Does oil price volatility scale with oil price?

James Hamski (james.hamski@spsmail.cuny.edu)

May 2017

1 Introduction

In a commodity trading market the price level is expected to be tied to commodity system dynamics, namely supply, demand, and delivery of the commodity being traded. Volatility, the variation in price over time, reflects uncertainty in the balance of these system factors. This research paper aims to answer a simply posed question: "does oil price volatility scale with price?"; i.e., can we expect to observe larger price swings when the price is near $100 per barrel vs $20 per barrel?

If oil price volatility reflects uncertainty about supply and demand dynamics, it isn't immediately clear whether we should expect volatility to depend on price level. Higher oil prices are associated with "tightness" in the supply market, meaning there is little excess capacity to increase production, therefore we may expect swings higher but some base price support that results in lower measured volatility. Likewise, low prices may suggest excess capacity that can buffer shocks to the oil delivery system, dampening volatility. Despite these "just so" arguments, a multitude of factors such as storage dynamics, supply chain disruptions, the ability of producers to increase production to bring more oil to market or shut in production capacity in response to prices ("rebalancing"), and market speculation complicate this picture and suggest it must be studied empirically.

If it is found that oil price volatility is dependent on price level, the relationship may follow a scaling formula. For instance, if we can expect volatility of $1/barrel when oil is at $20/barrel, can we expect volatility of $5/barrel at $100/barrel price levels via a simple linear scaling rule? Three methods are presented in this research paper to answer this question: (1) regression modeling of price and volatility, (2) viewing volatility within oil price regimes, and (3) using multivariate Generalized Autoregressive Conditional Heteroskedasticity (GARCH) modeling.

Note that in this research paper *oil price* is used to specifically mean spot-traded crude oil. This represents only one component of the oil markets, and 90%-95% of oil is sold under term contracts (Platts, 2010). Still, the spot price of oil is an important economic indicator used by the U.S. Federal Reserve as consideration for monetary policy. This research paper is concerned with understanding the energy system using pricing information. In this way, it differs from much of the published research in that it is not concerned with forecasting prices or volatility. Nor is it addressing exogeneous system elements such as equity markets or interest rates, though the literature shows that the crude oil market and larger economic indicators are intertwined (add cite). Instead, it contributes to our understanding of the system dynamics of an essential energy commodity.

2 Exploratory Data Analysis

2.1 Data Source

The data source is the West Texas Intermediate (WTI) nominal (i.e. not inflation adjusted) daily spot price record from the U.S. Energy Information Administration. Using nominal prices maintains the ability to assess volatility for a given time period without introducing an exogenous factor via inflation adjustment. The WTI series was filtered to the date range January 2, 1986 through December 30, 2016 (Table 1).

Table 1: Oil price series date and price ranges.

Date Range	Price Range
Min. :1986-01-03	Min. : 10.25
1st Qu.:1993-09-01	1st Qu.: 19.38
Median :2001-06-11	Median : 28.01
Mean :2001-06-19	Mean : 42.87
3rd Qu.:2009-03-31	3rd Qu.: 63.47
Max. :2016-12-30	Max. :145.31

2.2 Returns and Volatility

In this research paper, volatility is characterized two ways: (1) 5-day historic volatility and (2) 30-day historic volatility. Historic volatility is defined as:

$$V_{H,t} = \sqrt{N} sd(P)$$

In addition, the relationship between the returns themselves and price level is investigated. Daily returns were calculated as:

$$R_t = \ln(P_t - P_{t-1})$$

Note that the standard return formula takes the log of the daily price differences, meaning the day to day price differences are scaled using a natural logarithm. In some sections, the absolute value of "unscaled returns" (i.e. $|P_t - P_{t-1}|$) are used in order to investigate the scaling of the return series.

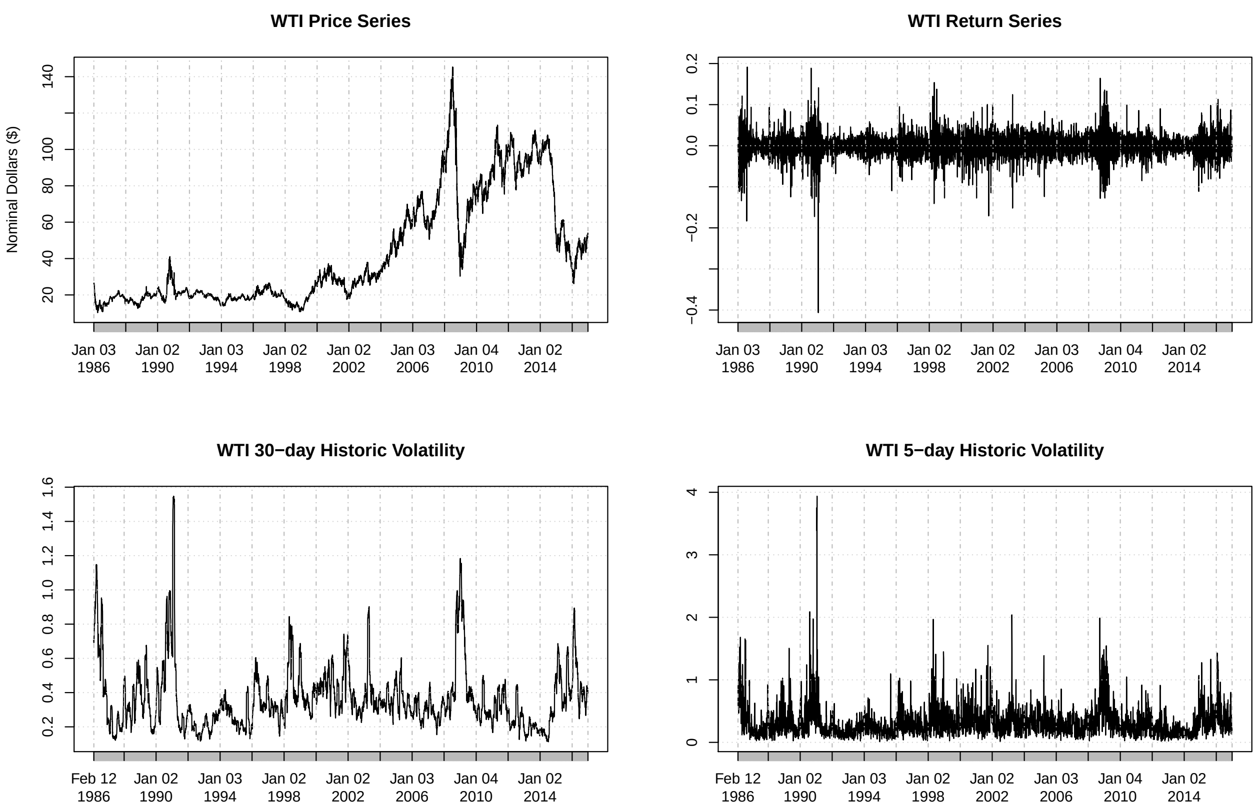

Figure 1: Price level, return, and volatility series plots of WTI spot oil prices.

A seen in Figure 1, most of the series from 1986 through 2004 contains prices between $10/barrel and $40/barrel. This results in a price series with a skewness value 0.98 and a long right tail (Figure 2), with most prices being around the series median of $28 per barrel. The return series exhibits fatter tails and a narrower center than a normal distribution of the same descriptive parameters. The kurtosis value of the return series is 13.52. Returns with excess kurtosis compared to normally distributed returns is a common characteristic in financial time series. This indicates that most returns are very close to the mean return, but some returns are dispersed very far from the mean.

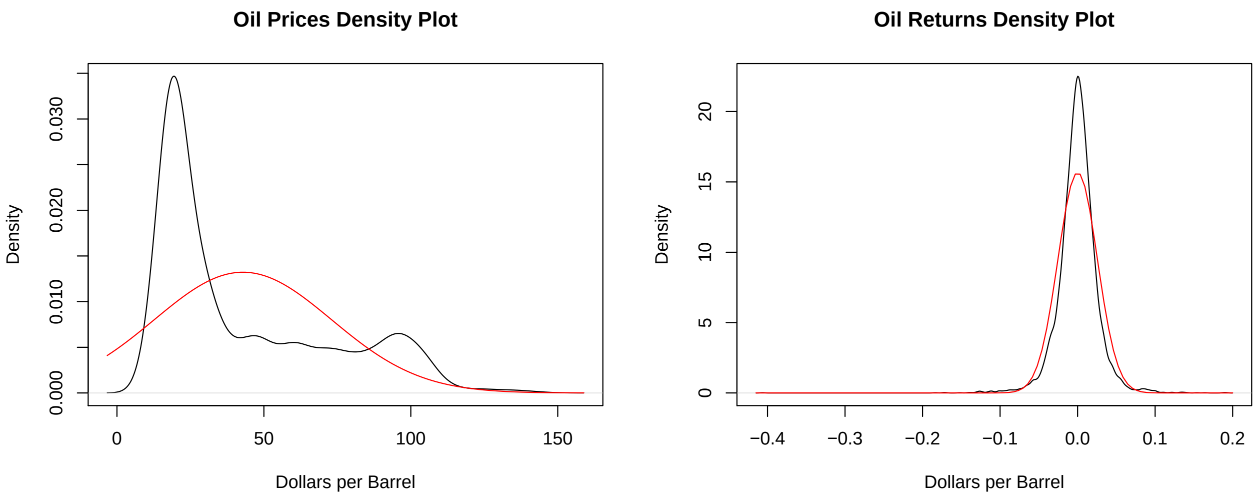

Figure 2: Price series and return density plots including normal distributions for comparison.

2.3 Autocorrelations

The autocorrelation plots for the oil price and return series indicate a high degree of correlation between adjacent observations. While we are not utilizing an autoregressive model here, partial autocorrelations are also shown in Figure 3 to show that partial autocorrelations do not exist in the price series after a shift of 1 day, compared to a longer partial autocorrelation decay in the return series.

3 Price-Volatility Regression Analysis

Relating price level and the measures of volatility at each time in the series is a simple exploration of the research problem. The covariance and correlation measures of vectors representing price versus returns, 30-day, and 5-day historic volatility indicate a weak, negative relationship (Table 2). However, the unscaled return series indicates a stronger positive covariance and correlation with the price series.

Table 2: Covariance and correlation of price level and returns, 30-day, and 5-day historic volatility.

	Covariance	Correlation
Return Series	0.0102677	0.0134110
Unscaled Return Series	12.0871446	0.4393373
30-day Historic Volatility	-1.0525476	-0.1860878
5-day Historic Volatility	-0.8388984	-0.1123822

In quantitative finance, a process where volatility scales with price is a lognormal process. When volatility is

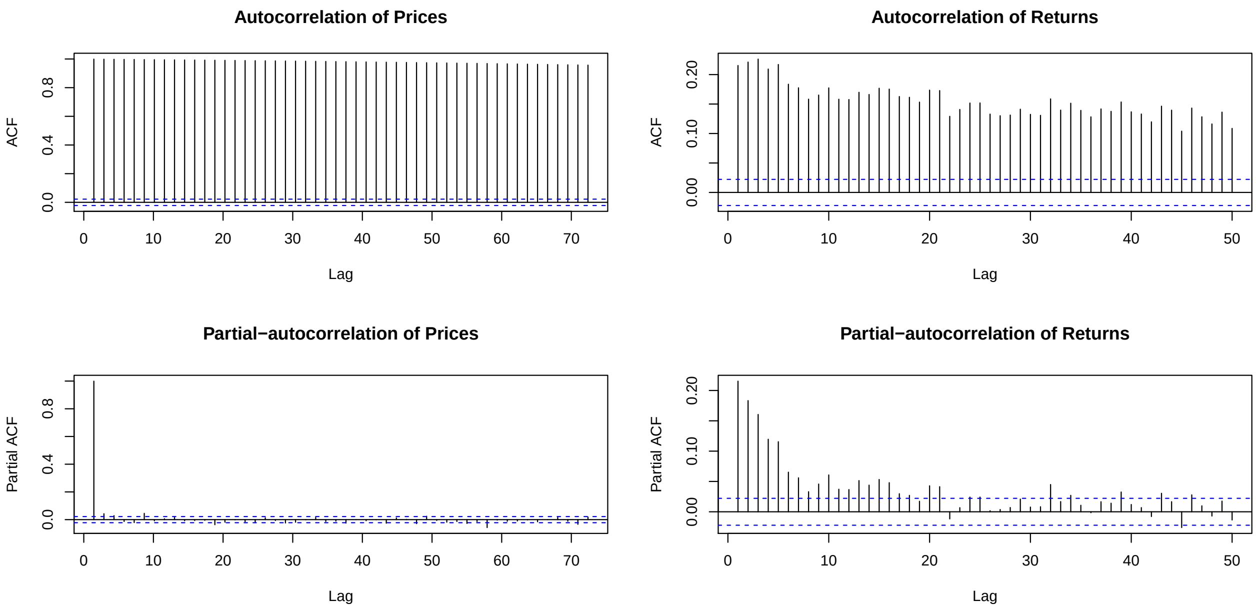

Figure 3: Autocorrelation plots for prices and returns.

independent of price, the process is normal (Ho and Lee, 2003). Regression of the price level and squared returns is identified as the method of distinguishing lognormal from normal processes in the literature. Figure 4 shows scatter plots of the two return series and two historic volatility measures fitted with a linear regression model.

The case of 30-day historic volatility indicates a negative relationship between price level and volatility. However, this result appears to be due to a cluster high volatility around $20 per barrel, creating a leverage point. In addition, it is obvious from even the scatter plot that residuals will not indicate stochastic error.

The unscaled returns appear to indicate an increasing relationship between day-over-day price change and base price. The resulting linear model is: $R_t = 0.013P_t + 0.14$ has an R^2 value of 0.19. Analyzing the residuals by calculating Cook's Distance indicates that massive price movements in the fall of 2008 resulting from the financial crises are acting as leverage points and the relationship between unscaled returns and price may be weaker than indicated by the linear model.

As seen in Figure 1, nominal oil prices have spent time as high as $145 per barrel. However, the majority of the time series is far lower, with a median price of $28 per barrel. This means that the data set is unbalanced and higher price levels represent a smaller portion of the data set. In addition, oil price (and financial time series in general) exhibits volatility clustering. Therefore, it is anticipated that this simple regression model based on price and volatility is not the best possible solution to the question of characterizing the dependency of volatility on price. A second time series was created by limiting WTI prices to the January 3, 1998 through December 31, 2016. This results in a distribution that, while significantly different than the normal distribution, is more evenly distributed across the price range from $10.82 per barrel to $145.30 per barrel (Figure 5).

Given the ever evolving economic, political, and technological landscape affect commodity prices, there is some trade off between looking at a large historic price record, which covers a larger data set of possible system states (and higher statistical power), and limiting to a more recent price record, which more fully represents the system in its current state. However, the linear models relating price level to returns and volatility resulting from the recent time series does not differ much from using the full data set. In general there is a weak, negative correlation between volatility and price, and a weak positive correlation between returns and price.

This simple method of exploring the relationship between price level and volatility does not provide a

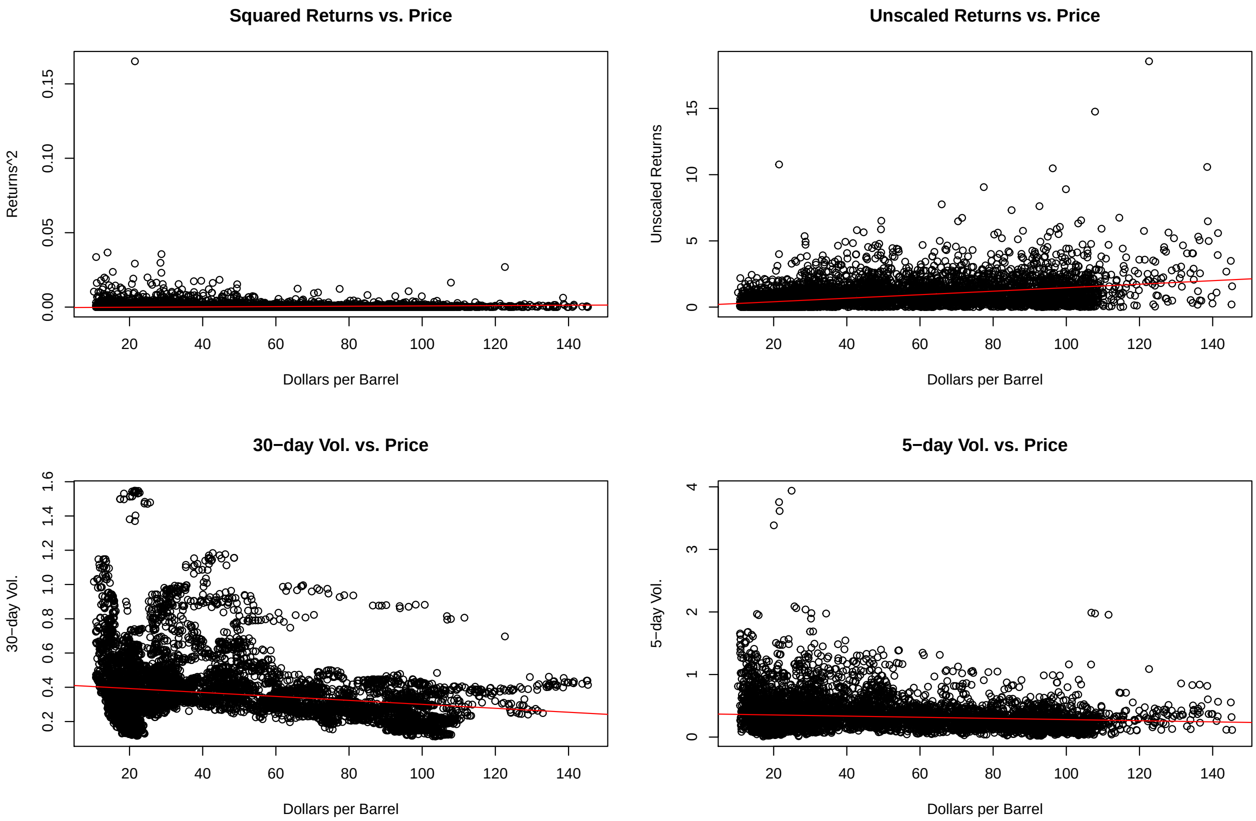

Figure 4: Scatterplots with a linear model relating spot oil price with returns, 30-day, and 5-day historic volatility.

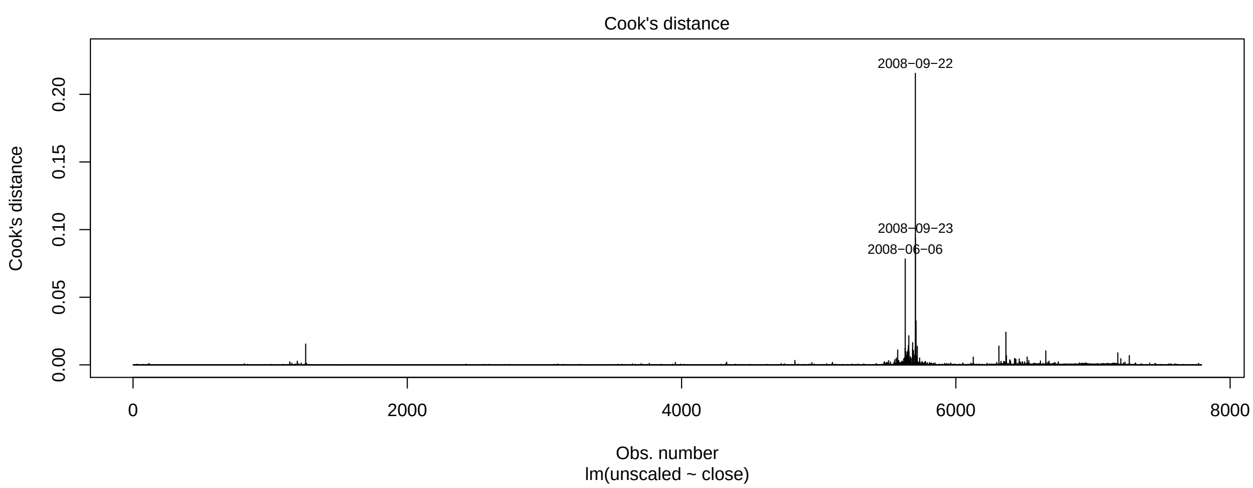

Figure 5: Cook's Distance plot indicating measurements acting as leverage points.

satisfactory explanation of whether we can expect more volatility, indicating more commodity system uncertainty, at high price levels. Volatility in financial time series tend to cluster. Typically, some event (called a "shock") occurs which results in an extremely high price movement. These are the large noticeable movements in the return series (Figure 1). Subsequent returns are also of higher magnitude than the typical return size, but taper off over time, eventually returning to a value close to the average return for the price series. This is referred to as volatility clustering, and indicates heterosketasticity in the variance component of a time series. This violates the assumption in the most frequently used time series model, the autoregressive integrated moving average (ARIMA) model.

These results using simple linear regression models to study how daily returns and historic volatility relate to price level may not adequately deal with this time-varying variance structure. If, for a given price level, we have a price shock, we expect the volatility to return back to its normal level. So at that price level we may have relatively few measurements indicating high volatility, resulting in regression models heavily weighted towards the baseline level of volatility. This "one to one" view of the price level - volatility relationship leaves the question open as to whether we can anticipate more volatility or larger price shocks at higher price levels.

4 Comparing Volatility across Price Regimes

Regime switching has been investigated in oil prices, however the focus has been on breaking volatility into regimes (Vo, 2009). In order to capture the volatility dynamics of the WTI price series, including volatility clustering and changes in base price, change point detection was used to break the series into price regimes. Change point detection aims to detect the point or points where the statistical properties of a sequence of observations change (Killick et al., 2016). Change point detection was used to estimate changes in the oil price mean throughout the period of record using the Pruned Exact Linear Time (PELT) algorithm (Killick et al. 2012). The time series between these change points represent "price regimes" (i.e. time series between change points) which have generally similar mean oil price compared to the entire record.

Change point detection proceeds by minimizing a cost function over possible locations and number of change points. The cost function:

$$\sum_{i=1}^{m+1} [C(y_{(\tau_{i-1}+1):\tau_i})] + \beta f(m)$$

Where m is the number of change points and their positions in the series at $\tau_{1:m} = (\tau_1, ..., \tau_m)$, within a series $y_{1:n} = (y_1, ..., y_n)$. C is a cost function for a segment and $\beta f(m)$ is a penalty to increase the threshold to identify a changepoint, thereby preventing overfitting (Killick ET AL., 2012). Minimizing this calculation can be computationally demanding, hence the development of algorithms such as PELT, however for the oil price series this is not a practical constraint.

The change point analysis was performed with a minimum regime length of 250 trading days, approximately one calendar year. The cost value was set to 10,000. This parameterization resulted in 15 price regimes (i.e. $m + 1$ change points). Then, within each regime the standard deviation (to indicate volatility) and median price were calculated.

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  wti.regimes.ret$returns and wti.regimes.ret$regime
## Fligner-Killeen:med chi-squared = 587.38, df = 14, p-value <
## 2.2e-16
```

The Fligner-Killeen test of homogeneity of variances determines if it is appropriate to reject the null hypothesis that the variances between each regime are the same. This test was chosen due to it's suitability for non-normal data. The chi-squared value of the test was 587.38 and the p-value was less than 2.2e-16, indicating it is appropriate to conclude the variances across the regimes are different.

The data set describing the regimes were then modeled using linear regression (Figure 6). The resulting model indicates a slope of 0.08, meaning that for every dollar increase in the median price of a price regime, we expect an increase in the standard deviation of that regime of 0.08. This model has an adjusted r-squared value of 0.45.

Table 3: Linear regression resulting from price as the predictor variable and regime variance as the response variable.

term	estimate	std.error	statistic	p.value
(Intercept)	1.9666240	1.5125362	1.300216	0.2161068
med	0.0828545	0.0252947	3.275565	0.0060244

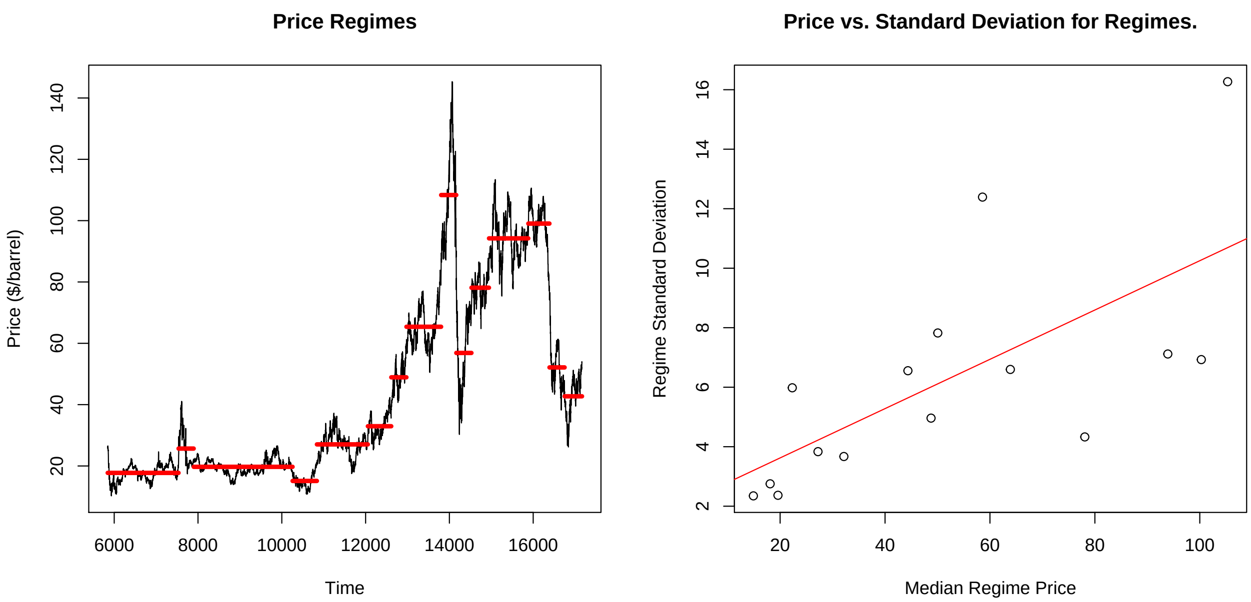

Figure 6: Price regimes and linear model relating median oil price for the regime to standard deviation within that price regime.

Analyzing the residuals of this model indicate some problems with the fit (Figure 7). Namely, residuals are not evenly distributed and the model is subject to suspected leverage point observations. This is typical in linear models with small sample sizes.

This section's analysis using changepoint analysis to break the oil price series into price regimes as defined by changes in the series mean is highly dependent upon model parameters. The PELT change point detection method identifies different regimes depending on the minimum segment length and penalty parameters, and based on the section of the time series used. An interactive Shiny App was created and may be access in order to try different parameterizations of the change point model and view the effects on the relationship between medial regime price and standard deviation within that regime: (https://jhamski.shinyapps.io/oil_price_regimes_and_volatility/).

5 Multivariate GARCH Model

Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) models attempt to model the variance component of a time series. They can account for asymmetric, clustered time series variance and has successfully been applied to model oil prices (Salisu and Fasanya, 2012; Narayan and Narayan, 2007). Most

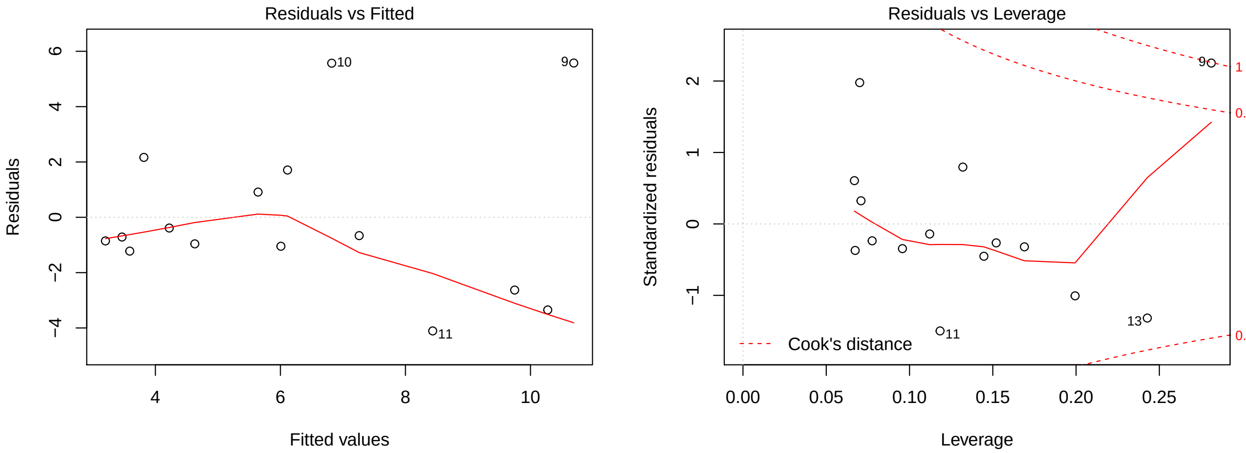

Figure 7: Residual Plots of the Linear Model fitted to Regime Price vs. Standard Deviation

of the literature exploring GARCH models is interested in predicting volatility and identifying volatility relationships across different time series, for instance volatility in oil price also occurring when there is volatility in the stock market.

Dynamic Conditional Correlation (DCC) models are based on univariate GARCH models. They parameterize the conditional correlations directly and are estimated in two steps – the first is a series of univariate GARCH estimates and the second the resulting correlation estimate. The correlation is "conditional" because it is based on the knowledge of the series at one time step before the time the correlation is calculated. Essentially, DCC is looking for "contemporaneous correlation" in the variance of two series.

This study is attempting a multivariate GARCH model using a return series, which is derived from the price series, and the price series itself. While using a derivative of a series is common in time series analysis (i.e. autoregressive models), this doe snot appear to be a common

A DCC model was constructed from an ARMA(1,1) and Glosten-Jagannathan-Runkle GARCH model. This means that we specified a base model for the conditional mean using an autoregressive–moving-average (ARMA) model with an autoregressive term of one and a moving average term of on. Note that the same terms used in Narayan and Narayan, 2007, an ARMA(12,12) model, were used. However all models built failed to converge. This may indicate that there is a problem with linear combinations resulting from building a model with price level and volatility.

The resulting conditional covariance chart (Figure 8) shows that during times of high volatility, the price level and returns have covariance values as high as 0.4 and 0.55. However, in general the value is less than 0.1. This result does not lead to a clear conclusion and the methodology laid out in this section requires more work.

6 Conclusion

This analysis suggests that the question "does oil price volatility scale with price?" requires further clarification to address.

Expected Returns at a Given Price Level

The regression of unscaled returns to price level indicates that we may expect returns to increase linearly with increasing price. The resulting linear model $R_t = 0.013P_t + 0.14$ indicates that for every one dollar increase in price, we may anticipate a 1.3 cent increase in expected returns. However, with an R^2 value of 0.19 the model does not explain most of the variation in the relationship between unscaled returns and base price.

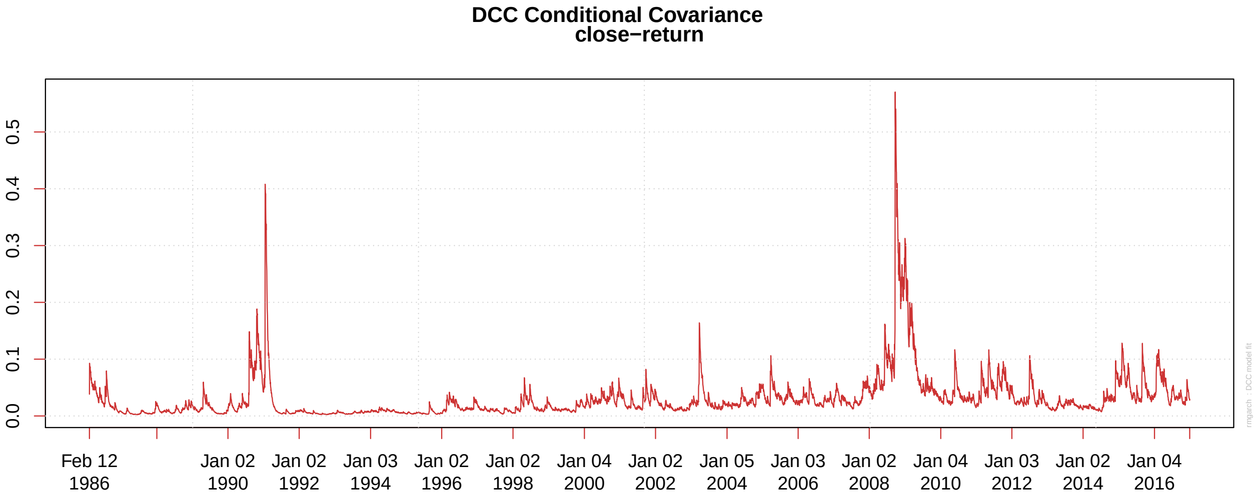

Figure 8: Dynamic Conditional Correlation between Price and Returns

Volatility Within Price Regimes
Detecting change points, breaking the price series into regimes, and comparing the standard deviation within those regimes provides the clearest evidence that we may expect higher volatility at higher prices. This method has the benefit of capturing volatility clustering and is less suseptable to imbalance due to the high kurtosis of the return series. Based on this method, we may expect the standard deviation of our returns (i.e. volatility within a price regime) to increase by 0.8 cents for every one dollar increase in that regime's median price.

Multivariate GARCH Models using Price Level, Returns, and Volatility

The method of using DCC models to investigate the relationship between oil price and volatility requires further exploration and theory development. It is not clear that the method of modeling volatility using a GARCH model will allow us to determine if price level is a an explanatory factor of volatility.

7 Acknowledgements

I would like to thank John Kemp, Thompson Reuters energy journalist, for posing the question investigated in this research paper. In addition, I thank Tancred Lidderdale and Mason Hamilton from the U.S. Energy Information Administration for additional information pertaining to the question.

8 Bibliography

Engle, R., 2002. Dynamic conditional correlation - a simple class of multivariate GARCH models. Journal of Business and Economic Statistics.

Ghalanos, Alexios, 2015. rmgarch: Multivariate GARCH models. R package version 1.3-0.

Killick, R., Fearnhead, P. and Eckley, I.A., 2012. Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association. 107(500), pp. 1590-1598

Killick R, Haynes K and Eckley IA, 2016. changepoint: An R package for changepoint analysis. R package version 2.2.2, <URL: https://CRAN.R-project.org/package=changepoint>.

Narayan, P. K., & Narayan, S., 2007. Modelling oil price volatility. Energy Policy, 35(12), 6549–6553. doi:10.1016/j.enpol.2007.07.020

Platts, 2010. The Structure of Global Oil Markets. Backgrounder marketing publication.

Thomas S. Y. Ho and Sang Bin Lee, 2004. The Oxford Guide to Financial Modeling: Applications for Capital Markets, Corporate Finance, Risk Management and Financial Institutions. ISBN: 9780195169621

Salisu A.A., and Fasanya, Y.O., 2012. Comparative Performance of Volatility Models for Oil Price

Vo, M.T., 2009. Regime-Switching Stochastic Volatility: Evidence from the Crude Oil Market. Energy Economics, 31, 779-788.