crude oil and its derivatives would be useful. For example, oil prices can have important microeconomic effects with commodity price risk having a potentially significant impact on a firm's profit. Knowledge of the underlying stochastic behavior of the underlying asset will aid in crude oil forecasting, investment decisions as well as the pricing of new oil-linked financial instruments. Furthermore, events in the market for oil are important for policymakers since oil price shocks are often followed by economic downturns (Hamilton, 2011). In this paper, we re-examine the assumption that the relative price returns of crude oil can be modeled using a continuous time process, employing daily data. This assumption of continuity is contrasted to a model that explicitly allows for discontinuities. Additionally, the potential presence and importance of volatility clustering, both in a continuous and discontinuous framework is examined. These elements lead us to investigate four stochastic processes to determine the appropriate data generating mechanism describing the evolution of crude oil price returns.

We draw several important conclusions from our analysis. The mixed jump diffusion model is preferred to the pure diffusion model, which is confirmed by the results of the Likelihood ratio tests. In examining the parameter estimates the importance of including 'jumps' in a model of crude oil returns is apparent. The probability of a jump occurring and the jump variance are consistently significant, particularly at the higher data frequencies. As well, we note that the arrival of information tends to lower oil prices on average. While including jumps is important, the results indicate that a model must also allow for time-varying volatility, a result reinforced by the *LR* tests. Furthermore, based on the results of the daily estimates, a model which includes both jumps and time-varying volatility seems to provide the most appealing results, with a high degree of significance across the jump and volatility parameters estimates. While intuitively this combined model seems the most appropriate, the results is not confirmed by the *LR* test. The sample period was also divided into two subsamples based on the results of the *LM* unit root test, which allowed for both a quadratic trend and endogenously determined structural breaks. With the division determined to be in the middle of 2004, the relative performance of the four models was maintained in both the pre- and post- break periods.

While a number of existing studies have used continuous process to model oil prices, and find that they are a good fit, these earlier studies generally used temporally aggregated data. We find that converting daily data to a more aggregated form, be it weekly, monthly or yearly, causes a loss in the ability to 'capture' jumps. One interpretation of these findings is that the higher volatility observed in data reflecting greater degrees of aggregation frequency data has a tendency to 'wash out' jumps – that is, such aggregation causes a reduction in power, which adversely impacts the ability to capture the discontinuous moves in oil prices. This interpretation has implications for the use of annual data, which has been used to investigate stochastic processes in the extant literature, and where the typical result indicates that smoothly evolving stochastic processes perform well. Our findings suggest that such results may be induced by the aggregation of a discontinuous process. We leave for future work what the implications of such an aggregated process might involve.

References:

Askari, H. and N. Krichene (2008). “Oil price dynamics (2002 – 2006).” Energy Economics 30 (5): 2134–2153.

Baker, M., Mayfield, S. and Parsons, J. 1998. “Alternative Models of Uncertain Commodity Prices for use with Modern Asset Pricing Methods,” *Energy Journal* 19: 115 – 148

Bollerslev, T. (1986) “Generalized Autoregressive Conditional Heteroskedasticity.” Journal of Econometrics 31: 307–327.

Casassus, J. and P. Collin-Dufresne (2005) “Stochastic Convenience Yield Implied from Commodity Futures and Interest Rates.” Journal of Finance 60(5): 2283-2331.

Coimbra, C. and P.S. Esteves (2004) “Oil Price Assumptions in Macroeconomic Forecasts: Should we Follow Futures Market Expectations?” OPEC Review 28: 87–106.

Conrad, J. M. and Kotani, K. 2005. “When to drill? Trigger Prices for the Arctic National Wildlife Refuge,” *Resource and Energy Economics* 27: 273 – 286

Cont, R. and P. Tankov (2004) Financial Modeling with Jump Processes. Boca Raton: Chapman - Hall

Cortazar, G. and L. Naranjo (2006) “An N-factor Gaussian Model of Oil Futures Prices.” Journal of Futures Markets 26 (3): 243–268.

Dixit, A. and R.S. Pindyck (1994) Investment under Uncertainty. Princeton, New Jersey: Princeton University Press.

Duan, J. C. (1997) “Augmented GARCH(*p,q)* Process and its Diffusion Limit.” Journal of Econometrics 79: 97–127.

Energy Information Administration, 2009. “Annual Oil Market Chronology.” Internet website: http://www.eia.doe.gov/cabs/AOMC/Overview.html . Downloaded 10/30/2009

Energy Information Administration, 2010a. “Spot Prices for Crude oil.” Internet website: http://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm. Downloaded 10/20/10

Energy Information Administration, 2010b. “NYMEX Futures Prices for Crude oil.” Internet website: http://www.eia.gov/dnav/pet/pet_pri_fut_s1_d.htm. Downloaded 10/20/10

Fackler, P. 2007. “Comment on Conrad and Kotani,” *Resource and Energy Economics* 29(2): 159 – 164

Hamilton, J.D. (2008) “Oil and the Macroeconomy.” In Durlauf, S. and L. Blume, eds., *New Palgrave Dictionary of Economics*, 2nd edition. Houndmills, UK: Palgrave McMillan Ltd.

Hamilton, J.D. (2011) "Historical Oil Shocks." University of California Center for Energy and Environmental Economics, Working Paper (WP-005). Available at http://uce3.berkeley.edu/WP_005.pdf

Huntington, H. G. (2007) "Oil Shocks and Real U.S. Income." The Energy Journal 28(4): 31–46.

Jimenez-Rodriguez, R., and M. Sanchez (2005) "Oil Price Shocks and Real GDP Growth: Empirical Evidence for Some OECD Countries." Applied Economics 37 (2): 201-228.

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

Lee, J. and M.C. Strazicich (2003) "Minimum Lagrange Multiplier Unit Root Test with Two Structural Breaks." The Review of Economics and Statistics 85(4): 1082–1089.

______ and M.C. Strazicich (2004) "Minimum Lagrange Multiplier Unit Root Test with One Structural Breaks." Working Paper, Department of Economics, Appalachian State University

______, J.A. List, and M.C. Strazicich (2006) "Non-renewable resource prices: Deterministic of Stochastic Trends?" Journal of Environmental Economics and Management 51: 354–370.

Lee, Y.H., H.N. Hu and J.S. Chiou (2010) "Jump Dynamics with Structural Breaks for Crude Oil Prices." Energy Economics 32(2): 343–350.

Li, J. and H. Thompson. (2010) "A Note on the Oil Price Trend and GARCH Shocks." The Energy Journal 31(3): 159-165.

MacKinnon, J. G. (1996) "Numerical Distribution Functions for Unit Roots and Cointegration Tests." Journal of Applied Econometrics 11: 601–618.

Maddala, G. S., and I. Kim (1998) Unit Roots, Cointegration and Structural Change. Cambridge, UK: Cambridge University Press.

Maslyuk, S. and R. Smyth (2008) "Unit Root Properties of Crude Oil Spot and Futures Prices." Energy Policy 36: 2591–2600.

Merton, R.C. (1976) "Option Pricing when the Underlying Stock Returns are Discontinuous." Journal of Financial Economics 3(1-2): 125–144.

Moshiri , S. and F. Foroutan (2006) "Forecasting Nonlinear Crude Oil Futures Prices." The Energy Journal 27(4): 81–95.

Narayan, P. K., and S. Narayan (2007) "Modelling oil price volatility." Energy Policy 35(12): 6549–6553.

Paddock, J., Siegel, D. and Smith, J. 1988. "Option Valuation of Claims on Real Assets: the Case of Offshore Petroleum Leases," *Quarterly Journal of Economics* 103 (3): 479 – 508

Perron, P. 1989. “The Great Crash, the Oil Price Shock and the Unit Root Hypothesis,” *Econometrica* 57: 1361 - 1401

Pindyck, R. (1999) “The long-run evolution of energy prices.” The Energy Journal 20(2):1–27.

Pindyck, R. (2004) “Volatility and Commodity Price Dynamics.” Journal of Futures Markets 24(11): 1029–1047.

Postali, F.A.S. and P. Picchetti (2006) “Geometric Brownian Motion and Structural Breaks in Oil Prices: A Quantitative Analysis.” Energy Economics 28(4): 506–522.

Sadorsky, P. (1999) “Oil Price Shocks and Stock Market Activity.” Energy Economics 21(5): 449–469.

Schwartz, E.S. (1997) “The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging." Journal of Finance 52(3): 923–973

Schwartz, E.S. and J.E. Smith (2000) “Short-term variations and long-term dynamics in commodity prices.” Management Science 47(2): 893-911.

Schmidt. P. and P.C.B. Phillips (1992). “*LM* Test for a Unit Root in the Presence of Deterministic Trends.” Oxford Bulletin of Economics and Statistics 54: 257–287.

Schoenberg, R. 2006. “Generalized AutoRegressive Conditional Heteroskedasticity Models”, Internet site: http://faculty.washington.edu/rons/garch.html Downloaded 08/26/2006

Serletis, A. and R. Rangel-Ruiz (2004) “Testing for Common Features in North American Energy Markets.” Energy Economics 26: 401–414.

Sivapulle, P. and I.A. Moosa (1999). “The Relationship Between Spot and Futures Prices; Evidence from the Crude Oil Market.” Journal of Futures Markets 19(2): 175–193.

Wang, P (2003) Financial Econometrics. New York, New York: Routledge

Wirl, F. (2008) “Why do oil prices jump (or fall)?” Energy Policy 36: 1029–1043.

Zivot, E. and D.W.K. Andrews (1992) “Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis.” Journal of Business and Economics Statistics 10: 251–270.

Figure 1: Histogram of Returns; WTI Spot Prices

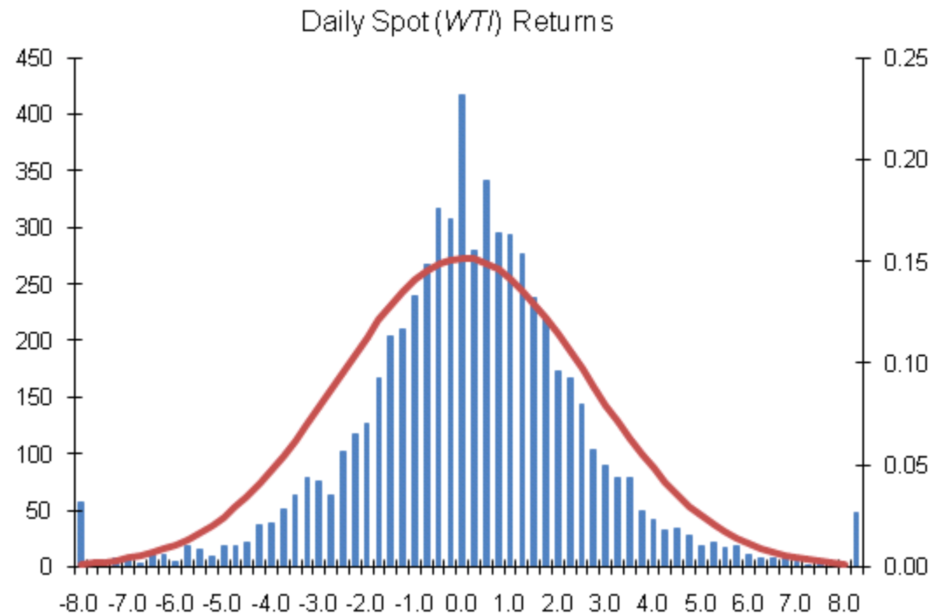

Figure 2: Implied Break Points for Crude Oil (*WTI*) Spot Prices

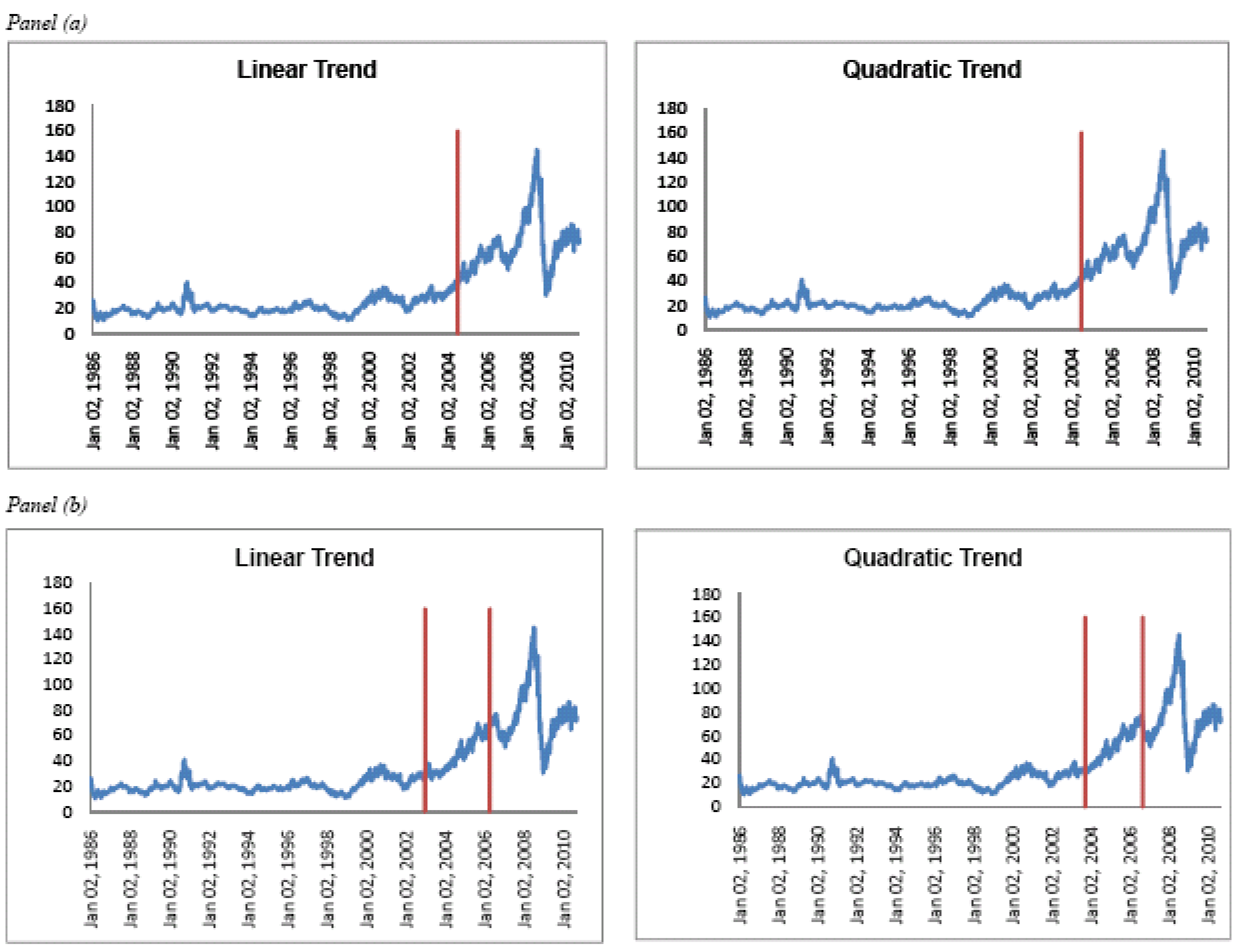

Table 1: Summary Statistics for Log-Price Relatives (Crude Oil Prices)

Data	Spot Price (WTI)			Spot Price (Brent)			Futures Price (1- month)		
	Daily	Weekly	Monthly	Daily	Weekly	Monthly	Daily	Weekly	Monthly
Median	0.0561	0.3114	1.2146	0.0262	0.3487	0.6391	0.0391	0.2082	0.9138
Maximum	19.15	25.12	39.22	18.13	23.36	45.89	16.41	25.53	37.72
Minimum	-40.64	-19.23	-39.48	-36.12	-23.16	-31.10	-40.05	-19.04	-39.64
Standard Deviation	2.63	4.54	8.89	2.39	4.37	9.13	2.45	4.18	8.50
Coefficient of Variation	15815%	5612%	2215%	10112%	3891%	1785%	18858%	6677%	3082%
Skewness	-0.770	-0.185	-0.424	-0.659	-0.255	-0.054	-0.844	-0.302	-0.398
Kurtosis	14.509	3.344	3.124	14.406	2.869	2.550	15.733	3.749	3.216
n	6223	1286	296	5917	1215	280	6873	1429	329
Jarque - Bera	34,958.1	13.7	9.1	32,501.7	14.0	2.5	47,247.4	55.1	9.3
probability	0.0000	0.0011	0.0108	0.0000	0.0009	0.2875	0.0000	0.0000	0.0094

Note: The data was obtained from the Energy Information association website of the Department of Energy and includes data available through the end of August 2010. The calculation of the log-price returns are given in equation (13)