Robust Estimation of Conditional Risk Measures for Crude Oil and Natural Gas Futures Prices in the Presence of Outliers[☆]

J. Byers[a,*], I. Popova[b], B. J. Simkins[a]

[a] *Department of Finance, Spears School of Business, Oklahoma State University, Stillwater, OK 74078*

[b] *Department of Finance & Economics, McCoy College of Business, Texas State University, San Marcos, TX 78666*

Abstract

In this study, we aim to build better risk models for energy commodities by employing statistical procedures to identify outliers in the prices for all crude oil and natural gas futures contracts traded on the CME over the period of December 2003 through March 2017. Empirical results for crude oil and natural gas futures contracts show including appropriate parameters for outlier effects is required when performing parametric estimation of risk parameters because outliers can have a large impact on the estimation of value at risk. We illustrate for actual crude oil and natural gas contract how the use of value at risk metrics based on raw data can lead to higher than expected actual losses. Our research demonstrates that it is crucial to include intervention parameters to address outlier impacts in order to obtain robust risk

[☆] We thank the participants at the Oxford Commodity and Energy Markets Conference 2017, the participants at the 35^{th} USAEE/IAEE North American Conference 2017, seminar participants at Ulsan National Institute of Science and Technology, Korea, Oklahoma State University, the participants at EFA 2018, FMA 2018.

[*] Corresponding author

Email addresses: `joe.w.byers@okstate.edu` (J. Byers), `ip12@txstate.edu` (I. Popova), `betty.simkins@okstate.edu` (B. J. Simkins)

metrics.

Keywords: Value at Risk, Conditional Value at Risk, Risk Capital

1. Introduction

It is well known that many commodity prices exhibit anomalous changes due to geo-political, weather, and industry specific events. For example, prices of crude oil changed drastically when on September 16, 2008 the OPEC cartel, which accounts for almost 50% of the world's oil, lowered its forecast for oil demand that year due to slowing economic growth. Similar volatile price behavior was observed on August 3, 2011 when oil settled low as investors continued to worry about weak consumer spending and sluggish economic growth; and on July 15, 2008 when President Bush lifted nearly two decades of executive orders banning drilling for crude oil and natural gas off the country shoreline.

The computation of robust risk metrics for any commodity firm with trading operations can be extensive. This is because the number of risk factors can be large, measuring in the thousands. For example, on any given day, natural gas futures have up to 120 risk factors[1] based on each open contract. Other risk factors include locations in the North American demand/supply centers like New York City Gate and Dominion South. An additional important feature is the broad number of transactions from exchange traded products to over-the-counter (OTC) and structured products that are part of a portfolio. Adding that commodity operations span multiple commodities in cross commodity transactions[2], the computational complexity of risk metrics increases substantially.

[1] At any point of time there are contracts with 10 years of maturity, 12 months per year.

[2] The exchange and OTC Crack Spread Contracts include two or more risk factors and lines of business.

Commodity firms like British Petroleum and financial firms like Goldman Sachs, hold commodity portfolios that contain open transactions numbering in the millions. The quantity of risk factors associated with the company's portfolios measure in the hundreds and even thousands. Each of these risk factors has a risk model that price outliers can influence. Additionally, a portfolio of commodity transactions also has covariation or correlation to address across time and commodities. Outliers could cause these diversification effects of commodity hedges to be incorrectly measured and hence costly for this firm.

Our motivation for this paper is the lack of research in estimating robust risk metrics in the presence of outliers for commodity prices. Identifying outliers is not equivalent to data winsorizing. Vast amount of research in finance winsorizes (and as a result removes and replaces) data points prior to applying any statistical estimation methodology. Outlier is a sudden or extraordinary discrete change termed an anomaly in the price sequence. These price events if not addressed, could lead to erroneous conclusions or inferences according to Tsay (1988).

Our empirical results using 14 years of daily settlements for commodity instruments from the CME group, show that a robust estimation (after properly modeling outliers) leads to an increase or decrease of VaR metrics. The increase in VaR occurred in 5% of crude oil contracts and 5.5% for natural gas contracts. These cases could potentially cause serious problems for a commodity trading firm. The reason is that the expected loss if VaR is exceeded, could be much larger than anticipated. These larger losses would require immediate risk capital to be deployed. Margin calls on exchange traded instruments, or posting of additional capital on over-the-counter transactions may be required. Additionally the firm may be found in violation of credit arrangements that could trigger a technical default. The larger than anticipated losses may also have an impact on the firm's credit receiving counterparties.

The risk parameters used to calculate credit metrics for counter parties could mean the firm has extended more credit to a counterparty than should have been allowed.

Outlier detection algorithms fall within two categories: Bayesian approaches (see McCulloch and Tsay (1994)), and non-Bayesian approaches (see Chen and Liu (1993)) as one of the most widely used procedures. The detection methodologies and algorithms that we use in this study belong to the non-Bayesian methodologies.

This study proceeds as follows. Section 2 discusses the background of this line of research and includes a literature review of the leading papers in the area. Section 3 describes the methodology for detecting outliers. Section 4 presents an outlier analysis for crude oil and natural gas futures contracts, discusses the impact that outliers have on computing different risk metrics, and illustrates the improvements after applying the outlier adjustments. Section 5 concludes.

2. Background and Literature Review

The study of outliers in time series starting with Fox (1972) has been very active. Detecting outliers is known by many terms including anomaly detection, event detection, novelty detection, deviant discovery, change point detection, fault detection, intrusion detection, and misuse detection according to Gupta, Gao, Aggarwal and Han (2014). Outlier research has been conducted on many types of data[3].

[3] Gupta, Gao, Aggarwal and Han (2014) discuss research in outliers that focused on temporal forms of data such as credit, financial, medical, judicial, astronomy, web usage, sensor, real and virtual traffic, and commercial transactions. Chang, Tiao and Chen (1988) review chemical process articles, and Burman and Otto (1998) study business division data series of retail and wholesale sales. Marczak and Proietti (2014) investigate industrial production for France, Germany, Spain, the United Kingdom, and the U.S. for 1991 to 2014 and find that outliers coincided with the Economic Crisis of 2009. Chen and Liu (1993) argue the outlier adjustment is an indispensable part of intervention analysis and applying their method allows for a more complete set of, though fewer, outliers with more robust results.

The major advancement in outlier research is the ability to quantitatively classify types of outliers and their impact on the data generating process (DGP). In his seminal work, Fox (1972) describes two types of outliers. Type I outlier is a gross error from a single observation and Type II outlier is an extreme innovation that affects observations following the event. Tsay (1988) extends the set of outliers to include level shift or change (LS) and temporary change (TC). A level shift is a step change in the time series. A temporary change is a temporary or transient level change in the time series. Consequently, the Type I and Type II outliers were renamed to additive outliers (AO) and innovative outliers (IO), respectively. Gupta, Gao, Aggarwal and Han (2014) note that AO and LS are single point events while IO and TC continue to impact the DGP following the event. Tsay (1988) introduces a variance change outlier (VC) and Watson, Tight, Clark and Redfern (1991) revisit it. Each of these outlier types affect the DGP.

Figure 2 shows how these innovations impact a time series. AO is on the top panel illustrated by two up and down single point events. TC is shown on the second panel where an extreme event continues to impact the DGP for a period of time. LS is in the third panel and the DGP shifts up after the event and stays there. IO is on the bottom panel where an extreme event causes temporary large deviations in the DGP until it slowly disappears.

Insert Figure 2 here

Tsay (1988), Chen and Liu (1993), and Fox (1972) argue that IO do not require adjustment since they impact the system after the outlier event occurs. Chen and Liu (1993) add that IO are not independent of the model and will decay with time for a stationary process. However, for a non-stationary process, after the event there may be an initial effect followed by a level shift (LS).

The standard approach for dealing with outliers is to locate and identify the types of outliers, then use intervention parameters to model the outlier effects. A simple way to think about an intervention model is to consider two different intervention variables - step and pulse response. Outliers can describe the dynamic pattern of untypical effects and can be captured by means of intervention variables[4].

Detailed discussion of the ARMA modeling of outliers can be found in Chen and Liu (1993), Chang, Tiao and Chen (1988), Muirhead (1986), Fox (1972), Sanchez and Pena (1997), and Tsay (1988). Fox (1972) and Chang, Tiao and Chen (1988) derived likelihood ratio statistics for determining AO or IO outliers. Tsay (1988) extends the outlier test statistics to include LS and TC outliers. Chen and Liu (1993) extend Chang, Tiao and Chen (1988) to include likelihood ratio statistic for LS and TC outliers.

In the commodity markets, Ju, Su, Zhou, Wu and Liu (2016) study annual crude oil prices relationship with CPI, GDP, and unemployment across 18 production and consumption economies. The study utilized two distance based and one artificial intelligence based outlier detection algorithms. Their study focuses on the interaction of independent variables and annual oil prices. It highlights a higher number of outliers found in oil consuming economies since 2008. Our study, using daily data, sheds additional light on the number and impact of outliers on commodity prices that could impact the macro economic data studied. Bagnai and Ospina (2018) study outliers as regime shifts in a multivariate cointegration setting. Their study uses monthly gasoline and crude oil prices series from the Energy Information Administration and

[4] In the presence of multiple outliers, a problem known as masked or shadow outliers may occur (see Chen and Liu (1993), Fox (1972), López-de Lacalle (2016), and Chang, Tiao and Chen (1988)). This happens when multiple outliers are detected and corrected independently. To account for this, shadow outliers require a final review of their statistical significance in an ARIMA model.

International Monetary Fund. It finds a single regime shift in 2008. The regime (outlier) detection techniques and the t − tests for symmetric elasticities are similar to Tsay (1988) and Chen and Liu (1993) LS algorithms and statistical tests. The results of their study expands the knowledge on regimes and interactions of crude oil and gasoline commodities. Our study, whilst not a multivariate analysis, utilizes data and techniques that can be applied to business risk decisions.

3. Methodology

In this section, we describe the methodology for detecting outliers. A three stage algorithm is presented together with the process for addressing outliers.

Following Chen and Liu (1993), let Y_t be a time series following an ARMA process without drift or trend:

$$Y_t = \frac{\theta(B)}{\varphi(B) \cdot \alpha(B)} a_t, t = 1, \cdots, n \tag{1}$$

where n is the number of observations, B is the back shift operator, $\theta(B)$ is a moving average component, $\varphi(B)$ is an auto regressive component, and $\alpha(B)$ is a difference component. All of these are polynomials of B. All roots of $\theta(B)$ and $\varphi(B)$ are outside the unit circle and all roots of $\alpha(B)$ are on the unit circle.

The following model describes a time series that is influenced by a nonrepetitive event:

$$Y^* = Y_t + \omega \cdot \frac{A(B)}{G(B) \cdot H(B)} \cdot I_t(t_1), \tag{2}$$

where Y_t follows a general ARMA process defined in Equation (1). Additionally, $I_t(t_1)$ is an indicator function equal to 1 if an outlier occurs and zero otherwise. The magnitude of the outlier effect is denoted by ω and $A(B)/G(B)\cdot H(B)$ represents the dynamic impact of outliers on the process.

By imposing a special structure on $A(B)/G(B)\cdot H(B)$, we can classify the outlier as IO if:

$$\frac{A(B)}{G(B)\cdot H(B)} = \frac{\theta(B)}{\alpha(B)\cdot\varphi(B)} \tag{3}$$

This is an outlier in the innovation series a_t that occurs at time $t = t_1$ and has a dynamic effect on Y_t^*. The effect of this outlier on $Y_{t_1+k}^*$ for $k \geq 0$ is equal to $\omega\Psi_k$ where ω is the initial effect and Ψ_k is the k^{th} coefficient of the polynomial $\Psi(B) = \dfrac{\theta(B)}{\alpha(B)\cdot\varphi(B)} = \Psi_0 + \Psi_1 B + \Psi_2 B^2 + \cdots,$. For stationary series the IO will produce a temporary effect since Ψ_j's will decay exponentially to zero.

We classify the outlier as AO if:

$$\frac{A(B)}{G(B)\cdot H(B)} = 1 \tag{4}$$

The outlier only affects $Y_{t_1}{}^*$.

We classify the outlier as TC if:

$$\frac{A(B)}{G(B)\cdot H(B)} = \frac{1}{(1-\delta B)} \tag{5}$$

This is a disturbance that affects $Y_t{}^*$, $\forall t \geq t_1$ but decays exponentially with rate $\delta, 0 < \delta < 1$ and initial impact ω.

The outlier is classified as LS if:

$$\frac{A(B)}{G(B)\cdot H(B)} = \frac{1}{(1-B)} \tag{6}$$

At time t_1 there is a permanent change of size ω.

Fox (1972), Burman and Otto (1998), Tsay (1988), Muirhead (1986), Watson (1991), Sanchez and Pena (1997), and Chang, Tiao and Chen (1988) develop algorithms for detecting and correcting outliers. Chen and Liu (1993) extend this early work for an algorithm to jointly detect outliers and estimate model parameters under a parametric ARIMA models. López-de Lacalle (2016) implements Chen and Liu (1993)'s algorithm and incorporate aspects of time series with ARIMA Noise, Missing Values, and Outliers (TRAMO) (Maravall (n.d.)) that automatically identifies, selects and corrects for outliers[5]. The benefit of this algorithm is that it satisfies Tsay (1988)'s comment that a "simple useful method to detect and handle outliers" is needed. This parametric modeling approach and the algorithms for automatic detection and adjustment would be classified as unsupervised (see Gupta, Gao, Aggarwal and Han (2014)).

The general algorithm for detecting and correcting outliers used in this research is:

Stage 1. Outlier detection is performed estimating ARIMA models and checking for significant outliers at different times based on t-statistics from the parametric estimation.

Stage 2. Filter outliers by joint estimation of ARIMA models with results from Stage 1. Outliers found to be insignificant are dropped from the initial set based on t-statistics by parameter estimation.

Stage 3. Iterate over Stages 1 and 2 to determine the adjusted series and the final outlier effects.

This process will select parametric ARIMA models with intervention parame-

[5]The outliers detection algorithm runs in stages. The first stage is to determine the initial set of outliers. Here we used the R *tsoutliers* package with the *auto.arima* method to locate potential outliers based on t-test of each potential outlier.

ters for the outlier effects based on the minimization of Akaike (AIC) and Bayesian (BIC) information criterion statistics (López-de Lacalle (2016)). The algorithm will return the final outlier set, the regression coefficients, adjusted data series, regression residuals, and intervention parameters. Sanchez and Pena (1997) highlight how this iterative process improves on the determination and overcomes limitations previously mentioned. Our analysis utilizes the R Analytical software and statistical packages for outliers and forecasts developed by López-de Lacalle (2016) and Hyndman (2017) for estimating the initial and final outliers in each time series. Additionally, Fox (1972) specifies that seasonal components must be filtered and Maruth and Ryan (2007) show that seasonality must be handled for geometric Brownian motion assumptions. This research utilizes software developed by Hyndman (2017) and Hyndman and Khandakar (2008) that filters seasonal ARIMA components.

If outliers were found, the completion of the three stages on log returns will result in parametric specification of the time series, composed of two components as follows: an ARIMA model specification of the log returns and functional specifications if the intervention parameters for outliers identified that are based on Equations (4) to (6) with a decay rate of 0.7. Only the ARIMA specification is returned if no outliers are found, and this is the best fit model of the time series.

In order to illustrate the effects of different types of outliers, consider a model of the difference in commodity price log returns generated by an ARIMA(1,1,1) process with drift and an AO outlier at time step $t = k$. This model will have the following parametric specification:

$$\nabla Y_t = a_0 + \phi_0 \nabla Y_{t-1} + \theta_0 e_{t-1} + w_t I_t + e_t,$$
$$I_t = \begin{cases} 1 & \text{if t = k} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where ∇Y_t is the difference in commodity price log returns. A model of the difference in commodity price log returns generated by an ARIMA(2,1,1) process with drift that have an LS outlier at time step $t = k$ would be:

$$\nabla Y_t = a_0 + \phi_0 \nabla Y_{t-1} + \phi_1 \nabla Y_{t-2} + \theta_0 e_{t-1} + w_t I_t + e_t,$$
$$I_t = \begin{cases} 1 & \text{if t > k} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

As can be seen from the above two examples, outlier intervention analysis does not remove the outliers. It separates them from the ARIMA model and for the case of AO a one time shock at time k is included. For the LS, the price shock at time k has an impact on the DGP for $t > k$.

4. Analysis

4.1. Outlier Simulation Example

Here we use simulated data to illustrate the steps of the outlier intervention analysis. In the next section we show results with actual CL and NG daily futures prices.

We perform a simulation analysis to demonstrate how the algorithm described in Section 3 would detect different types of outliers. We simulate a single random walk without a drift with annualized volatility of 25% for 100 days. We then introduce outliers of each type in the simulation, which are described in Table 1. The AO jumps up 0.8 (80%) then back down by the same amount on days 33 and 67 and Panel 1 of Figure 1 displays the pattern. TC are demonstrated with a jump of 0.1 (10%) that decays 0.7 (70%) each day over days 45 to 65 and this pattern is shown in Panel 3 of Figure 1. LS are simulated with a jump of 0.1 (10%) on day 45 and are shown in Panel 2 Figure 1.

Insert Table 1 and Figure 1 here

The outlier detection methodology was applied to the five simulated time series. The upper section of Table 2 shows the summary statistics or the raw financial data series, and the lower section is summary statistics that include intervention parameters for outliers. The mean, standard deviation, and annualized standard deviation are highlighted before and after the outlier detection algorithm was applied. The annualized results show a significant difference from the 25% volatility used to simulate the data, and the means exhibit non-zero values in three cases. The mean of the base data is close to zero, as is the AO series.[6] The arithmetic means for the LS, TC, and IO simulated data are different from our base at -2.28%, 5.65%, and 0.06%, respectively. The annualized standard deviations are 25% for the random walk base case, but 180%, 154%, 84%, and 202% for AO, TC, LS and IO, respectively.

Insert Table 2 here

The skewness and excess kurtosis of the simulated outlier data series indicate that the data may be non-normal. The Jarque-Bera and Shapiro-Wilk tests for normality

[6] The value of -0.04% for each data series is a simulation error in a small sample. AO is not different from our base case by construction because two AO outliers occur at different times and are of equal, but opposite, magnitude thereby, offsetting each other in the mean.

were performed and the results are shown in Table 3. All raw data series reject normality with the exception of the base case raw data, which is by design.

Insert Table 3 here

The time series methodology, which includes intervention parameters for outliers, was applied to the simulated data series. All data series final model time series specifications were ARIMA(0,1,0) corresponding to our base case initial random walk without a mean or trend. The outlier detection algorithm identified all outliers in each data series at the correct time step with intervention parameters, and a default critical value of 3.5. The mean of the modeled residuals and the annualized standard deviations are all close to the original simulation parameters of zero mean and 25% volatility. The normality tests show that residuals fail to reject normality as shown in the lower panel of Table 3. IO is the exception to these where the original volatility is not fully recovered and normality is rejected as well. IO outliers are difficult to simulate and this example uses a deterministic process based on up and down movements decaying to zero. Identifying the outliers produces data that is normal, or closer to normal, and recovers the underlying DGP.

Figure 1, Panels 1 to 4, shows the results of the methodology for each outlier series. The upper graph in Panel 1 is the original and adjusted data series that shows the impact on the path of the mean over time for an AO outlier. The lower graph in Panel 1 shows the AO outlier effects on the mean, which is added to the adjusted data to recover the original simulated series. Panels 2 and 3 show that TC and LS outliers are similar in that the mean shifts up then down temporarily for TC, a transient impact; but for LS the mean shifts up permanently. The lower graphs in Panels 2 and 3 show the outlier effects for TC and LS. The outlier effects of the IO in Panel 4 show the decay that results. The resulting ARMA models estimated by the outlier adjustment algorithm attempts to render the residuals as white noise.

Table 4 reports the model parameter estimates for the base case of a random walk, ARIMA models of best fit, and the outlier adjusted models. The parametric estimations for ARIMA best fit models include only auto-regressive (AR) and moving average (MA) terms and are based on minimization of the AIC (Akaike Information Criteria) and BIC (Bayesian Information Criteria) statistics. The models with outlier intervention parameters and time series ARMA terms are the results of the three stage algorithm described earlier. All models exhibit improved fits as shown by the more negative AIC and BIC statistics and lower RMSE (Root Mean Squared Error) compared to the base random walk model. The estimated models based on the three stage algorithm recover the original simulation parameters for AO, TC, and LS. The IO example has an AR(1) and MA(1) component along with the outlier parameters for the IO and the decay component. A similar case was described by Chen and Liu (1993) for a stationary process where the IO occurs and then decays to zero.

Insert Table 4 here

To illustrate the results in Table 4, we present in the Appendix the estimated models for AO, TC, LS and IO.

4.2. Analysis of Crude Oil and Natural Gas Futures Prices

The data for this analysis is from the CME Group daily settlements for commodity instruments[7] and the contracts are monthly for each commodity. The specific CME instruments are outright futures contracts for NYMEX natural gas (NG) and

[7]CME (2017a), CME (2017b)

WTI crude oil (CL). The data starts on 12/31/2003 and ends on 3/20/2017. The CME Group lists CL futures contracts nine forward years with monthly listing for the current year and following five years. Year 6 and out are listed for the June and December contracts. Additional months are added annually when the December contract expires to keep nine years of the combination of monthly and biannual contracts listed. NG is listed monthly for the current year plus the following twelve calendar years with a new year added when the December contract expires for the current year.

There are 198 CL contracts and 276 NG contracts active during the studied time period. The first expiration for CL contracts is January 2007 and the last expiration is December 2025. For NG contracts, the first expiration is January 2007 and the last expiration is December 2029. There are 196,301 cumulative observations across all CL contracts and 376,429 across all NG contracts.

Table 5 shows summary statistics for all CL and NG contracts.

Insert Table 5 here

The original raw data has the following summary statistics. The average annualized mean return across all CL contracts is -7% and -22% for NG. The range for CL is -45% to 45% and -105% to 32% for NG. For CL, the January 2022 contract has the largest decline in value and the June 2008 contract has the largest increase in value. For NG, the September 2009 contract has the largest loss and the July 2008 contract has the largest gain. The average annualized standard deviation of all CL contracts is 24% with a range from 15% to 38%. The June 2025 and April 2009 are the two contracts on each end of the range. The average annualized standard deviation of all NG contracts is 17% with a range from 9% to 34%. The January 2027 and November 2009 are the two contracts on each end of the range. Overall, both commodity futures exhibit skewness and excess kurtosis. The range for skewness is -1.4 to 0.9 for CL and -1.8 to 1.1 for NG.

The robust estimation with outlier methodology with intervention parameters described in Section 3 was applied to the historical daily CL and NG futures price returns. Table 5, bottom panel, shows results for the estimated models with outlier intervention parameters. The modeled results have the following summary statistics. The average annualized mean return across all crude oil contracts is -0.7% and -53% for natural gas. The range for crude oil is -73% to 50% and -848% to 52% for natural gas. The changes in the mean return summary statistics is difficult to explain, but for natural gas, if the last active contract year 2029 is omitted, the average becomes -0.0005 and the range -0.0014 to 0.0014. The average annualized standard deviation of all crude oil contracts is 22% with a range from 13% to 37%. The average annualized standard deviation of all natural gas contracts is 15% with a range from 7% to 31%. The annualized volatility decreased by over 1.5 standard deviations, and is as high as 2.7 standard deviations for the October 2009 natural gas contract. The range for skewness is -1.0 to 0.3 for crude oil and -0.58 to 1.2 for natural gas. The range for excess kurtosis is -0.6 to 8.0 for crude oil and 0.03 to 9.64 for natural gas. Both skewness and excess kurtosis decreased for both commodities, which is indicative that the residuals are close to white noise.

Normality tests include the Jarque Bera and Shapiro Wilk tests at the 0.001 confidence level. Table 6 summarizes these tests. Before correcting for outliers (see the raw data result), the Jarque Bera (Shapiro Wilk) test show there are 31 (42) of 198 CL futures contracts that failed to reject normality, or 16% (21%) of the contracts. All NG futures rejected the normality assumption. For the models with outlier intervention results in the table, there are 45 (47) of 198 CL futures contracts that failed to reject normality based on the Jarque Bera (Shapiro Wilk) tests, or

23% (24%) of the contracts. This is an increase of 45% (12%) based on the Jarque Bera (Shapiro Wilk) tests. NG, after including outlier intervention parameters, has 22 (31) contracts that failed to reject normality or 8% (11%) with the Jarque Bera (Shapiro Wilk) tests[8].

Insert Table 6 here

The lower section of Table 6 shows that there are 13,481 potential outliers for CL and 15,079 for NG. This is almost 7% of the data for CL and slightly more than 4% for NG. These numbers are well within common Extreme Value Theory observation data reserves of 5% to 15% of the data. Stages 2 and 3 of the algorithm determine the final set of outliers based on outlier t-statistics and model specification tests. There are 1,357 outliers for CL and 3,071 outliers for NG. This represents 0.7% of the data for CL and 0.8% for NG. The final outlier set was reduced almost 90% for CL from the initial candidates and almost 80% for NG. This highlights the need for outlier analysis and is consistent with Chen and Liu (1993)'s statement of fewer but more meaningful set of outliers. If we compare our approach to winsorizing where outliers are usually removed and replaced, we are not removing any data points. The three stage algorithm identifies the crucial set of outliers that have a significant impact on the DGP.

Table 7 presents a summary of outliers by type. The algorithm initially identifies 4,575 AO for CL and 6,054 for NG. However, the final set of AO is reduced by 85% for CL and 76% for NG. Interestingly, IO are not detected in the first stage but the final set includes 310 IO for CL and 618 IO for NG. The biggest reduction in number of outliers is for LS, 99% for CL and 98% for NG; the next is TC, 92% reduction

[8]Improvements in these normality test statistics are consistent with López-de Lacalle (2016) results for consumer price indexes.

for CL and 83% reduction for NG. It is clear from the summary results that the algorithm effectively identifies a set of outliers that have an impact on the DGP, and addressess masking issues that are know to exists with multiple outlier events (Chen and Liu (1993), Fox (1972), López-de Lacalle (2016), and Chang, Tiao and Chen (1988))

Insert Table 7 here

Similar to the simulated data example, we apply the three stage algorithm to the CL and NG data. Table 8 shows summary statistics for the AIC and BIC model specification improvements from the outlier adjusted models (OAM) and best fit ARIMA (BFA) models. The base case is a random walk model. The BFA is superior in 229(127) of 276 NG contracts based on the AIC (BIC) statistic, representing an 83% (46%) of the total NG contracts. The OAM is superior in 258 (253) contacts using the AIC (BIC) statistics, representing 94% (92%) of the contracts. Similar results are found for CL contracts. Based on AIC (BIC), 91% (70%) of contracts exhibited improvement for BFA. The OAM model specification fits shows 99% improvement based on AIC and 94% for BIC. The percentage change medians of the AIC (BIC) statistic for NG is 3.1% (1.3%) for OAM and 2.1% (1.4%) for CL.

Insert Table 8 here

The resulting parametric specifications from the outlier modeling process for all contracts is too numerous to discuss. We present in the Appendix in more detail the November 2020 CL and the May 2022 NG contracts.

4.3. Computation of VaR

To analyze the impact of outliers on the forecast error and address the cost implication, modified VaR risk metrics were calculated based on Cornish-Fischer approximations (see Favre and Galeano (2002)). Table 9 shows the average percentage change in each risk measure as well as the range of each. CL VaR and CVaR decreased on average of 8.6% and 8.9%, respectively with NG decreasing on average of 14.4% and 16.7%, respectively. Notably, the risk metrics of some contracts increased. Table 10 shows that 8 to 15 out of 198 contracts risk increased after outlier intervention model for CL and 15 to 17 out of 276 for NG. Each risk measure is stand alone for a long position in each contract based on one million barrels of CL and 1 Bcf (billion cubic feet which is equivalent to 1,000,000 mmBTU) of NG.

Insert Tables 9 and 10 here

These changes in risk metrics indicate that outlier intervention model is relevant. To get a holistic appraisal of the impact, a VaR / Volatility Elasticity is reported in Table 11. This elasticity is the percentage change in each risk measure divided by the percentage change in residual volatility. Table 11 shows that on average for CL the elasticity is 0.90% to 0.94%, or for a 1% change in volatility, risk decrease by 0.009 to 0.0094. NG elasticity for a 1% change in volatility is 1.2% to 1.3%.

Insert Table 11 here

There are instances where the outlier adjusted risk metrics exceeded the risk metrics based on raw data. This occurred for 5% of CL contracts and 5.5% for NG contracts. These cases could potentially cause serious problems for a firm. Backtests will also suffer showing that the VaR is exceeded, instead of not, more than the predicted number of times per year. This will imply an inadequate risk metric. The distributional characteristics also change, but here the tails are larger than originally estimated with the raw data. As a result, the expected loss if VaR is exceeded, could be much larger than anticipated. These larger losses would require immediate risk capital to be deployed, such as a margin call on exchange traded instruments, posting additional capital on over-the-counter transactions, or being in violation of credit arrangements resulting in technical default. There are credit impacts for a firm that extends credit to counterparties. These risk parameters are used to calculate credit metrics for counter parties and the outlier adjusted parameters could mean the firm has extended more credit to a counterparty than should have been allowed.

We can explore these impacts in more detail by reviewing the contracts for CL and NG that exhibit the maximum increases in the risk metrics for VaR and CVaR. These risk metrics are based on the MTM value of the previously discussed 1,000,000 BBLs of crude oil or 1000 contracts for each crude oil futures contract These contracts were identified as the 2008 June (2008M) and 2008 October (2008V) futures for CL. The MTM value of the 2008M CL portfolio is $129 million and the MTM value of 2008V is $121 million. The base, or raw, Gaussian VaR (GVaR) for these futures positions was $2.9 million and $3.3 million respectively and the intervention modeled GVaR for both was $3.1 million, as seen in Panel A of Table 12.

The Modified Gaussian VaR (MGVaR) for these futures positions was $2.9 and $ 2.8 million respectively with the intervention modeled MGVaR of $3.1 million for both contracts, as shown in Panel B of Table 12. The GVaR for 2008M increased 4.23% and 2008V decreased 7.64%, with the MGVaR increasing for both contracts at 4.22% and 7.47% respectively. The increase in the GVaR for 2008M creates an issue in that the risk measure is under estimating the risk of the portfolio. The decrease in the GVaR for the 2008V does not create an issue, but the increase in the MGVaR does. This is a problem when the Modified GVaR is needed to compute skewness and kurtosis in commodity, non-normal returns.

The GVaR and MGVaR are risk metrics that specify the minimum threshold for a VaR risk loss. These do not estimate the actual loss. The Conditional VAR

(CVaR)and Modified Conditional VaR (MCVaR) are risk measures of the expected loss given that the GVaR limit is exceeded. The CVaR for 2008M exceeded the GVaR by $784 thousand for the raw data case. When the intervention modeled CVaR is compared to the base case, the value is $899M for an additional expected loss of $115 thousand or 3.93% higher,and the MCVaR for this contract was similar. The 2008V contract GVaR metrics decreased meaning that our raw data risk measures estimated risk higher than when intervention modeling in employed. But, this contract exhibited non-normality meaning the Modified GVaR and CVAR would be more appropriate risk measures. The MCVAR was $576 thousand greater than the MGVaR based on raw data and $989 thousand based on the intervention model results, for an additional expected loss of $413 thousand or 14.53% higher. We can see that in cases where estimating risk measures in the presence of outliers can cause serious problems. A firm that has set risk limits and expected loss limits based on raw data can be surprised by the actual occurrence of a loss and final loss that the firm incurs due to biased risk modeling in the presence of outliers.

Similar results are seen with the two NG contracts of 2007 July (2007N) and 2010 April (2010J) where the portfolio for each of these contracts was 1,000,000 mmbtu of NG or 100 contracts. The value of the 2007N portfolio is $6.9 million with a GVaR of $178 thousand, and $3.8 million MTM value for the 2010J portfolio with a GVaR of $107 thousand. The 2007N intervention modeled GVaR increased by 4.80% and 3.38% for the 2010J contract, as seen in Panel C of Table 12. The CVaR for these contracts also increased 4.37% and 3.75% respectively, as seen in Panel D of Table 12. Both of these contract exhibited non-normality in their returns meaning the MGVaR and MCVaR were more appropriate risk measures. The intervention modeled MGVaR increases slightly higher than the GVaR at rates of 4.83% for the 2007N contract and 3.98% for the 2010J contract. These GVaR and MGVaR are risk thresholds or loss limits, as previously mentioned. The CVaR and MCVaR provide an estimated of the loss, and the percentage increases of the intervention model over the raw data for these risk metrics are similar to the GVaR and MGVaR. The CVaR is expected to exceed the GVaR for 2007N by $46 thousand based on raw data and $56 thousand using the intervention model for an additional expected loss of $9,800 or 5.51% higher, and the 2010J contract was 4.69% greater. The MCVaR was $66 thousand greater than the MGVaR for 2007N based on raw data and $56 thousand using clean data for an additional expected loss of $9,992 or 5.56% higher, and the 2010J contract was 5.88%.

The above examples of risk increasing after implementing time series models with outlier intervention parameters show that a firm could potentially be faced with a catastrophic risk event that was completely masked by the outliers. The risk metrics based on the raw data will understate the risk and the actual loses will be greater than expected. The backtests that use the wrong VaR could indicate failures of the risk metrics. Also, contractual covenants, governance, compliance, and other controls may be violated putting the firm at a substantial financial risk.

Insert Table 12 here

5. Conclusion

We show that detecting outliers is an important step in identifying the true DGP from a risk measurement point of view. The algorithm was able to address common issues with outliers of masking/shadowing as seen by the substantial reduction in each contact's set of final outliers from the initial set. The analysis demonstrated that risk could be separated between the DGP and outlier impacts. We showed that managing outliers using time series models of the data with intervention parameters can improve normality tests, which is similar to López-de Lacalle (2016) results.

The analysis showed that risk metrics like VaR can be inaccurately reported, which could impact hedging cost and hedging decisions from the changes in the 2^{nd}, 3^{rd}, and 4^{th} moments of the DGP. The analysis of residual variance or forecast error was similar to Tsay (1988) findings where the 95th percentile decreased by 50% in his research. Our research is the first step towards demonstrating the following four issues in commodity risk management in the presence of outliers: biased statement of risk, additional cost to hedge the risk, inappropriate and inadequate hedges and misstatement of risk associated with extreme events. With inappropriate risk models that do not include outlier intervention impacts, a firm may be reporting flawed risk metrics and not maintaining adequate risk capital. As a result, a firm may be placing itself unknowingly at precarious financial risk.

6. Appendix

6.1. Outlier Simulation Estimated Models for AO, TC, LS and IO

To illustrate the results in Table 4, we present the estimated models for AO, TC, LS and IO. The parametric model estimates from Table 4 specify an outlier adjusted estimated model for AO based on Equations (1) through (6) as:

$$\nabla Y_t = .7893\ I_{AO}^{33} - .7912\ I_{AO}^{67} + e_t,$$
$$I_{AO}^{33,67} = \begin{cases} 1 & \text{if } t\text{=33, 67} \\ 0 & \text{otherwise} \end{cases} \qquad (9)$$

The equation estimated for TC using Equations (1) through (6) and replacing $\frac{1}{(1-.7B)} I_t$

with w_t is:

$$\nabla Y_t = -.6983\ w_t + e_t, \qquad w_t = \begin{cases} 1 & \text{if } t\text{=45} \\ .7w_{t-1} & \text{if } t \geq 45 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

The equation for LS using Equations (1) through (6) and setting $I_t = \dfrac{1}{(1-B)} I_t$ is:

$$\nabla Y_t = 0.1010\ I_t + e_t, \qquad I_t = \begin{cases} 1 & \text{if } t \geq 45 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Using Equations (1) through (6) and rewriting $\dfrac{1}{(1 + 1.2813B + 0.4575B^2)} I_t$ as w_t^{IO} and $\dfrac{1}{(1-.7B)} I_t$ as w_t^{TC}, we obtain the model for IO:

$$\nabla Y_t = -1.2813\ \nabla Y_{t-1} - 0.4575\ \nabla Y_{t-2} + .2356\ w_t^{IO} - .1771\ w_t^{TC} + e_t,$$

$$w_t^{IO} = \begin{cases} 1 & \text{if } t\text{=46} \\ -1.2813 w_{t-1}^{IO} - 0.4575 w_{t-2}^{IO} & \text{if } t \geq 47 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \tag{12}$$

$$w_t^{TC} = \begin{cases} 1 & \text{if } t\text{=47} \\ .7w_{t-1}^{TC} & \text{if } t \geq 47 \\ 0 & \text{otherwise} \end{cases}$$

6.2. November 2020 CL and the May 2022 NG contracts

The resulting parametric specifications from the outlier adjustment process for all contracts is too numerous to discuss. We present in more detail the November 2020 CL and the May 2022 NG contracts. The resulting ARIMA and outlier parametric specifications based on outlier t-statistics and minimizing information criteria for CL are:

$$\begin{aligned}\nabla Y_t = -.001 - 0.4907\ \nabla Y_{t-1} - 0.0718\ I_{AO}^{(179)} + 0.0437 \frac{1}{(1-.7B)}\ I_{TC}^{(176)} \\ + 0.4704\ e_{t-1} + 0.1277\ e_{t-2} + e_t,\end{aligned} \tag{13}$$

and for NG are:

$$\begin{aligned}\nabla Y_t = -0.0007 + 0.1465\ \nabla Y_{t-1} + 0.0292\ I_{AO}^{(6)} + 0.0358\ I_{AO}^{(76)} + 0.0983\ I_{AO}^{(229)} \\ + 0.0364\ I_{AO}^{(323} + 0.053\ I_{AO}^{(659)} + 0.0308\ I_{AO}^{(1010)} - 0.0394\ I_{AO}^{(1031)} + 0.045\ I_{AO}^{(1047)} \\ + 0.0303\ I_{AO}^{(1679)} - 0.1652 \frac{1}{(1-.7B)}\ I_{TC}^{(165)} - 0.0390 \frac{1}{(1-.7B)}\ I_{TC}^{(173)}\end{aligned} \tag{14}$$

$$\begin{aligned}+ 0.0251 \frac{1}{(1-.7B)}\ I_{TC}^{(181)} - 0235 \frac{1}{(1-.7B)}\ I_{TC}^{(201)} + 0.0323 \frac{1}{(1-.7B)}\ I_{TC}^{(846)} \\ + 0.0260 \frac{1}{(1-.7B)}\ I_{TC}^{(1052)} - 0.0237 \frac{1}{(1-.7B)}\ I_{TC}^{1257} - 0.0351(1+0.1465B)\ I_{IO}^{(231)} \\ + 0.0340(1+0.1465B)\ I_{IO}^{(644)} + 0.0584(1+0.1465B)\ I_{IO}^{(657)} + e_t,\end{aligned}$$

The indicator parameter is $I_x^{(k)}$ where x is AO, TC, IO, or LS, and k is the time step of the event where the value is 1 and 0 otherwise.

Bibliography

Bagnai, A., Ospina, C.A.M., 2018. Asymmetries, outliers and structural stability in the us gasoline market. Energy Economics 69, 250–260.

Burman, J.P., Otto, M.C., 1998. Outliers in time series. Bureau of the Census Report, SRD/RR-88/14.

Chang, I., Tiao, G.C., Chen, C., 1988. Estimation of time series parameters in the presence of outliers. Technometrics 30, 193–204.

Chen, C., Liu, L.M., 1993. Joint estimation of model parameters and outlier effects in time series. Journal of the American Statistical Association 88, 284–297.

CME, 2017a. Crude oil futures.

CME, 2017b. Natural gas futures.

Favre, L., Galeano, J., 2002. Mean-modified value-at-risk optimization with hedge funds. The Journal of Alternative Investments 5, 21–25.

Fox, A., 1972. Outliers in time series. Journal of the Royal Statistical Society, Series B (Methodological) 34, 350–363.

Gupta, M., Gao, J., Aggarwal, C.C., Han, J., 2014. Outlier detection for temporal data: A survey. IEEE Transactions on Knowledge and Data Engineering 26.

Hyndman, R., 2017. forecast: Forecasting functions for time series and linear models, r package version 8.0. `http://github.com/robjhyndman/forecast`.

Hyndman, R., Khandakar, Y., 2008. Automatic time series forecasting: The forecast package for r. Journal of Statistical Software 27, 1–22.

Ju, K., Su, B., Zhou, D., Wu, J., Liu, L., 2016. Macroeconomic performance of oil price shocks: Outlier evidence from nineteen major oil-related countries/regions. Energy Economics 60, 325–332.

López-de Lacalle, J., 2016. R package tsoutliers. `https://cran.r-project.org/package=tsoutliers`.

Maravall, A., n.d. Notes on programs tramo and seats.

Marczak, M., Proietti, T., 2014. Outlier detection in structural time series models: The indicator saturation approach. SSRN Electronic Journal .

McCulloch, R.E., Tsay, R.S., 1994. Bayesian analysis of autoregressive time series via the gibbs sampler. Journal of Time Series Analysis 15, 235–250.

Muirhead, C.R., 1986. Distinguishing outlier types in time series. Journal of the Royal Statistical Society. Series B (Methodological) 48, 39–47.

Sanchez, M., Pena, D., 1997. The identification of multiple outliers in arima models. Statistics and Econometric Series 27, Universidad Politecnica de Madrid, Working Paper.

Tsay, R.S., 1988. Outliers, level shifts, and variance changes in time series. Journal of Forecasting 7, 1–20.

Watson, S., Tight, M., Clark, S., Redfern, E., 1991. Detection of outliers in time series. ITS Working Paper 362 University of Leeds.

Table 1: Hypothetical example Gaussian simulation inputs.

Outlier	AO	TC	LS	IO
Up Event	0.8	0.7	0.1	0.7
Down Event	-0.8			-0.7
Decay Factor		0.7		0.8
Time Period (Days) Affected	33,67	45,65	45,100	45,65

We simulate a single random walk without a drift with annualized volatility of 25% for 100 days. We then introduce outliers of each type in the simulation. The AO jumps up 0.8 then back down by the same amount on days 33 and 67. TC are demonstrated with a jump of 0.1 that decays 0.7 each day over days 45 to 65. LS are simulated with a jump of 0.1 on day 45. The decay factor will exponentially decay to zero from the event over the time period.

Table 2: Summary statistics of simulated 100 day returns for base (GBM), AO, IO, TC, and LS examples.

Statistic	Base	AO	TC	LS	IO
Observations	100	100	100	100	100
Minimum (%)	-3.54	-79.12	-67.80	-3.54	-69.59
Median (%)	0.01	0.01	-0.24	8.34	0.01
Arithmetic Mean (%)	0.05	0.05	-2.28	5.65	0.06
Geometric Mean (%)	0.04	-0.94	-3.05	5.52	-0.96
Maximum (%)	3.91	78.93	3.79	13.91	69.93
Standard Deviation (%)	1.59	11.34	9.69	5.29	12.00
Skewness	0.195	-0.037	-4.943	-0.233	-0.165
Kurtosis	-0.294	45.07	26.394	-1.591	19.866
Annualized Standard Deviation (%)	25.24	180.02	153.82	83.98	201.61
Summary Statistics from Outlier Intervention Modeled Results					
Arithmetic Mean (%)	0.052	0.054	0.048	-0.005	0.689
Standard Deviation (%)	1.584	1.594	1.591	1.590	6.430
Annualized Standard Deviation (%)	25.138	25.296	25.259	25.236	102.080

The top panel contains summary statistics based on the simulated data. The bottom panel contains statistics based on the results after processing the data for outliers based on the resulting Outlier Intervention Models.

Table 3: Normality tests for simulated prices.

	Base	AO	TC	LS	IO
		Raw Data Series			
Jarque Beta Statistic	0.995	8,463.588	3,309.870	11.460	1,644.787
	(0.60820)	(0.00000)*	(0.00000)*	(0.00325)	(0.00000)*
Shapiro Wilk Statistic	0.992	0.261	0.397	0.869	0.455
	(0.79073)	(0.00000)*	(0.00000)*	(0.00000)*	(0.00000)*
		Results from Outlier Intervention Models			
Jarque Beta Statistic	0.995	0.913	0.977	1.085	6592.081
	(0.60820)	(0.63346)	(0.61359)	(0.58140)	(0.00000)*
Shapiro Wilk Statistic	0.992	0.991	0.992	0.991	0.589
	(0.79073)	(0.78066)	(0.82752)	(0.71050)	(0.00000)*

Entries are the estimated normality test statistics and their p-values in parentheses.
* denotes significance at the 1% level that indicates rejection of normal hypothesis.

Table 4: Estimated models for base case of random walk, ARIMA best fit, and outlier intervention models.

Model	Statistic	Base	AO	TC	LS	IO
Random Walk						
	RMSE	0.01584	0.11286	0.09905	0.07723	0.12641
	Log-L	272.655	76.263	89.315	114.202	64.928
	df	1.00	1.00	1.00	1.00	1.00
	AIC	-543.31	-150.52	-176.63	-226.40	-127.86
	BIC	-540.70	-147.92	-174.03	-223.80	-125.25
ARIMA	ϕ_0			0.6770	-0.1551	-0.7608
Best				(0.00000)	(0.44951)	(0.00000)
Fit	ϕ_1				-0.5338	
Model					(0.00064)	
	θ_0				-0.4043	
					(0.05176)	
	θ_1				0.4011	
					(0.05096)	
	RMSE	0.01584	0.11286	0.07267	0.01982	0.08135
	Log-L	272.655	76.263	120.490	249.447	109.071
	df	1.00	1.00	2.00	5.00	2.00
	AIC	-543.31	-150.52	-236.98	-488.89	-214.14
	BIC	-540.70	-147.92	-231.77	-475.92	-208.93

continued

Table 4: (Continued)

Model	Statistic	Base	AO	TC	LS	IO
Outlier intervention Model	ϕ_0					-1.2813
						(0.00000)
	ϕ_1					-0.4575
						(0.00000)
	I_0		0.7893*	-0.6983**	0.1010**	0.2356 ***
			(0.00000)	(0.00000)	(0.00000)	(0.00000)
	t-Statistic		50.036	-61.760	47.738	8.931
	I_1		-0.7912*			-0.1771
			(0.00000)			(0.00000)
	t-Statistic		-50.152			-8.525
	RMSE	0.01584	0.01594	0.01591	0.01590	0.06430
	Log-L	272.655	273.039	272.681	272.770	133.584
	df	1.00	3.00	2.00	2.00	5.00
	AIC	-543.31	-540.08	-541.36	-541.54	-257.17
	BIC	-540.70	-532.26	-536.15	-536.33	-244.14

The random walk is the base case with zero mean, no drift DGP, ARIMA is the estimated model of each outlier data series, and Outlier Intervention Model is the final model estimated. ϕ_i is an autoregressive parameter and θ_0 is a moving average parameter. The entries report the estimated model summary statistics. The P-values of the estimates are in parenthesis. I_i designates an indicator function that is 1 or 0 at that time step for AO, TC, and LS, while for IO this is a decay factor for all time steps after the event. * is for times steps 33 and 67, ** is the indicator function starting at time step 45, and *** is the indicator function change at time step 46. t-Statistic is the calculated t-statistic for the outlier.

Table 5: Summary statistic comparison of log returns of the raw and outlier intervention modeled CL and NG commodity contracts.

	CL				NG		
Contracts	198	Min	Max	Contracts	276	Min	Max
Observations	196,301	79	2,213	Observations	376,429	75	2,232
	Average	Range			Average	Range	
Original Data							
Annualized				Annualized		-	
Mean(%)	-7.06	-44.89	45.22	Mean(%)	-22.70	104.81	32.21
Median(%)		-0.14	0.17	Median(%)		-0.09	0.07
Annualized				Annualized			
StDev(%)	23.78	15.42	38.43	StDev(%)	16.77	8.83	33.92
Skewness		-1.40	0.91	Skewness		-1.751	1.123
Kurtosis		-0.56	11.10	Kurtosis		1.009	22.773
Results from Outlier Intervention Models							
Annualized				Annualized			
Mean(%)	-0.68	-72.74	49.93	Mean(%)	-52.58	-847.9	52.32
Median(%)		-0.06	0.18	Median(%)		-2.28	0.12
Annualized				Annualized			
StDev(%)	22.27	13.02	36.92	StDev(%)	15.13	7.19	31.19
Skewness		-1.03	0.32	Skewness		-0.58	1.02
Kurtosis		-0.56	7.97	Kurtosis		0.03	9.64

Table 6: Analysis of the number and percentage of futures contracts that failed to reject normality.

				% of Contracts	
	Contracts	Jarque Bera	Shapiro Wilk	Jarque Bera	Shapiro Wilk
			Raw Data		
CL	198	31	42	16%	21%
NG	276	0	0	0%	0%
		Results from Outlier Intervention Models			
CL	198	45	47	23%	24%
NG	276	22	31	8%	11%
	Initial	% Total	Final	% Total	% Change
CL	13,481	6.87%	1,357	0.69%	-89.93%
NG	15,079	4.01%	3,071	0.82%	-79.63%

Table 7: Summary of initial and final outliers by type.

		AO	IO	LS	TC	Total
CL	Initial	4,575	0	4,827	4,079	13,481
	Final	703	310	33	311	1,357
	Change	-85%		-99%	-92%	-90%
NG	Initial	6,054	0	3,435	5,590	15,079
	Final	1,465	618	62	926	3,071
	Change	-76%		-98%	-83%	-80%

Table 8: Model specification improvement comparing AIC and BIC statistics for random walk, ARIMA best fit, and outlier intervention models.

		ARIMA Best Fit	Outlier Intervention Model	Percentage Change ARIMA Best Fit	Percentage Change Outlier Intervention Model
Natural Gas					
AIC	# Models Improved	229	258	82.97%	93.48%
	Median Change	-5.24	-265.16	0.07%	3.05%
BIC	# Models Improved	127	253	46.02%	91.67%
	Median Change	1.36	-183.67	-0.02%	1.27%
Crude Oil					
AIC	# Models Improved	180	196	90.91%	98.99%
	Median Change	-3.18	-110.82	0.08%	2.11%
BIC	# Models Improved	138	186	69.70%	93.94%
	Median Change	-6.26	-83.41	0.09%	1.40%

These statistics are derived from the estimated models of 276 NG contracts and 198 CL contracts.

Table 9: Percentage change in value at risk and expected shortfall metrics.

Risk Metric	Average	Min		Max	
Crude Oil					
Gaussian VaR	-8.66%	-27.19%	2022.CLF	4.23%	2008.CLM
Modified VaR	-8.91%	-40.39%	2015.CLG	7.47%	2008.CLV
Gaussian CVaR	-8.58%	-26.83%	2022.CLF	3.10%	2008.CLM
Modified CVaR	-8.66%	-26.64%	2022.CLF	12.08%	2008.CLV
Natural Gas					
Gaussian VaR	-15.00%	-48.94%	2029.NGZ	4.80%	2007.NGN
Modified VaR	-16.85%	-65.43%	2029.NGX	6.26%	2022.NGK
Gaussian CVaR	-14.37%	-47.04%	2029.NGZ	4.37%	2007.NGN
Modified CVaR	-14.98%	-56.55%	2029.NGX	4.74%	2010.NGJ

Table 10: Analysis of risk metrics that increased after outlier intervention modelling.

Risk Metric	Risk Change > 0	Percentage Change
Crude Oil		
Gaussian VaR	9	4.55%
Modified VaR	10	5.05%
Gaussian CVaR	15	7.58%
Modified CVaR	8	4.04%
Total Contracts	198	
Natural Gas		
Gaussian VaR	15	5.43%
Modified VaR	17	6.16%
Gaussian CVaR	15	5.43%
Modified CVaR	15	5.43%
Total Contracts	276	

Table 11: Percentage change in VaR/volatility elasticity.

Risk Metric	Average	Min		Max	
		Crude Oil			
Gaussian VaR	0.899	-5.374	2008.CLJ	1.730	2020.CLX
Modified VaR	0.928	-5.462	2008.CLJ	2.860	2015.CLF
Gaussian CVaR	0.921	-4.041	2008.CLJ	1.587	2020.CLX
Modified CVaR	0.940	-4.178	2008.CLJ	1.865	2015.CLJ
		Natural Gas			
Gaussian VaR	1.259	-4.182	2029.NGZ	4.362	2007.NGN
Modified VaR	1.326	-4.285	2029.NGX	6.398	2022.NGK
Gaussian CVaR	1.209	-3.163	2029.NGZ	3.697	2007.NGN
Modified CVaR	1.203	-3.564	2029.NGX	4.012	2010.NGJ

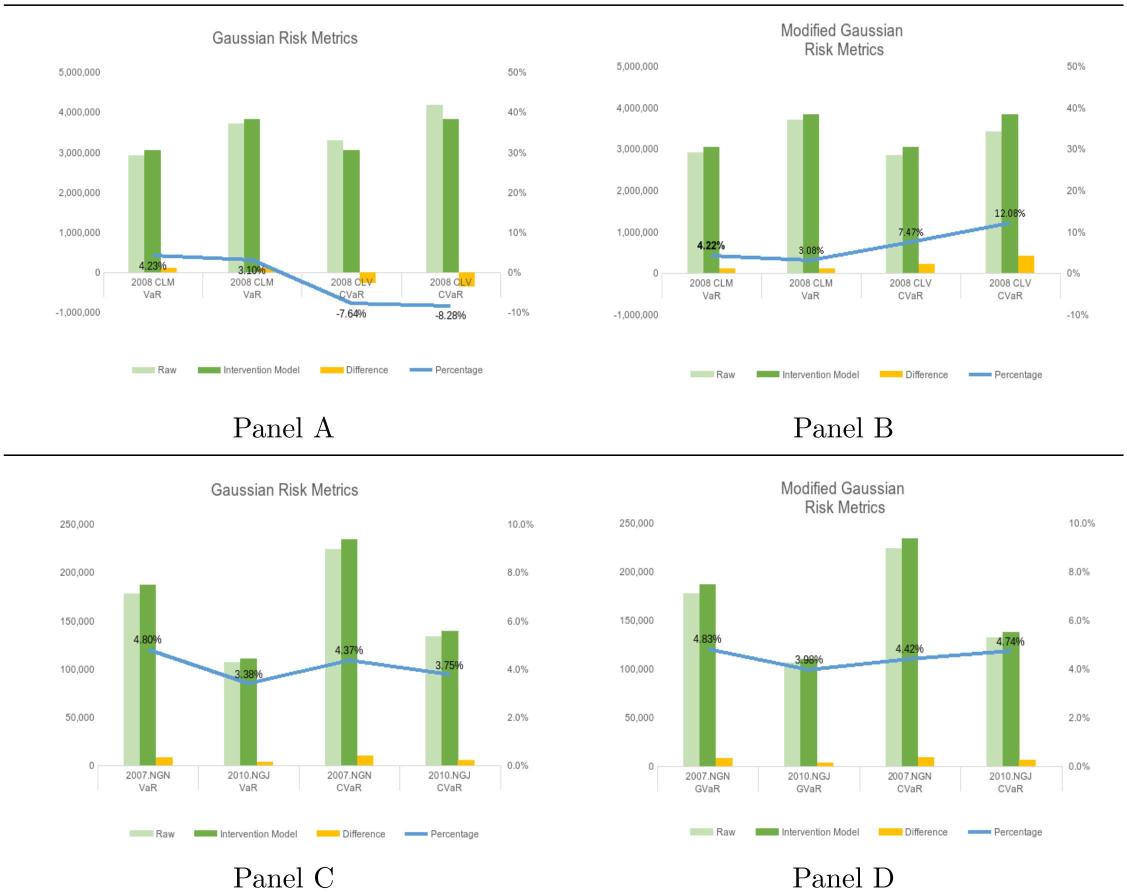
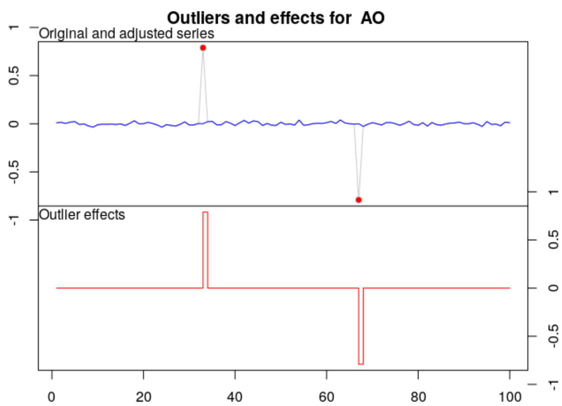
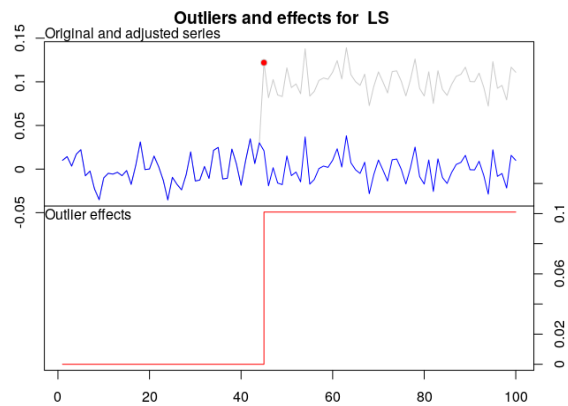


Table 12: Crude Oil and NG Value at Risk metrics for contracts with significant increases from raw and intervention modeled risk metrics.

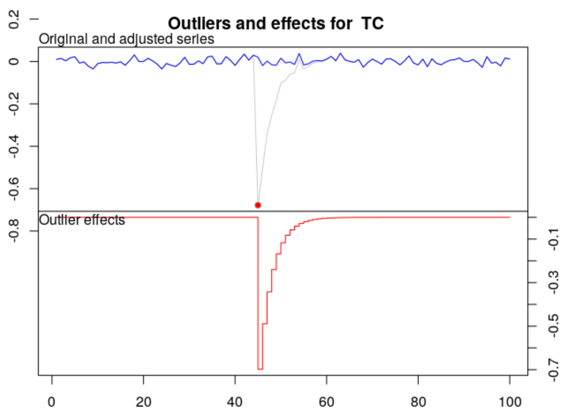
Figure 1: Examples of Outlier and effects for AO, LS, TC, and IO.



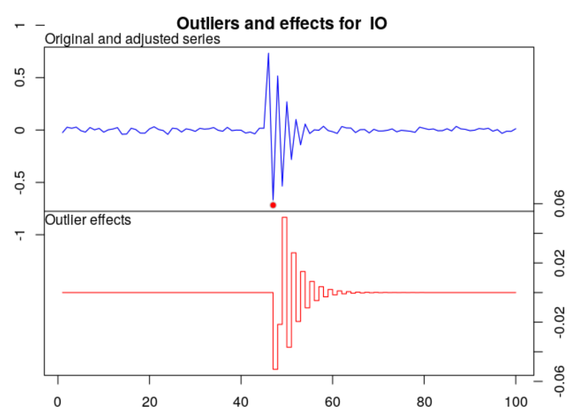
Panel 1 Additive Outlier Example



Panel 2 Level Shift Outlier Example



Panel 3 Temporary Change Outlier Example



Panel 4 Innovative Outlier Example

Figure 2: Example of outlier types.