VPIN, Jump Dynamics, Inventory Announcements in Energy Futures Markets

By

Johan Bjursell*, George H. K. Wang** and Hui Zheng***

August 1, 2016

JEL Classification: G14; G12

Keywords: VPIN, Order flow toxicity, Price jumps, Inventory announcements, Crude oil futures, Natural gas futures

* Credit Suisse, Tokyo, Japan. E-mail: johan.bjursell@gmail.com.
** Finance Area, School of Business, George Mason University, Fairfax, VA USA 22030. E-mail: gwang2@gmu.edu.
*** Department of Finance, Faculty of Business and Economics, University of Sydney, Sydney, Australia, E-mail: hui.zheng@sydney.edu.au.

VPIN, Jump Dynamics, Inventory Announcements in Energy Futures Markets

Abstract

The Volume-Synchronized Probability of Informed Trading (VPIN) metric is proposed by Easley et al. (2011, 2012) as a real-time measure of order flow toxicity in an electronic trading market. This paper examines the performance of VPIN around inventory announcements and price jumps in crude oil and natural gas futures markets with a sample period from January 2009 to May 2015. We have obtained several interesting results: (1) VPIN increased significantly around the inventory announcements with price jumps (scheduled events) and at jumps not associated with any scheduled announcements (unscheduled events). (2) VPIN did not peak prior to the events but shortly after. (3) A minor variation of VPIN based on exponential smoothing significantly improved the early warning signal property of VPIN. Moreover, this estimate of toxicity returns faster to the pre-event level. (4) In general, the VPIN estimate of the toxicity level is higher in natural gas futures than in crude oil futures during our sample period.

1 Introduction

High frequency trading (HFT) accounts for a major portion of trading volume in the U.S. equity and futures markets. In electronic limit order markets, there are no designated market markers, and liquidity arises endogenously from the orders submitted by HFT and non-HFT market participants. Technological advances in computation and communication allow HFT traders to play a crucial role in liquidity supply and demand in the trading environment. For example, Hendershott et al. (2011) present empirical evidence that algorithmic trading improves liquidity for large stocks; and Hasbrouck and Saar (2013) analyze low-latency activity and find that HFT improves market quality measures such as liquidity in the limit order book. Brogaard et al. (2014) provide evidence that HFT trading accelerates price efficiency and provision of liquidity at stressed times such as during the most volatile days. This literature focuses primarily on normal market conditions.

It has been recognized that when HFT participants have significant exposure to large downside market moves and if the toxicity increases, they may become liquidity consumers rather than providers or even abandon market-making activities. This will result in illiquid markets and induce an increase in short term price volatility.[1] Easley et al. (2011, 2012) present the Volume Synchronized Probability of Informed Trading (VPIN) metric as a real-time indicator for measuring "order flow toxicity" faced by market makers in HFT trading environments. The order flow is regarded as toxic when market makers face strong adverse selection risk. They may be unaware of when such market conditions arise resulting in them providing liquidity at a loss. Hence, market markers' estimate of the real time-varying toxicity

[1] May 6, 2010, the so-called market flash crash, is an example (see Kirilenko et al. (2015) and Easley et al. (2012)).

level becomes a crucial factor to managing their liquidity provision. VPIN is a timely new innovation developed to meet the demand to measure the order flow toxicity for market makers, exchanges and regulators. Easley et al. (2012) have successfully demonstrated that VPIN reached the highest level of order flow toxicity in E-mini futures contracts two hours prior to the so-called flash crash on May 6, 2010. They also provided evidence that VPIN achieved very high levels (the cumulative distribution function (CDF) of VPIN was equal or greater than 0.9) on May 5, 2011, when speculators unwound their large speculative positions in WTI crude oil futures. The unwinding of massive positions led them to seek liquidity, and as market makers realized that the selling pressure was persistent, they started to withdraw, which in turn increased the high level of order flow toxicity. Andersen and Bondarenko (2014, 2015), conversely, documented in their empirical investigation that VPIN is a poor predicator of short-run volatility, and that VPIN did not reach an all-time high prior to the flash crash on May 6 but rather following the event. They suggest that the predictive power of VPIN is mainly due to a mechanical relationship with underlying trading intensity. In a rejoinder, Easley et al. (2014) point out there is a confusion with the analysis Andersen and Bondarenko (2014) carry out explaining the contradictory conclusion. Wu et al. (2013) analyze five and half years of data from the 100 most liquidity futures contracts traded worldwide in major exchanges. Their test results confirm that VPIN is a strong predictor of liquidity-induced volatility. With selection of parameter choices, the false positive rates are about 7% averaged over all futures contracts in their data set. When the CDF of VPIN rises above 0.99, the volatility in the subsequent time windows is higher than 93% on average. Using 120 stocks in NASDAQ for 2008 and 2009, Yildiz et al. (2013) document that the order flow toxicity in volume bucket $\lambda - 1$ is positively related to the volatility in bucket λ even after controlling for trade intensity variables. Cheung et al. (2015) study the behavior of VPIN around the mandatory call events of callable bull/bear option contracts at the Hong Kong Option Exchange. They conclude that high VPIN around mandatory call events indicates the existence of large volume imbalances.

In short, there is an ongoing debate on the predictive power of VPIN and future liquidity-induced short run volatility.[2] In general, most previous literature assesses the usefulness of VPIN as a signal for order flow toxicity at a selected single trading day events such as the May 6 market flash crash.[3]

Observers of energy futures markets have long noted that energy futures prices are very volatile and often exhibit jumps (price spikes) at inventory news releases. The theory of storage (see Kaldor (1939), Working (1948, 1949), Brennan (1958), Telser (1959) and others) demonstrates that the level of inventory is one of major factors determining spot and futures prices of consumption-based commodities.[4] Volatility behavior of energy futures prices has been investigated by Mu (2007), Chan et al. (2010) and others. Mu (2007) finds that extreme weather conditions and low inventories are important factors affecting natural gas futures volatility within a single equation model with a GARCH error process. Chen et al. (2010) studies the common jump dynamics in natural gas futures and spot markets within a bivariate autoregressive jump intensity GARCH framework. They particularly examine the role of weather as a short-run demand factor and inventory as a short-run supply factor in explaining price spikes and time varying volatility in spot and futures returns.

Previous papers examining price behavior and volatility surrounding inventory

[2] For other empirical works related to using VPIN refer to Wei et al. (2013) and others.
[3] See Easley de Prado and O'Hara (2012)
[4] Crude oil and natural gas are classified as consumption-based commodities. Furthermore, we should mention that convenience yield has an inverse relationship with level of inventory.

announcements of energy stocks include Linn and Zhu (2004), Gay, Simkins and Turac (2009), Bjursell et. al (2015). Linn and Zhu (2004) report an increase in volatility before and after the release of inventory reports by the Energy Information Administration. Gay, Simkins and Turac (2009) demonstrate that one percent unexpected increase in natural gas inventory results in an approximately one percent drop in the natural gas price. Furthermore, they provide evidence that prices react most strongly to forecasts of analysts with better prior forecast accuracy. Bjursell et. al (2015) apply nonparametric methods to identify jumps in futures prices and intraday jumps surrounding inventory announcements of crude oil, heating oil and natural gas contracts traded on the New York Mercantile Exchange with a sample period of intraday data from January 1990 to January 2008. They obtained several interesting empirical results. (1) Large volatility days are often associated with large jump components and large jump components are often associated with the Energy Information Administration's inventory announcement dates. (2) The volatility jump component is less persistent than the continuous sample path component. (3) Volatility and trading volume are higher on days with a jump at the inventory announcement than on days without a jump at the announcement. Furthermore, the intraday volatility returns to normal faster following inventory announcements with jumps than after announcements without jumps. Based on previous results, we can expect that the order flow becomes more toxic due to high volatility and trading volume during inventory announcement periods. Therefore, we have an ideal empirical test setting for examining the performance of VPIN as a real-time indicator of order flow toxicity and early warning indicator for market turbulence around repetitive scheduled information and liquidity events.

The major purposes of this paper are twofold: (1) We examine the behavior of VPIN around inventory announcements with price jumps (scheduled events) and price jumps not associated with scheduled events in crude oil and natural gas futures markets during a recent sample period spanning from January 1, 2009 to May 31, 2015; and (2) we propose a minor variation of VPIN by applying exponential smoothing in the last stage of the calculation. We believe this will increase the sensitivity of VPIN to capture recent order flow toxicity. We have obtained several interesting results:

1. VPIN estimates increase significantly around the inventory announcements period with price jump (scheduled events) as well as at jumps not associated with any inventory announcements (unscheduled events).[5]
2. VPIN does not peak prior to the scheduled inventory announcements but rather after these events.
3. In general, the values of VPIN in natural gas futures are higher than the VPINs in crude oil futures during our sample period. These results are consistent with previous finding by Bjursell et. al (2015) that volatility in natural gas futures are higher than volatility in crude oil futures.
4. A minor variant of VPIN with exponential smoothing significantly improves the early warning signals property of the VPIN and the modified VPIN estimate returns faster to normal levels after the events.
5. The updated version of VPIN proposed in this paper outperforms VPIN in temporal linkages between toxicity and short-term volatility in terms of alternative correlation measures and a conditional probabilities framework.

The organization of the paper is as follows. Section 2 presents the empirical methodology on

[5] Unscheduled events refer to jumps which cannot be associated with any event listed in Bloomberg's economic calendar.

identification of intraday price jumps and estimation of the VPIN metric. Section 3 discusses inventory announcements, the contract specifications and the data. Section 4 presents empirical results. We conclude the paper in Section 5.

2 Empirical Methodology

This section consists of two parts. Section 2.1 presents the statistical procedure used to identify the intraday timing of price jumps. Section 2.2 describes the computational algorithm of Volume-Synchronized Probability of Informed Trading (VPIN) metric proposed by Easley et al. (2012) (ELO) and a new proposed minor variation of the metric.

2.1 Asset Price Dynamics and Jumps Statistics

Let $X_t = \log S_t$ denote the logarithmic price where S_t is the observed price at time *t*. Assume that the logarithmic price process follows a continuous-time diffusion process X_t coupled with a discrete process defined as,

$$\mathrm{d}X_t = \mu_t \mathrm{d}_t + \sigma_t \mathrm{d}W_t + \kappa_t \mathrm{d}q_t \qquad (1)$$

where μ_t is the instantaneous drift process and σ_t is the diffusion process; W_t is the standard Wiener process; $\mathrm{d}q_t$ is a Poisson jump process with intensity λ_t, that is $\mathrm{P}(\mathrm{d}q_t = 1) = \lambda_t \mathrm{d}_t$; and κ_t is the logarithmic size of the price jump at time *t* if a jump occurred. If X_{t-} denotes the price immediately prior to the jump at time *t*, then $\kappa = X_t - X_{t-}$.

We use a nonparametric test developed by Lee and Mykland (2008), which identifies the significant intraday jump returns and thus provides the intraday arrival time, realized size and direction of jumps. The test statistic is applied to the intraday continuous returns, r_{t_i}, by comparing their magnitude to the local variation (or instantaneous volatility) of the return process at time t_i. Specifically, the realized intraday return is compared to an estimate of the instantaneous volatility of the price process observed immediately prior to the tested return, r_{t_i}. Lee and Mykland (2008) suggest estimating the volatility by using a variation of the realized bipower variation, (see Barndorff-Nielsen and Shephard (2004) and (2006)),

$$BV_t = \frac{\pi}{2}\frac{m_t}{m_t - 1}\sum_{j=0}^{m_t} \left|r_{t_j}\right| \left|r_{t_{j-1}}\right|, \quad (2)$$

which is robust to jumps. The jump detection statistic is calculated as $L_{t_i} = r_{t_i}\hat{\sigma}_{t_i}^{-1}$ where $\hat{\sigma}_{t_i}^{1} = (K-2)^{-1}\sum_{j=i-K+2}^{i-1} \left|r_{t_j}\right| \left|r_{t_{j-1}}\right|$. Hence, the volatility is estimated based on the K intraday returns preceding $r_{t_i}$where a sufficiently large window size is chosen so that the impact from previous jumps is minimized. Lee and Mykland (2008) report that there exists a range of values of *K* such that larger values only make a marginal contribution. The appropriate choice of *K* depends on the sampling interval. We apply the statistic to five-minute intraday returns, and follow the recommendation by Lee and Mykland (2008) and calculate the statistic based on the past 270 returns.

Lee and Mykland (2008) obtain a rejection region by deriving the limiting distribution of the maximum of the statistic under the null hypothesis of no jump. The statistic is calculated as ($|L_t|$ $- C_n) / S_n$ where,

$$C_n = \frac{(2\log n)^{\frac{1}{2}}}{c} - (\log \pi + \log\log n)/\left(2c(2\log n)^{1/2}\right), \quad (3)$$

$c = \sqrt{2}/\sqrt{\pi}$ and $S_n = 1/\left(c(2\log n)^{1/2}\right)$. The cumulative distribution function is derived as, $P(\xi \leq x) = e^{-e^{-x}}$. Thus, for a given significance level, we can solve for *X* to determine the threshold for significant jumps. We report empirical results for the one percent significance level. Hence, we reject the null hypothesis of no jump for values of the maximum statistic larger than $\beta_{0.01} = -\log(-\log(0.99)) = 4.60$.

2.2 VPIN Metric

Volume-Synchronized Probability of Informed Trading (VPIN) is calculated following the algorithm described in Easley et al. (2012) and Abad and Yagüe (2012), and outlined here.

1. Time bars: Initially, trades are aggregated based on one-minute intervals into time bars. We also produce results based on ten-second intervals. The trade volume is aggregated per time bar and the closing price is recorded in order to calculate the return per time bar. The overnight return is omitted; instead, the open to close return is used for the first time bar per day. The trade activity within the time bar is then treated as if the contracts were traded at the closing price and thus have the same return.
2. Volume buckets and bulk classification: A volume bucket is obtained by adding trading volume from consecutive time bars until the total volume reaches the volume bucket size (VBS). Hence, depending on the trade activity, a volume bucket may require multiple time bars or just a fraction of one time bar. The remaining trades from a time bar are applied to the subsequent volume bucket. VBS is set to the average number of daily traded contracts divided by 50 following ELO's work.[6]
3. The trade direction is then determined per time bar in probabilistic terms where the buy volume is obtained by multiplying the trade volume by $Z(\Delta P / \sigma\Delta P)$ where $\sigma\Delta P$ denotes the standard deviation of all price changes for the whole sample. Similarly, the sell

---

[6] The results reported are based on using the daily average across the whole sample. We also divided the sample into two subsets and updated the VBS based on these but obtained qualitatively analogous results.

volume is given by the volume multiplied by $1 - Z\,(\Delta P\,/\sigma\Delta P)$. The order imbalance, OI_λ, is then calculated as the absolute difference between buy and sell volumes.

4. Finally, VPIN is calculated based on n consecutive volume buckets and is given by

$$\frac{\sum_{i=1}^{n} \mathrm{OI}_\lambda}{\mathrm{nVBS}} \qquad (3)$$

The time series of VPIN estimates are obtained using a moving window of volume buckets. That is, the first VPIN is calculated using the volume buckets [1, n]. The next estimate is based on [2, n + 1] and so on.

We can rewrite the VPIN equation (3) as follows,

$$\mathrm{VPIN} = \sum_{\lambda=1}^{n} \frac{1}{n} \frac{\mathrm{OI}_\lambda}{VBS} \qquad (4)$$

From equation (4), we see that VPIN is based on a simple moving average with equal weight (1/n) given to current and past observations. VPIN is designed to have a forecasting property; thus it may be desirable to give more weight to recent observations in order to capture newly arrived information. For this reason, we propose to use an exponential weighted moving average to calculate VPIN instead of a simple moving average with equal weights. The VPIN with exponential smoothing, $\mathrm{EXPS\_VPIN}_\alpha$, where α is the smoothing constant is described as follows. Let,

$$v_\lambda = \frac{OI_\lambda}{\mathrm{VBS}}, \qquad (5)$$

$\mathrm{EXPS\_VPIN}_{\alpha=0.1}$ is then defined as,

$$\mathrm{EXPS\_VPIN}_{\alpha,\lambda} = \alpha v_\lambda + (1-\alpha)\mathrm{EXPS\_VPIN}_{\alpha,\lambda-1} \qquad (6)$$

Given a moving window of size n, the initial value of $\mathrm{EXPS\_VPIN}_\alpha$ is then based on the first n values of v_λ. We need to select the smoothing constant α. The higher value of α, the more weight is given to the current and most recent observations.[7] In this paper, we specify α=0.1 (EXPS_VPIN$_{\alpha=0.1}$) and a moving window of n = 50 observations to calculate VPIN.

3 Contract Specifications, Data and Inventory Announcements

3.1 Contract Specifications and Data

We examine price series for two contracts from the U.S. energy futures markets. The contracts are on natural gas and crude oil which are traded on the New York Mercantile Exchange (NYMEX).

The futures contract on crude oil began trading in 1983. The contract calls for delivery of both domestic as well as international crude oils of different grades in Cushing, Oklahoma. The contract, which is listed nine years forward, trades in units of 1,000 U.S. barrels (42,000 gallons) and is quoted in U.S. dollars and cents per barrel.

The natural gas futures contract began trading on April 3, 1990 and is based on delivery at the Henry Hub in Louisiana.[8] The contract trades in units of 10,000 million British thermal units (mmBtu) and is quoted in dollars and cents per mmBtu. Contracts are traded for about thirteen years forward (the current calendar year plus the next twelve years). Appendix A.I presents detailed specifications for these contracts.

The price series range from January 1, 2009 to May 31, 2015. Each transaction includes a date and time stamp and the transaction price. Since January 31, 2007, the trading hours have

---

[7] Further discussions on exponential smoothing moving average procedures and the statistical properties are referred to in Brown (1962), Chatfield et al. (2001) and Diebold (2007).

[8] The natural gas futures contract is commonly cited as the benchmark for the spot market, which accounts for nearly 25 percent of the energy consumption in the U.S.

been set at 9:00 AM to 2:30 PM. The contracts began trading electronically via the Globex trading platform in the spring of 2007. The electronic trading became consistently higher than pit trading around September of 2007 for these contracts, and has since remained the predominant trading platform. Hence, we use prices from this platform. The electronic trading takes place from 6:00 PM to 5:45 PM the following day; however, for consistency we consider only the transactions for the same hours during which the pit trading takes place as this is the most liquid time. Furthermore, we use the data series from nearby contract months. During the maturity month, we shift to the first deferred contract month, using the daily trade volume as the switch indicator. The data are filtered to limit any biased results due to illiquid trading.

3.2 Inventory Announcements

EIA releases weekly reports on the inventory status of crude oil and natural gas. Since 2003, a smaller version of the inventory report for crude oil with highlights and summarizing tables is released at 10:30 AM on Wednesdays; a full report is published after 1:00 PM on the same day. The EIA also compiles and releases a weekly natural gas storage report with estimates of natural gas in underground storage. EIA releases the report at 10:30 AM on Thursdays.[9]

Data on the market's expectation on the weekly changes in inventories in these commodities are obtained from Bloomberg, who reports weekly surveys of market analysts' forecasts of the inventory levels. The reports include statistics such as mean, median, low and high values of the forecasts where the number of analysts ranges from fifteen to thirty for these commodities. The surveys also include actual inventory levels and, hence, allow us to obtain the surprise at time t defined as the difference between the actual value, A_t, and the consensus forecast, F_t, where the

[9] Further discussion on the inventory reports is referred to EIA's website: http://www.eia.doe.

median is chosen as the forecast. Since small differences between actual and forecast values can be expected without materially impacting the market, we focus on significant surprises, which we define as surprises larger than one standard deviation, σ_t (i.e., standard deviation of the differences between actual and forecast values). To compare the size of surprises across products, we define standardized surprises as,

$$\frac{(A_t - F_t) - \mu_{surp,t}}{\sigma_t} \qquad (7)$$

where $\mu_{surp,t}$ is the mean of the differences between actual and forecast.

4 Empirical Results

Section 4.1 reports summary statistics of VPIN, and the time series behavior of VPIN and price returns over the sample period. Section 4.2 documents the toxicity metrics' behavior on a particular day, and their properties around price jumps at inventory announcements and jumps not associated with any scheduled event. We examine temporal linkages among toxicity, liquidity and return volatility in Section 4.3

4.1 Exploratory Data Analysis

Figure 1 plots daily time series of VPIN and daily continuous returns for crude oil (Panel A) and natural gas (Panel B). The black lines are the continuous returns based on close to close prices. The green lines denote the daily VPIN represented by the last VPIN estimate per day. The time series suggest that higher values of VPIN can often be associated with volatile market conditions.

[Insert Figure 1]

Table 1 summarizes the number of trading days with significant positive and negative jumps per year, and shows that (1) natural gas futures have a greater number of jumps than crude oil futures; and (2) there are more negative than positive jumps in both commodities. These empirical results are consistent with empirical results reported by Bjursell et al. (2015) using sample data from 1990 to January 2008.

[Insert Table 1]

Table 2 reports statistical properties of VPIN and EXPS_VPIN$_{\alpha=0.1}$ for crude oil and natural gas futures for the complete sample period and two sub periods. Results denoted by EXPS_VPIN$_a$ are based on the sample period from 2009 to 2011 and EXPS_VPIN$_b$ from 2012 to 2015, May 31. Based on skewness and kurtosis values, we reject that VPIN and its variants have a normal distribution. The result of the Augmented Dickey-Fuller test confirms that they are stationary time series. We also include the 0.1, 0.25, 0.5, 0.75 and 0.9 percentiles of the empirical distributions of VPIN and EXPS_VPIN$_{\alpha=0.1}$. Comparing EXPS_VPIN$_{\alpha=0.1,a}$ and EXPS_VPIN$_{\alpha=0.1,b}$ show that the distributions have been relatively stable over time. Furthermore, VPIN is higher for natural gas as well as more volatile.

[Insert Table 2]

Table 3 reports the average number of contracts traded for crude oil and natural gas futures for the whole sample period and two the sub periods. We observe that both contracts have been more actively traded over the last three years as indicated by the increased average daily trading volume. Following Easley et al. (2012), we set the daily number of volume buckets to 50. That is, the volume buckets size (VBS) is equal to average daily volume divided by 50. VBS has increased in both commodities over the latter half of the sample data. Nevertheless, the main conclusions henceforth do not change based on whether the empirical analysis is based on the VBS on all data or updated per subset. We only include analysis based on the whole sample to preserve space.

[Insert Table 3]

4.2 The Behavior of VPIN around News Events and Price Jumps

In this section we first examine VPIN versus $EXPS\_VPIN_{\alpha=0.1}$ on May 5, 2011, for crude oil futures. Easley et al. (2011) analyzed VPIN's behavior on this day when there was a large selling pressure due to market participants taking profits. Table 4 presents details of the behavior of the intraday returns, VPIN, $EXPS\_VPIN_{\alpha=0.1}$ and ECDF(VPIN) and ECDF($EXPS\_VPIN_{\alpha=0.1}$) for this day. Figure 2 plots the intraday dynamics of VPIN and $EXPS\_VPIN_{\alpha=0.1}$ for May 5, 2011, in the crude oil market.

[Insert Table 4 and Figure 2]

We observe that the intraday returns (continuous line in the top panel) starts falling shortly after the open but initially at a relatively slow pace. Around 10:40AM, the return begins to drop faster and continues to drop until 11:12AM, after which the price stabilizes for a while and even increases a bit before dropping for the remainder of the day. Referring to the level and ECDF of $EXPS\_VPIN_{\alpha=0.1}$ (dashed lines in the middle and bottom panels), the toxicity starts increasing rapidly around 10:15AM and peaks around 10:55AM. At this time the futures markets is dropping quickly but has yet to drop 2% in the next 15 minutes and 4% for the day; hence $EXPS\_VPIN_{\alpha=0.1}$ provides an indication to widen bid-asks spreads significantly or even get out of the market. While both VPIN and $EXPS\_VPIN_{\alpha=0.1}$ start increasing rapidly shortly before 11AM, EXPS VPIN$_{\alpha=0.1}$ increases faster by putting more weight on more recent observations. EXPS VPIN$_{\alpha=0.1}$ peaks for the day at around 11:15AM whereas VPIN peaks around 12:00AM. In short, we find that EXPS VPIN$_{\alpha=0.1}$ significantly improves the early warning signals in comparison with VPIN for crude oil futures on May 5, 2011.

[Insert Table 4 and Figure 2]

Next, we broaden the analysis by looking at the behavior and predictive power of VPIN versus EXPS_VPIN$_{\alpha=0.1}$ to detect adverse conditions at inventory releases. In particular, we test whether VPIN on average increases prior to these events conditioned on significant jumps and surprises in the announcement. We use a one-way analysis of variance regression model to estimate the mean behavior of VPINs around inventory announcements with negative jumps,

$$\text{VPIN}_{i,k} = \beta_0 + \sum_{j\in J} \beta_j D_j + \epsilon \quad , \qquad (8)$$

The dependent variable VPIN$_{i,k}$ is the i$^{\text{th}}$ VPIN estimate associated with the kth event. Since VPIN estimates occur asynchronously, they are indexed in order relative to the timing of the event. The dummy variable D_j is equal to one for the j^{th} VPIN observation where j =-19,..., -1,1,...,60 and zero otherwise. That is, we include 20 VPIN estimates prior to the event and 60 following the event where D_{-1} and D_1 are associated with the VPIN estimates just prior to and after the event. D_{20} serves as the benchmark assumed to be absent any information about the content of the event; hence, the value of D_{20} is the intercept in the regression. We include 60 subsequent observations since VPIN is calculated based on a moving window with 50 observations.

Table 5 presents the regression results for crude oil and natural gas for the scheduled event when the inventory report is released. We only consider cases with a surprise in the forecast greater than one standard deviation and a negative and significant jump at the time of the release. There are three sets of results per commodity. Column I gives the results for the original VPIN calculation using a simple average based on the 50 past observations. Columns II and III report estimates using exponential moving average with smoothing parameter α set to 0.05 and 0.10. Panels A (crude oil) and C (natural gas) in Figure 3 plot the mean values (coefficient of intercept + coefficient of dummy variable) of VPIN, $EXPS\_VPIN_{\alpha=0.05}$ and $EXPS\_VPIN_{\alpha=0.1}$.

Table 6 documents the results of the regression model for VPIN versus $EXPS\_VPIN_{\alpha=0.05}$ and $EXPS\_VPIN_{\alpha=0.1}$ at jumps which are not associated with any inventory reports. The time series behavior of the mean values of VPIN, $EXPS\_VPIN_{\alpha=0.05}$, and $EXPS\_VPIN_{\alpha=0.1}$ derived from the regression coefficients reported in Table 6 are plotted in Panels A (crude oil) and C (natural gas) in Figure 4.

Table 7 reports estimates from the regression model, equation 8, with the empirical cumulative densities (ECDF) of VPIN, $EXPS\_VPIN_{\alpha=0.05}$ and $EXPS\_VPIN_{\alpha=0.1}$ for the same events as considered in Table 5. Table 8 presents the equivalent results for jumps which are not associated with any scheduled events. Panels B and D in Figures 3 and 4 plot these estimates where the plotted lines are mean values of ECDF(VPIN) and ECDF($EXPS\_VPIN_{\alpha=0.1}$) around inventory announcements derived from the regression results in Tables 7 and 8.

Summarizing the regression results above, we first note that the values of VPIN and $EXPS\_VPIN_{\alpha=0.1}$ increased significantly around the inventory announcements period with price jumps (scheduled events) and jumps without inventory announcements (unscheduled events). These results suggest that toxicity metrics can provide a signal that order flow becomes more toxic around the release of new events and at price jumps.

Second, on average, VPIN and $EXPS\_VPIN_{\alpha=0.1}$ did not reach local maxima prior to these events but shortly after. Comparing the two toxicity metrics, we see that VPIN did not reach its highest value until around the 25th observation for crude oil after the new release with jumps while $EXPS\_VPIN_{\alpha=0.1}$ reached its highest value at the 5th observation. Furthermore, the value of $EXPS\_VPIN_{\alpha=0.1}$ is in most cases higher than VPIN prior to the events and also in a few instances statistically different from the benchmark value prior to the jumps in crude oil futures. Similar results hold for natural gas futures in Table 5 and Figure 3 as well as for price jumps without inventory releases reported in Table 6 and Figure 4. Thus, these empirical results support that $EXPS\_VPIN_{\alpha=0.1}$ improve on signaling for adverse order flow over VPIN.

Third, the ECDF results from Table 7 and Panels B and D in Figure 3 corroborate the findings based on the level in toxicity. As ELO emphasize, it is more appropriate to refer to the ECDF of the toxicity metrics rather than the absolute level in toxicity to identify critical levels. We observe that ECDF(VPIN) never goes above 0.9 whereas ECDF($EXPS\_VPIN_{\alpha=0.1}$) surpasses 0.98 shortly after the inventory report is released. Table 8 and Panels B and D in Figure 4 for unscheduled events show a similar story though the ECDF values do not reach as high levels.

In summary, we find that VPIN is a useful tool to signal periods of turbulent price behaviors during news releases and price jumps, but VPIN does not demonstrate its early warning signal property in crude oil and natural gas futures. Hence, our results are consistent with the finding of Andersen and Bondarenko (2014, 2015)'s assessment based on S&P 500 E-mini futures. The proposed variation of VPIN ($EXPS\_VPIN_{\alpha=0.1}$) strengthens the signal of toxicity and also returns faster to the normal pre-event levels following the turbulent period.

4.3 Temporal Linkages among Toxicity, Liquidity and Return Volatility

It is generally agreed that as order flow toxicity increases, market makers face potential losses, and so may opt to reduce providing liquidity or even abandoning market-making activities. The ensuing reduction in liquidity in turn suggests that high levels of VPIN should signal greater short run volatility induced by shortage of liquidity. In this section, we follow ELO (2013)'s methodology to examine these temporal linkages among VPINs, liquidity and return volatility.

As a first step, Table 9 reports Pearson correlations of $|r_\lambda|$ with the variables $\text{VPIN}_{\lambda-1}$, $\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}$, $\text{RV}_{\lambda-1}$ (realized volatility), $\text{NT}_{\lambda-1}$ (number of traders), $\text{TS}_{\lambda-1}$(trade size) and BAS_λ (bid-ask spread). Notice that all variables but $|r_\lambda|$ and BAS_λ are lagged one period in λ. Panel A tabulates the point estimates of the Pearson correlations with the associated 95% confidence intervals in Panel B. We use Fisher Z-transformation to construct these confidence intervals[10]. We highlight two observations. First the point estimates of $corr\left(\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}, |r_\lambda|\right)$ are greater than $corr(\text{VPIN}_{\lambda-1}, |r_\lambda|)$ in both crude oil and natural gas futures. Furthermore, they are statistical different at the 95% level since the corresponding confidence intervals of $corr\left(\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}, |r_\lambda|\right)$ and $corr(\text{VPIN}_{\lambda-1}, |r_\lambda|)$ do not intersect. Second, $corr\left(RV_{\lambda-1}, |r_\lambda|\right)$ are greater than both $corr\left(\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}, |r_\lambda|\right)$ and $corr(\text{VPIN}_{\lambda-1}, |r_\lambda|)$ in both commodities futures.

[Insert Table 9]

Since both series are auto correlated, the criterion for applying the Fisher Z-transformations is not fully satisfied. Hence, to further explore and validate the correlation structure, we carry out a pair analysis on the differences of the moving window correlations,

[10] The rationale and the procedure of using Fisher Z transformation to construct confidence interval for non-zero correlations in population is referred to Snedecor and Cochran (1980).

$corr(\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}, |r_\lambda|)$ and $corr(\text{VPIN}_{\lambda-1}, |r_\lambda|)$ by employing a window length set to 50 volume buckets. Figure 5 displays the time series behavior of the difference of the two moving window correlations over the sample period. The paired t test result for crude oil is 217 with the mean difference equal to 0.12; for natural gas, the t tests is 254 and mean difference 0.16. These results support that $corr(\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}, |r_\lambda|)$ are statistically and significantly greater than $corr(VPIN_{\lambda-1}, |r_\lambda|)$ in both futures markets.

[Insert Figure 5]

We apply the cross correlation technique to examine lead and lag correlations between VPIN and EXPS_VPIN$_{\alpha=0.1}$ with $|r_\lambda|$ at various lags in λ. The cross correlations are defined as follows:

$$C_{x,y} = \sum_{\lambda=1}^{\lambda-k} (x_\lambda - \bar{x})(y_{\lambda+k} - \bar{y}) / \lambda \quad ,k = 1,2,\ldots 50 \quad k = -1,-2,\ldots -50.$$

where $x_\lambda = |r_\lambda| and \quad y_{\lambda+k} = VPIN_{\lambda+K} \quad or \quad , \quad y_{\lambda+k} = EXPS\_VPIN_{\lambda+k}$

Figure 6 presents cross correlations for crude oil (Panel A) and natural gas (Panel B) between absolute returns and the toxicity metrics. The x-axis labels denote lags in λ between the absolute returns and VPIN. We draw attention to two interesting results. First, the correlation of EXPS_VPIN$_{\alpha=0.1}$ for negative lags are greater than corresponding correlations of VPIN for first ten negative lags while the reverse relationship holds from the 11$^{\text{th}}$ and onwards in both futures markets. This is not surprising since EXPS_VPIN$_{\alpha=0.1}$ are more sensitive to recent observations of volume buckets than VPIN which assigns equal weight to all buckets within the window. Second, the positive lags, which denote that the absolute return is lagged to VPIN

(EXPS_VPIN$_{\alpha=0.1}$), are substantially lower than the cross correlations for negative lags. These results suggest that VPIN (EXPS_VPIN$_{\alpha=0.1}$) plays a stronger leading indicator to absolute returns while the feedback effect is relatively weaker.

[Insert Figure 6]

Following the methodology laid out by ELO (2013), we also evaluate the temporal linkages between VPIN (EXPS_VPIN$_{\alpha=0.1}$) and $|r_\lambda|$ in a model-free framework based on conditional probabilities. By obtaining conditional probabilities, we aim to answer the questions: (1) given high values of $\text{VPIN}_{\lambda-1}$ (EXPS_VPIN$_{\alpha=0.1,\lambda-1}$), what is the subsequent behavior by $|r_\lambda|$; and (2) when $|r_\lambda|$ is high, what is the probability distribution prior to that in $\text{VPIN}_{\lambda-1}$(EXPS_VPIN$_{\alpha=0.1,\lambda-1}$).

We first create a discrete joint probability distribution between $|r_\lambda|$ and $\text{VPIN}_{\lambda-1}$ (EXPS_VPIN$_{\alpha=0.1,\lambda-1}$) by grouping VPINs into five-percentile bins and absolute returns $|r_\lambda|$ in bins of size 0.25% up to 2%. From the discrete joint distribution, we derive the two types of conditional probability distributions.

Panels A and B in Table 10 (Table 11) present $prob(|r_\lambda|/VPIN_{\lambda-1})$ and $prob(|r_\lambda|/EXPS\_VPIN_{\lambda-1,\alpha=0.1})$ for crude oil (natural gas) respectively. There are twenty conditional distributions for the toxicity metrics lagged by one period relative to the absolute value of returns. Summarizing the major findings, we first note that a low value of $\text{VPIN}_{\lambda-1}$ (EXPS_VPIN$_{\alpha=0.1,\lambda-1}$) is often associated with relative high probabilities of low values of $|r_\lambda|$. For example, when VPIN is in the 0.05 percentile, the probability associated with the 0-0.25%

bin for absolute returns is 95%, while there is a zero probability for bins of $|r_\lambda|$ greater than 1.0%. Second, as the value of $\text{VPIN}_{\lambda-1}$ ($\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}$) increases, the conditional distribution of subsequent $|r_\lambda|$ gradually disperses across the buckets. Referring to the 0.90 percentile (Table 10 Panel A), the probability associated with the 0.25% bin for $|r_\lambda|$ is reduced to 70.2% and the remaining 29.8% is distributed over larger values of absolute returns. Third, low values of $\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}$ are coupled with higher probabilities of low values of $|r_\lambda|$ compared to the probabilities associated with $\text{VPIN}_{\lambda-1}$ and $|r_\lambda|$. Furthermore, it remains true that high values of $\text{EXPS\_VPIN}_{\alpha=0.1}$ are associated with lower probabilities of low values of $|r_\lambda|$ as well as relatively high probabilities are associated with high values of $|r_\lambda|$ compared to the analogous relationships between $\text{VPIN}_{\lambda-1}$ and $|r_\lambda|$. For example, looking at the results for natural gas in Panel A in Table 11, given the 1.00 percentile value of $\text{VPIN}_{\lambda-1}$, the associated probabilities with the 0.25% and 1.50% bins of absolute returns are 58.1% and 0.8 % respectively. The analogous values for $\text{EXPS\_VPIN}_{\alpha=0.1}$ are 47.9% and 2.0% (see Panel B in Table 11).

[Insert Table 10 and 11]

Table 12 and 13 present $prob(VPIN_{\lambda-1} / |r_\lambda|)$ in Panel A and $prob(EXPS\_VPIN_{\lambda-1} / |r_\lambda|)$ in Panel B for crude oil and natural gas, respectively. There are nine conditional probability distributions for given values of $|r_\lambda|$. Two important findings stand out. First, conditioned on the range 0-0.25% of $|r_\lambda|$, the probabilities associated with prior values of $\text{VPIN}_{\lambda-1}$ ($\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}$) are low and dispersed nearly uniformly across the range of percentiles. However, for bins greater than 1.50%, the probabilities associated with VPIN percentiles less than 0.5% is zero with the probabilities shifted to higher percentiles of VPINs. This pattern holds for both crude oil and natural gas futures. Second, for absolute value of returns in the bins 2.0% and $\geq 2.00\%$, the probability associated with $\text{VPIN}_{\lambda-1}$ in the 1.00 percentile are 36% and 30.8% while the corresponding probabilities for $\text{EXPS\_VPIN}_{\alpha=0.1}$ are 72.% and 69.2% for crude oil futures. Analogous patterns hold for natural gas futures. Accordingly, we provide evidence that $\text{EXPS\_VPIN}_{\alpha=0.1}$ has stronger temporal linkages with subsequent absolute value of returns compared to $\text{VPIN}_{\lambda-1}$.

[Insert Tables 12 and 13]

Finally, we examine the predicative power of $\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}$ and $\text{VPIN}_{\lambda-1}$ for short-term volatility ($|r_\lambda|$) and bid-ask spreads (BAS_λ) in a regression framework with and without the control variables: number of trades ($\text{NT}_{\lambda-1}$), trade size ($\text{TS}_{\lambda-1}$) and realized volatility ($\text{RV}_{\lambda-1}$)).

Table 14 reports six regressions of BAS_λ on $\text{VPIN}_{\lambda-1}$, $\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}$ and the control variables for crude oil (Panel A) and natural gas (Panel B) futures. We find that both $\text{VPIN}_{\lambda-1}$ and $\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}$ have predictive power on subsequent bid-ask spreads with and without control variables for natural gas futures, but the coefficients of $\text{VPIN}_{\lambda-1}$ and $\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}$ in equation I and II are incorrect and statistically significant for crude oil futures. However, the coefficients of these two VPINs become positive and significant with control of other variables in regressions for crude oil futures.

Table 15 reports six regressions of $|r_\lambda|$ on $\text{VPIN}_{\lambda-1}$, $\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}$ and the control variables for crude oil (Panel A) and natural gas (Panel B) futures. All regression coefficients are consistent with prior expectations and statistically significant at 1% at least. These regression results provide additional evidence that both $\text{VPIN}_{\lambda-1}$and $\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}$ have predictive power for $|r_\lambda|$ within the regression framework, Furthermore, we observe $\text{EXPS\_VPIN}_{\alpha=0.1,\lambda-1}$ have stronger predicate power than $\text{VPIN}_{\lambda-1}$ in terms of the magnitude of regression coefficients.

[Insert Tables 14 and 15]

5 Summary and Conclusions

This paper assesses the performance of VPIN versus a variant of VPIN, ($\text{EXPS\_VPIN}_{\alpha=0.1}$) around inventory announcements with price jumps and price jumps not associated with inventory announcements in crude oil and natural gas futures markets. Our sample period spans from January 1, 2009 to May 31, 2015. We believe that over six years of intraday sample data provide reliable and robust empirical results rather than relying on single day or a short time period as previous evaluations of the properties of VPIN have done. We obtain several interesting empirical results.

First, we document that VPIN has increased significantly around the inventory announcements period with price jumps (scheduled events) and jumps without inventory announcements (unscheduled events). These results suggest that order flow gains more toxicity during the release of events and periods with price jumps.

Second, we find that VPIN did not reach local maxima prior to the events but rather after the occurrences of the events, which is consistent with previous findings by Andersen and Bondarenko (2014, 2015).

Third, in general, the values of VPIN in natural gas futures are higher than observed in crude oil futures market during our sample period. These results are consistent with previous findings by Bjursell et al. (2015) who observed that the volatility in natural gas futures markets is higher than in crude oil futures markets.

Fourth, we demonstrate that a minor variant of VPIN with exponential smoothing (EXPS_VPIN$_{\alpha=0.1}$) significantly improves the early warning signals property of the VPIN metric and returns faster than VPIN to normal levels after the event time.

Fifth, following the model free approach used by ELO (2013), we find EXPS_VPIN outperform VPIN in temporal linkages between toxicity and short term volatility in terms of alternative correlation measures and a conditional probability framework.

In summary, the contribution of this paper is to provide large scale empirical evidence to an ongoing debate on the predictive power of VPIN and futures liquidity induced volatility. We find that VPIN is a useful tool to signal periods of turbulent price behaviors during news releases and price jumps, but VPIN does not demonstrate an early warning signal property in crude oil and natural gas futures. However, a minor variant of VPIN (EXPS_VPIN$_{\alpha=0.1}$) proposed in this work, significantly improves the early warning signals of toxicity. Furthermore, it returns faster to the normal pre-event levels following the turbulent period.

References

Abad, D. and Yagüe, J. (2012). From PIN to VPIN: An introduction to order flow toxicity. *The Spanish Review of Financial Economics*, 10(2):74-83.

Andersen, T. G. and Bondarenko, O. (2014). VPIN and flash crash. *Journal of Financial Markets*, 17:1-46.

Andersen, T. G. and Bondarenko, O. (2015). Assessing measures of order ow toxicity and early warning signals for market turbulence. *Review of Finance*, 19:1-54.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2:1-48.

Barndorff-Nielsen, O. E. and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4(1):1-30.

Bjursell, J., Gentle, J. E., and Wang, G. H. (2015). Inventory announcements, jump dynamics, volatility and trading volume in U.S. energy futures markets. *Energy Economics*, 48(1):50-72.

Brennan, M. J. (1958). The supply of storage. *American Economic Review*, 48(1):50-72.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-frequency trading and price discovery. *Review of Financial Studies*, 27(8):2267-2306.

Brown, R. G. (1962). *Smoothing, Forecasting and Prediction of Discrete Time Series*. Prentice-Hall, New York.

Chan, W. H., Wang, G. H. K., and Yang, L. (2010). Weather, inventory, and common jump dynamics in natural gas futures and spot markets. *Review of Futures Markets*, 18(4):364-384.

Chatfeld, C., Koehler, A., Ord, J., and Ord, R. (2001). A new look at models for exponential smoothing. *The Statistician*, 50:147-159.

Cheung, W. M., Chou, R. K., and Lei, A. C. (2015). Exchange-traded barrier option and VPIN: Evidence from Hong Kong. *Journal of Futures Markets*, 35(6):561-581.

Diebold, F. (2007). *Elements of Forecasting*. Thomson-South-Western, 4 edition.

Easley, D., de Prado, M. L., and O'Hara, M. (2011). The microstructure of the flash crash flow toxicity, liquidity crashes and the probability of informed trading. *Journal of Portfolio Management*, 37:118-128.

Easley, D., de Prado, M. L., and O' Hara, M. (2014). VPIN and ash crash: A rejoinder. *Journal of Financial Markets*, 17:47-52.

Easley, D., de Prado, M. M. L., and O'Hara, M. (2012). Flow toxicity and liquidity in a high-frequency world. *Review of Financial Studies*, 25:1457-1493.

Gay, G. D., Simkins, B., and Turac, M. (2009). Analyst forecasts and price discovery in futures markets: The case of natural gas storage. *Journal of Futures Markets*, 29(5):451-477.

Hasbrouck, J. and Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, 16:646-679.

Hendershott, T., Jones, C. M., and Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1):1-33.

Jiang, G. J., Lo, I., and Verdelhan, A. (2011). Information shocks and bond price jumps: Evidence from the u.s. treasury market. *Journal of Financial and Quantitative Analysis*, 146(2):527-551.

Kaldor, N. (1939). Welfare propositions in economics and interpersonal comparisons of utility. *Economic Journal*, 49(145):549-552.

Kirilenko, A. A., Kyle, A. S., Samadi, M., and Tuzun, T. (2015). The flash crash: The impact of high frequency trading on an electronic market. SSRN Manuscript, www.ssrn.com/abstract=1686004.

Lee, S. S. and Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies*, 21(6):2535-2563.

Linn, S. C. and Zhu, Z. (2004). Natural gas prices and the gas storage report: Public news and volatility in energy futures markets. *Journal of Futures Markets*, 24(3):283-313.

Mu, X. (2007). Weather, storage, and natural gas price dynamics: Fundamentals and volatility. *Energy Economics*, 29(1):46-63.

Snedecor, G. W and Cochran, W. G (1980), Statistical Methods, 7th edition. The Iowa State University Press, Ames, Iowa.

Telser, L. G. (1958). Futures trading and the storage of cotton and wheat. *Journal of Political Economy*, 66(3):233-255.

Wei, W. C., Gerace, D., and Frino, A. (2013). Informed trading, ow toxicity and the impact on intraday trading factors. *Australasian Accounting Business and Finance Journal*, 7(2):3-24.

Working, H. (1948). The theory of inverse carrying charge in futures markets. *Journal of Farm Economics*, 30:1-28.

Working, H. (1949). The theory of the price of storage. *American Economic Review*, 39:1254-1262.

Wu, K., Bethel, W., Gu, M., Leinweber, D., and Ruebel, O. (2013). A big data approach to analyzing market volatility. *Algorithmic Finance*, 2:241-267.

Yildiz, S., Ness, R. A. V., and Ness, B. F. V. (2013). Analysis determinants of VPIN, HFTs order ow toxicity and impact on stock price variance. Presented at 2013 FMA annual meetings. October. Chicago, Illinois.

Tables

Table 1: Yearly summary per commodity. No. Days denotes the number of days with trade data. Total presents the number of significant jumps with the number of positive and negative jumps in the two following columns. The time series for 2015 ends at May 31.

		Crude Oil - No. Jumps			Natural Gas - No. Jumps		
Year	No. Days	Total	Positive	Negative	Total	Positive	Negative
2009	258	10	7	3	34	17	17
2010	258	22	9	13	41	14	27
2011	258	18	8	10	55	24	31
2012	258	21	7	14	41	18	23
2013	258	8	2	6	46	23	23
2014	258	19	9	10	44	21	23
201505	101	7	7	0	13	5	8

Table 2: Summary statistics for `VPIN` and $\texttt{EXPS\_VPIN}_{\alpha=0.10}$ for crude oil and natural gas futures. The third ($\texttt{EXPS\_VPIN}_{\alpha=0.10,a}$) and fourth columns ($\texttt{EXPS\_VPIN}_{\alpha=0.10,b}$) per commodity are results for subsets of the sample data using the exponential smoothing approach. $\texttt{EXPS\_VPIN}_{\alpha=0.10,a}$ is based on data from January 2009 to December 2011, and $\texttt{EXPS\_VPIN}_{\alpha=0.10,b}$ is based on data from January 2012 to May 2015. Kurtosis denotes estimates of the excess kurtosis. AR(1) is the auto correlation for lag 1 for the `VPIN` time series. ADF is the augmented Dickey-Fuller test statistic. No. Obs denotes the number of observations. The values labelled CDF are the percentiles based on the ECDF of the `VPIN` values.

	Crude Oil				Natural Gas			
	VPIN	EXPS_VPIN	EXPS_VPINa	EXPS_VPINb	VPIN	EXPS_VPIN	EXPS_VPINa	EXPS_VPINb
Mean	0.12	0.12	0.13	0.11	0.30	0.30	0.30	0.29
Std Dev	0.04	0.06	0.06	0.05	0.09	0.12	0.10	0.12
Skew	1.59	1.93	1.69	2.30	0.84	1.07	0.94	1.14
Kurtosis	4.34	6.81	4.95	9.93	0.99	1.37	1.13	1.38
AR(1)	0.996	0.972	0.970	0.973	0.996	0.977	0.969	0.980
ADF	−26.00	−27.28	−20.99	−22.74	−29.39	−29.73	−22.57	−24.58
No. Obs	82424	82424	37850	44574	82435	82435	31635	50800
CDF(0.1)	0.08	0.07	0.08	0.06	0.19	0.17	0.19	0.16
CDF(0.25)	0.09	0.09	0.10	0.08	0.23	0.21	0.23	0.20
CDF(0.5)	0.12	0.11	0.12	0.10	0.28	0.27	0.29	0.26
CDF(0.75)	0.14	0.15	0.16	0.13	0.34	0.35	0.36	0.35
CDF(0.9)	0.18	0.19	0.21	0.18	0.41	0.45	0.44	0.47

Table 3: Statistics for the futures contracts for the full sample period (2009-2015) and the subperiods 2009-2011 and 2012-2015. Avg Vol denotes the average daily trading volume. VBS denotes the volume bucket size per period and commodity.

	Crude Oil			Natural Gas		
	2009-2015	2009-2011	2012-2015	2009-2015	2009-2011	2012-2015
Avg Vol	153144	149935	155982	67929	55592	78842
VBS	3063	2999	3120	1359	1112	1577

Table 4: The table lists intraday time series of returns, VPIN and $\texttt{EXPS\_VPIN}_{\alpha=0.10}$ for crude oil on May 5, 2011. The first column lists the closing timestamps in hours and minutes of the VPIN observations. The second column presents the intraday returns, $\log(p_{t_i}/p_{t_{i-1}})$. The two following columns are VPIN and $\texttt{EXPS\_VPIN}_{\alpha=0.10}$. The last two columns are the ECDF(VPIN) and ECDF($\texttt{EXPS\_VPIN}_{\alpha=0.10}$). The VPIN and $\texttt{EXPS\_VPIN}_{\alpha=0.10}$ calculations are based on one-minute time bars and averaged over a window with 50 observations.

				ECDF	
	Return	VPIN	EXPS_VPIN	VPIN	EXPS_VPIN
09:01	0.000	0.15	0.08	0.77	0.22
09:02	−0.000	0.14	0.11	0.73	0.48
09:03	−0.001	0.13	0.11	0.67	0.47
09:06	−0.003	0.13	0.11	0.67	0.50
09:07	−0.003	0.14	0.16	0.75	0.82
09:08	−0.002	0.13	0.15	0.65	0.75
09:11	−0.002	0.13	0.14	0.62	0.70
09:16	−0.001	0.12	0.12	0.58	0.61
09:18	0.001	0.13	0.15	0.64	0.78
09:21	0.001	0.13	0.14	0.63	0.70
09:26	0.002	0.13	0.13	0.62	0.67
09:32	0.001	0.13	0.12	0.61	0.59
09:35	−0.000	0.13	0.13	0.63	0.62
09:38	−0.001	0.13	0.12	0.64	0.61
09:45	0.000	0.12	0.12	0.58	0.59
09:50	0.000	0.12	0.11	0.56	0.49
09:55	−0.001	0.12	0.11	0.54	0.46
09:58	−0.003	0.12	0.11	0.56	0.50
10:01	−0.001	0.12	0.12	0.55	0.57
10:05	−0.000	0.12	0.12	0.53	0.55
10:11	−0.001	0.12	0.11	0.51	0.46
10:15	−0.004	0.12	0.12	0.55	0.55
10:17	−0.004	0.12	0.13	0.59	0.62
10:19	−0.004	0.12	0.12	0.56	0.60
10:21	−0.006	0.12	0.14	0.59	0.69
10:25	−0.004	0.13	0.14	0.62	0.74
10:30	−0.007	0.12	0.15	0.60	0.74
10:31	−0.006	0.13	0.16	0.63	0.80
10:35	−0.005	0.13	0.15	0.64	0.77
10:37	−0.005	0.13	0.16	0.67	0.81
10:41	−0.006	0.12	0.16	0.57	0.80
10:45	−0.006	0.12	0.15	0.59	0.77
10:50	−0.008	0.12	0.14	0.54	0.70
10:51	−0.009	0.13	0.17	0.62	0.83
10:52	−0.010	0.13	0.18	0.66	0.88
10:53	−0.011	0.13	0.19	0.68	0.90
10:54	−0.014	0.15	0.25	0.78	0.96
10:55	−0.018	0.16	0.31	0.86	0.99
10:56	−0.016	0.17	0.33	0.89	0.99
10:56	−0.016	0.18	0.35	0.91	0.99
10:57	−0.016	0.19	0.34	0.92	0.99
10:59	−0.015	0.19	0.32	0.92	0.99
11:01	−0.017	0.19	0.32	0.93	0.99
11:03	−0.019	0.20	0.31	0.94	0.99
11:05	−0.025	0.21	0.35	0.96	0.99
11:05	−0.025	0.23	0.42	0.97	1.00
11:05	−0.025	0.25	0.47	0.98	1.00
11:06	−0.025	0.25	0.44	0.98	1.00
11:07	−0.024	0.25	0.41	0.98	1.00
11:09	−0.025	0.25	0.38	0.98	1.00

Table 4 continue

				ECDF	
	Return	VPIN	EXPS_VPIN	VPIN	EXPS_VPIN
11:10	−0.025	0.25	0.34	0.98	0.99
11:11	−0.027	0.25	0.37	0.98	0.99
11:12	−0.033	0.27	0.41	0.99	1.00
11:13	−0.041	0.28	0.47	0.99	1.00
11:13	−0.041	0.29	0.52	0.99	1.00
11:13	−0.041	0.31	0.57	1.00	1.00
11:14	−0.039	0.32	0.56	1.00	1.00
11:14	−0.039	0.34	0.58	1.00	1.00
11:15	−0.035	0.34	0.61	1.00	1.00
11:16	−0.035	0.35	0.58	1.00	1.00
11:17	−0.033	0.36	0.57	1.00	1.00
11:19	−0.030	0.37	0.57	1.00	1.00
11:20	−0.028	0.37	0.54	1.00	1.00
11:22	−0.028	0.37	0.50	1.00	1.00
11:24	−0.027	0.37	0.46	1.00	1.00
11:25	−0.022	0.39	0.50	1.00	1.00
11:26	−0.024	0.39	0.46	1.00	1.00
11:28	−0.028	0.39	0.45	1.00	1.00
11:30	−0.027	0.39	0.42	1.00	1.00
11:33	−0.030	0.40	0.40	1.00	1.00
11:37	−0.026	0.40	0.39	1.00	1.00
11:40	−0.026	0.40	0.36	1.00	0.99
11:43	−0.025	0.40	0.36	1.00	0.99
11:47	−0.020	0.40	0.33	1.00	0.99
11:48	−0.019	0.41	0.36	1.00	0.99
11:49	−0.022	0.41	0.34	1.00	0.99
11:52	−0.023	0.41	0.32	1.00	0.99
11:55	−0.024	0.41	0.30	1.00	0.99
11:58	−0.029	0.41	0.32	1.00	0.99
11:59	−0.030	0.42	0.33	1.00	0.99
12:01	−0.026	0.42	0.34	1.00	0.99
12:04	−0.028	0.42	0.30	1.00	0.99
12:09	−0.028	0.42	0.28	1.00	0.98
12:13	−0.029	0.42	0.27	1.00	0.98
12:19	−0.031	0.41	0.26	1.00	0.97
12:23	−0.032	0.41	0.24	1.00	0.96
12:28	−0.028	0.40	0.23	1.00	0.95
12:35	−0.031	0.38	0.22	1.00	0.94
12:38	−0.031	0.37	0.21	1.00	0.93
12:43	−0.031	0.37	0.21	1.00	0.93
12:49	−0.031	0.36	0.19	1.00	0.90
12:53	−0.034	0.36	0.19	1.00	0.90
12:56	−0.038	0.37	0.24	1.00	0.96
12:58	−0.037	0.37	0.23	1.00	0.95
13:00	−0.037	0.36	0.22	1.00	0.94
13:04	−0.033	0.34	0.23	1.00	0.95
13:09	−0.034	0.33	0.22	1.00	0.94
13:13	−0.030	0.33	0.23	1.00	0.95
13:20	−0.031	0.33	0.21	1.00	0.93
13:27	−0.031	0.33	0.21	1.00	0.92

Table 4 continue

	Return	VPIN	EXPS_VPIN	ECDF VPIN	ECDF EXPS_VPIN
13:27	−0.031	0.33	0.21	1.00	0.92
13:36	−0.036	0.33	0.19	1.00	0.90
13:39	−0.037	0.32	0.20	1.00	0.91
13:42	−0.038	0.31	0.20	1.00	0.91
13:45	−0.038	0.29	0.19	0.99	0.89
13:49	−0.038	0.27	0.17	0.99	0.86
13:51	−0.041	0.26	0.21	0.99	0.93
13:53	−0.043	0.27	0.25	0.99	0.96
13:53	−0.043	0.25	0.23	0.98	0.95
13:56	−0.046	0.24	0.22	0.98	0.94
13:57	−0.048	0.25	0.27	0.98	0.97
13:58	−0.046	0.24	0.25	0.98	0.97
14:00	−0.045	0.23	0.25	0.97	0.96
14:02	−0.044	0.23	0.24	0.97	0.96
14:04	−0.048	0.24	0.26	0.97	0.97
14:05	−0.051	0.25	0.32	0.98	0.99
14:07	−0.053	0.24	0.31	0.98	0.99
14:08	−0.051	0.25	0.33	0.98	0.99
14:09	−0.052	0.24	0.33	0.98	0.99
14:11	−0.057	0.25	0.35	0.98	0.99
14:11	−0.057	0.26	0.38	0.99	1.00
14:12	−0.053	0.26	0.37	0.99	0.99
14:13	−0.055	0.26	0.35	0.99	0.99
14:14	−0.054	0.26	0.34	0.99	0.99
14:16	−0.053	0.26	0.33	0.99	0.99
14:18	−0.054	0.26	0.33	0.98	0.99
14:20	−0.051	0.26	0.33	0.99	0.99
14:22	−0.050	0.26	0.33	0.99	0.99
14:24	−0.052	0.26	0.30	0.99	0.99
14:25	−0.054	0.27	0.33	0.99	0.99
14:27	−0.057	0.27	0.36	0.99	0.99
14:28	−0.059	0.27	0.36	0.99	0.99
14:29	−0.057	0.27	0.34	0.99	0.99
14:29	−0.057	0.28	0.36	0.99	0.99
14:30	−0.058	0.28	0.34	0.99	0.99

Table 5: The table presents changes in VPIN surrounding the release of the inventory reports for crude oil and natural gas. Since VPIN estimates are asynchronous, they are indexed in order relative to the timing of the inventory release where index -1 and 1 denote the first observations immediately before and after the release, respectively. Only events with a negative and significant jump and a surprise greater than one standard deviation are included in the regression where 20 VPIN estimates prior to the inventory release and 60 observations following the inventory are considered. The estimates are obtained via OLS of the regression equation,

$$\mathrm{V}_{t,k} = \beta_0 + \sum_{i \neq l} \beta_i \mathrm{D}_i + \epsilon_{t,k},$$

where D_i is a dummy variable for the ith VPIN estimates. There is no dummy variable D_0. The regression table reports estimates for the dummy variables $D_{-10} - D_{14}$ to preserve space. Regression I includes results for VPIN; Regression II and III are results for $\texttt{EXPS\_VPIN}_{\alpha=0.05}$ and $\texttt{EXPS\_VPIN}_{\alpha=0.10}$ respectively. The VPIN calculation is based on one-minute time bars. All VPIN calculations are based on a moving window with 50 VPIN observations.

	Crude Oil			Natural Gas		
	I	II	III	I	II	III
Intercept	0.10	0.09	0.09	0.25	0.25	0.24
	(7.64)	(6.70)	(4.51)	(19.36)	(16.88)	(12.85)
D_{-10}	-0.004	0.004	0.008	0.001	0.01	0.02
	(-0.23)	(0.22)	(0.28)	(0.06)	(0.50)	(0.70)
D_{-9}	-0.004	0.004	0.007	$4e-04$	0.01	0.02
	(-0.23)	(0.20)	(0.25)	(0.02)	(0.49)	(0.65)
D_{-8}	-0.002	0.006	0.01	$-2e-04$	0.01	0.02
	(-0.11)	(0.30)	(0.40)	(-0.01)	(0.51)	(0.67)
D_{-7}	-0.002	0.008	0.02	-0.003	0.005	0.007
	(-0.09)	(0.43)	(0.57)	(-0.17)	(0.26)	(0.27)
D_{-6}	-0.003	0.007	0.01	-0.006	0.001	-0.001
	(-0.15)	(0.36)	(0.46)	(-0.30)	(0.05)	(-0.04)
D_{-5}	$-3e-06$	0.01	0.03	-0.008	-0.004	-0.01
	$(-2e-04)$	(0.70)	(0.92)	(-0.43)	(-0.20)	(-0.41)
D_{-4}	$5e-05$	0.01	0.02	-0.01	-0.007	-0.02
	(0.003)	(0.61)	(0.77)	(-0.52)	(-0.36)	(-0.62)
D_{-3}	-0.001	0.008	0.01	-0.01	-0.01	-0.03
	(-0.07)	(0.39)	(0.44)	(-0.61)	(-0.62)	(-0.98)
D_{-2}	-0.002	0.004	0.004	-0.01	-0.02	-0.03
	(-0.12)	(0.19)	(0.15)	(-0.69)	(-0.78)	(-1.16)
D_{-1}	0.003	0.02	0.03	-0.009	-0.007	-0.01
	(0.18)	(0.96)	(1.25)	(-0.49)	(-0.34)	(-0.42)
D_1	0.02	0.05	0.10	0.006	0.03	0.06
	(0.98)	(2.71)	(3.65)	(0.34)	(1.37)	(2.26)
D_2	0.03	0.08	0.16	0.02	0.06	0.12
	(1.61)	(4.28)	(5.69)	(1.05)	(2.96)	(4.59)
D_3	0.04	0.11	0.20	0.03	0.09	0.18
	(2.27)	(5.51)	(7.14)	(1.77)	(4.37)	(6.56)
D_4	0.05	0.12	0.21	0.04	0.12	0.22
	(2.68)	(6.13)	(7.70)	(2.42)	(5.54)	(8.05)
D_5	0.05	0.12	0.20	0.06	0.13	0.24
	(2.77)	(5.96)	(7.13)	(3.01)	(6.43)	(9.06)
D_6	0.05	0.12	0.19	0.06	0.15	0.27
	(2.90)	(5.98)	(6.88)	(3.52)	(7.20)	(9.84)
D_7	0.05	0.12	0.18	0.07	0.17	0.29
	(2.91)	(5.91)	(6.53)	(4.03)	(8.02)	(10.70)
D_8	0.05	0.12	0.18	0.08	0.18	0.30
	(2.95)	(6.14)	(6.62)	(4.43)	(8.58)	(11.12)
D_9	0.06	0.12	0.18	0.09	0.19	0.31
	(3.13)	(6.23)	(6.53)	(4.80)	(9.14)	(11.54)
D_{10}	0.06	0.12	0.17	0.09	0.19	0.31
	(3.28)	(6.18)	(6.25)	(5.11)	(9.32)	(11.39)
D_{11}	0.06	0.12	0.16	0.10	0.20	0.30
	(3.36)	(5.99)	(5.81)	(5.37)	(9.54)	(11.32)
D_{12}	0.06	0.11	0.15	0.10	0.20	0.30
	(3.45)	(5.86)	(5.47)	(5.67)	(9.68)	(11.15)
D_{13}	0.06	0.11	0.15	0.11	0.20	0.29
	(3.57)	(5.86)	(5.35)	(5.86)	(9.65)	(10.74)
D_{14}	0.07	0.11	0.14	0.11	0.20	0.27
	(3.63)	(5.65)	(4.95)	(6.04)	(9.51)	(10.21)
R^2 Adj	0.56	0.66	0.70	0.38	0.37	0.43
F Stat	6.08	8.82	10.32	21.29	20.29	25.78

Table 6: The table presents changes in `VPIN` at jumps which are not associated with the inventory reports for crude oil and natural gas. Since `VPIN` estimates are asynchronous, they are indexed in order relative to the timing of the inventory release where index -1 and 1 denote the first observations immediately before and after the release, respectively. Only events with a negative and significant jump and a surprise greater than one standard deviation are included in the regression where 20 `VPIN` estimates prior to the inventory release and 60 observations following the inventory are considered. The estimates are obtained via OLS of the regression equation,

$$\mathrm{V}_{t,k} = \beta_0 + \sum_{i \neq l} \beta_i \mathrm{D}_i + \epsilon_{t,k},$$

where D_i is a dummy variable for the ith `VPIN` estimates. There is no dummy variable D_0. The regression table reports estimates for the dummy variables $D_{-10} - D_{14}$ to preserve space. Regression I includes results for `VPIN`; Regression II and III are results for $\texttt{EXPS\_VPIN}_{\alpha=0.05}$ and $\texttt{EXPS\_VPIN}_{\alpha=0.10}$ respectively. The `VPIN` calculation is based on one-minute time bars. All `VPIN` calculations are based on a moving window with 50 `VPIN` observations.

	Crude Oil			Natural Gas		
	I	II	III	I	II	III
Intercept	0.11	0.11	0.11	0.26	0.25	0.25
	(9.43)	(8.98)	(7.13)	(17.77)	(17.28)	(14.15)
D_{-10}	0.007	0.01	0.02	0.005	0.003	0.006
	(0.43)	(0.75)	(1.17)	(0.24)	(0.13)	(0.23)
D_{-9}	0.008	0.01	0.03	0.004	0.004	0.008
	(0.50)	(0.81)	(1.22)	(0.22)	(0.17)	(0.31)
D_{-8}	0.008	0.01	0.03	0.005	0.005	0.01
	(0.52)	(0.81)	(1.19)	(0.26)	(0.24)	(0.42)
D_{-7}	0.01	0.02	0.03	0.006	0.008	0.02
	(0.60)	(0.95)	(1.37)	(0.30)	(0.37)	(0.62)
D_{-6}	0.01	0.02	0.04	0.006	0.009	0.02
	(0.75)	(1.18)	(1.71)	(0.27)	(0.45)	(0.74)
D_{-5}	0.01	0.02	0.04	0.008	0.01	0.03
	(0.85)	(1.29)	(1.84)	(0.38)	(0.68)	(1.10)
D_{-4}	0.02	0.02	0.04	0.007	0.01	0.03
	(0.92)	(1.35)	(1.88)	(0.33)	(0.66)	(1.04)
D_{-3}	0.02	0.02	0.04	0.008	0.02	0.03
	(0.96)	(1.41)	(1.92)	(0.38)	(0.86)	(1.32)
D_{-2}	0.02	0.02	0.04	0.006	0.02	0.03
	(0.95)	(1.33)	(1.73)	(0.28)	(0.75)	(1.10)
D_{-1}	0.02	0.03	0.05	0.007	0.02	0.04
	(1.15)	(1.70)	(2.28)	(0.36)	(0.99)	(1.46)
D_1	0.03	0.05	0.09	0.01	0.04	0.07
	(1.69)	(2.95)	(4.27)	(0.69)	(1.89)	(2.91)
D_2	0.04	0.07	0.13	0.02	0.06	0.11
	(2.27)	(4.15)	(6.08)	(1.02)	(2.78)	(4.27)
D_3	0.04	0.08	0.15	0.02	0.07	0.12
	(2.66)	(4.83)	(6.94)	(1.18)	(3.23)	(4.83)
D_4	0.05	0.09	0.16	0.03	0.08	0.14
	(2.96)	(5.31)	(7.45)	(1.50)	(3.93)	(5.80)
D_5	0.05	0.10	0.16	0.04	0.09	0.16
	(3.21)	(5.60)	(7.63)	(1.81)	(4.49)	(6.49)
D_6	0.06	0.10	0.16	0.04	0.10	0.17
	(3.37)	(5.68)	(7.47)	(2.04)	(4.88)	(6.87)
D_7	0.06	0.10	0.15	0.05	0.11	0.18
	(3.50)	(5.76)	(7.35)	(2.24)	(5.19)	(7.13)
D_8	0.06	0.10	0.15	0.05	0.11	0.17
	(3.65)	(5.88)	(7.29)	(2.36)	(5.26)	(6.98)
D_9	0.06	0.10	0.15	0.05	0.11	0.17
	(3.75)	(5.95)	(7.18)	(2.47)	(5.22)	(6.66)
D_{10}	0.06	0.10	0.15	0.05	0.11	0.17
	(3.87)	(5.98)	(7.03)	(2.60)	(5.34)	(6.64)
D_{11}	0.06	0.10	0.14	0.06	0.11	0.17
	(3.94)	(5.90)	(6.69)	(2.75)	(5.48)	(6.67)
D_{12}	0.07	0.10	0.14	0.06	0.11	0.16
	(4.08)	(5.95)	(6.61)	(2.84)	(5.51)	(6.52)
D_{13}	0.07	0.10	0.14	0.06	0.12	0.16
	(4.21)	(5.94)	(6.43)	(2.98)	(5.58)	(6.45)
D_{14}	0.07	0.10	0.13	0.06	0.11	0.15
	(4.30)	(5.85)	(6.14)	(3.07)	(5.51)	(6.18)
R^2 Adj	0.14	0.12	0.15	0.08	0.10	0.15
F Stat	8.32	7.42	9.51	5.03	6.53	9.37

Table 7: The table presents changes in the empirical cumulative density function of the VPIN surrounding the release of the inventory reports for crude oil and natural gas. Since VPIN estimates are asynchronous, they are indexed in order relative to the timing of the inventory release where index −1 and 1 denote the first observations immediately before and after the release, respectively. Only events with a negative and significant jump and a surprise greater than one standard deviation are included in the regression where 20 VPIN estimates prior to the inventory release and 60 observations following the inventory are considered. The estimates are obtained via OLS of the regression equation,

$$\mathrm{V}_{t,k} = \beta_0 + \sum_{i \neq l} \beta_i \mathrm{D}_i + \epsilon_{t,k},$$

where D_i is a dummy variable for the ith ECDF(VPIN) estimates. There is no dummy variable D_0. The regression table reports estimates for the dummy variables D_{-10}–D_{14} to preserve space. Regression I includes results for VPIN; Regression II and III are results for $\mathrm{EXPS\_VPIN}_{\alpha=0.05}$ and $\mathrm{EXPS\_VPIN}_{\alpha=0.10}$ respectively. The VPIN calculation is based on one-minute time bars. All VPIN calculations are based on a moving window with 50 VPIN observations.

	Crude Oil			Natural Gas		
	I	II	III	I	II	III
Intercept	0.31	0.26	0.29	0.36	0.36	0.38
	(3.80)	(3.46)	(3.37)	(10.35)	(9.61)	(9.55)
D_{-10}	−0.05	0.04	0.07	0.004	0.03	0.06
	(−0.46)	(0.42)	(0.57)	(0.09)	(0.61)	(1.02)
D_{-9}	−0.06	0.04	0.06	$9e-04$	0.03	0.06
	(−0.49)	(0.35)	(0.50)	(0.02)	(0.59)	(0.98)
D_{-8}	−0.04	0.06	0.10	$-3e-04$	0.03	0.06
	(−0.32)	(0.54)	(0.84)	(−0.007)	(0.61)	(0.98)
D_{-7}	−0.04	0.08	0.15	−0.01	0.01	0.02
	(−0.31)	(0.80)	(1.22)	(−0.20)	(0.27)	(0.41)
D_{-6}	−0.05	0.07	0.12	−0.02	−0.003	−0.003
	(−0.41)	(0.65)	(0.98)	(−0.40)	(−0.06)	(−0.06)
D_{-5}	−0.02	0.14	0.23	−0.03	−0.02	−0.04
	(−0.17)	(1.35)	(1.89)	(−0.58)	(−0.47)	(−0.66)
D_{-4}	−0.02	0.12	0.19	−0.03	−0.04	−0.06
	(−0.16)	(1.18)	(1.61)	(−0.66)	(−0.72)	(−1.02)
D_{-3}	−0.03	0.07	0.11	−0.04	−0.06	−0.09
	(−0.27)	(0.71)	(0.91)	(−0.82)	(−1.17)	(−1.63)
D_{-2}	−0.04	0.03	0.04	−0.04	−0.08	−0.11
	(−0.34)	(0.28)	(0.29)	(−0.91)	(−1.41)	(−1.93)
D_{-1}	0.02	0.19	0.27	−0.03	−0.04	−0.04
	(0.20)	(1.84)	(2.23)	(−0.66)	(−0.76)	(−0.72)
D_1	0.18	0.45	0.53	0.03	0.11	0.21
	(1.56)	(4.25)	(4.38)	(0.56)	(2.02)	(3.74)
D_2	0.28	0.55	0.61	0.08	0.24	0.38
	(2.44)	(5.24)	(5.03)	(1.63)	(4.55)	(6.71)
D_3	0.37	0.62	0.66	0.13	0.35	0.47
	(3.24)	(5.94)	(5.49)	(2.73)	(6.49)	(8.27)
D_4	0.43	0.68	0.69	0.18	0.41	0.51
	(3.77)	(6.43)	(5.71)	(3.70)	(7.70)	(8.95)
D_5	0.45	0.68	0.69	0.22	0.45	0.53
	(3.96)	(6.51)	(5.72)	(4.55)	(8.45)	(9.29)
D_6	0.47	0.69	0.69	0.26	0.48	0.54
	(4.11)	(6.53)	(5.70)	(5.27)	(8.94)	(9.47)
D_7	0.47	0.68	0.68	0.29	0.50	0.55
	(4.09)	(6.50)	(5.66)	(5.99)	(9.43)	(9.67)
D_8	0.46	0.68	0.68	0.32	0.52	0.55
	(4.07)	(6.49)	(5.62)	(6.45)	(9.71)	(9.78)
D_9	0.47	0.67	0.67	0.34	0.53	0.56
	(4.11)	(6.41)	(5.53)	(6.91)	(9.96)	(9.87)
D_{10}	0.47	0.66	0.65	0.36	0.53	0.55
	(4.14)	(6.31)	(5.42)	(7.25)	(9.99)	(9.78)
D_{11}	0.48	0.66	0.64	0.37	0.53	0.55
	(4.19)	(6.25)	(5.32)	(7.48)	(9.97)	(9.64)
D_{12}	0.48	0.65	0.63	0.38	0.53	0.54
	(4.24)	(6.19)	(5.24)	(7.81)	(9.98)	(9.53)
D_{13}	0.49	0.66	0.64	0.39	0.53	0.53
	(4.31)	(6.23)	(5.27)	(7.95)	(9.94)	(9.42)
D_{14}	0.50	0.65	0.62	0.40	0.52	0.52
	(4.36)	(6.15)	(5.15)	(8.12)	(9.84)	(9.23)
R^2 Adj	0.65	0.68	0.62	0.46	0.42	0.40
F Stat	8.49	9.42	7.46	29.70	24.86	23.04

Table 8: The table presents changes in the empirical cumulative density function of the VPIN at jumps which are not associated with the inventory reports for crude oil and natural gas. Since VPIN estimates are asynchronous, they are indexed in order relative to the timing of the inventory release where index -1 and 1 denote the first observations immediately before and after the release, respectively. Only events with a negative and significant jump and a surprise greater than one standard deviation are included in the regression where 20 VPIN estimates prior to the inventory release and 60 observations following the inventory are considered. The estimates are obtained via OLS of the regression equation,

$$\mathrm{V}_{t,k} = \beta_0 + \sum_{i \neq l} \beta_i \mathrm{D}_i + \epsilon_{t,k},$$

where D_i is a dummy variable for the ith ECDF(VPIN) estimates. There is no dummy variable D_0. The regression table reports estimates for the dummy variables D_{-10}–D_{14} to preserve space. Regression I includes results for VPIN; Regression II and III are results for $\mathtt{EXPS\_VPIN}_{\alpha=0.05}$ and $\mathtt{EXPS\_VPIN}_{\alpha=0.10}$ respectively. The VPIN calculation is based on one-minute time bars. All VPIN calculations are based on a moving window with 50 VPIN observations.

	Crude Oil			Natural Gas		
	I	II	III	I	II	III
Intercept	0.40	0.42	0.41	0.36	0.35	0.37
	(9.04)	(9.28)	(9.19)	(7.82)	(7.79)	(8.52)
D_{-10}	0.05	0.04	0.07	0.004	−0.005	−0.003
	(0.74)	(0.65)	(1.14)	(0.05)	(−0.07)	(−0.05)
D_{-9}	0.05	0.04	0.07	0.001	−0.006	−0.003
	(0.87)	(0.67)	(1.16)	(0.02)	(−0.09)	(−0.04)
D_{-8}	0.05	0.04	0.07	$7e-04$	−0.007	−0.001
	(0.81)	(0.63)	(1.08)	(0.01)	(−0.11)	(−0.02)
D_{-7}	0.06	0.04	0.06	0.007	0.007	0.02
	(0.90)	(0.63)	(0.99)	(0.10)	(0.10)	(0.39)
D_{-6}	0.07	0.04	0.06	0.008	0.02	0.03
	(1.05)	(0.62)	(0.94)	(0.13)	(0.24)	(0.57)
D_{-5}	0.07	0.04	0.05	0.02	0.03	0.06
	(1.10)	(0.55)	(0.82)	(0.28)	(0.52)	(1.03)
D_{-4}	0.08	0.04	0.07	0.02	0.04	0.07
	(1.20)	(0.69)	(1.06)	(0.26)	(0.57)	(1.07)
D_{-3}	0.08	0.06	0.09	0.02	0.05	0.08
	(1.25)	(0.92)	(1.39)	(0.37)	(0.82)	(1.36)
D_{-2}	0.08	0.05	0.08	0.02	0.05	0.07
	(1.20)	(0.82)	(1.20)	(0.30)	(0.77)	(1.15)
D_{-1}	0.10	0.10	0.17	0.03	0.07	0.10
	(1.55)	(1.58)	(2.64)	(0.40)	(1.03)	(1.68)
D_1	0.15	0.23	0.34	0.05	0.13	0.20
	(2.47)	(3.67)	(5.40)	(0.77)	(2.02)	(3.30)
D_2	0.22	0.34	0.45	0.07	0.19	0.28
	(3.49)	(5.34)	(7.07)	(1.16)	(2.99)	(4.62)
D_3	0.25	0.37	0.46	0.09	0.22	0.31
	(3.98)	(5.78)	(7.29)	(1.35)	(3.48)	(5.09)
D_4	0.27	0.38	0.46	0.11	0.27	0.35
	(4.30)	(5.93)	(7.27)	(1.74)	(4.17)	(5.70)
D_5	0.28	0.38	0.45	0.13	0.29	0.36
	(4.52)	(5.89)	(7.05)	(2.07)	(4.51)	(5.89)
D_6	0.29	0.37	0.43	0.15	0.31	0.37
	(4.64)	(5.83)	(6.83)	(2.33)	(4.82)	(6.05)
D_7	0.29	0.37	0.42	0.17	0.32	0.37
	(4.67)	(5.75)	(6.67)	(2.56)	(5.04)	(6.16)
D_8	0.30	0.36	0.41	0.18	0.33	0.37
	(4.74)	(5.66)	(6.48)	(2.71)	(5.12)	(6.12)
D_9	0.29	0.36	0.40	0.19	0.33	0.37
	(4.66)	(5.58)	(6.31)	(2.86)	(5.13)	(6.01)
D_{10}	0.29	0.35	0.39	0.19	0.33	0.36
	(4.68)	(5.46)	(6.07)	(3.01)	(5.18)	(5.97)
D_{11}	0.29	0.34	0.38	0.20	0.34	0.36
	(4.67)	(5.41)	(5.93)	(3.15)	(5.30)	(5.99)
D_{12}	0.30	0.35	0.38	0.21	0.34	0.36
	(4.79)	(5.53)	(6.02)	(3.21)	(5.35)	(5.93)
D_{13}	0.31	0.36	0.40	0.21	0.34	0.35
	(4.90)	(5.69)	(6.24)	(3.30)	(5.28)	(5.72)
D_{14}	0.31	0.36	0.39	0.22	0.33	0.34
	(4.98)	(5.68)	(6.12)	(3.36)	(5.22)	(5.54)
R^2 Adj	0.13	0.12	0.14	0.10	0.11	0.12
F Stat	8.24	7.56	8.42	6.23	6.61	7.78

Table 9: Panel A presents intraday Pearson correlation for crude oil and natural gas. Panel B tabulates 95% confidence intervals of the correlation estimates. The intervals are obtained by Fisher Z-transformation. λ denotes the intraday timing for VPIN observations. The variables are absolute value of return ($|r|$), bid-ask spread obtained using the CFTC estimator (BAS), number of trades (NT), average trade size (TS), and realized volatility (RV).

Panel A: Pearson Correlation

	$\lvert r_\lambda \rvert$	BAS_λ	$\text{VPIN}_{\lambda-1}$	$\text{EXPS\_VPIN}_{\lambda-1}$	$\text{NT}_{\lambda-1}$	$\text{TS}_{\lambda-1}$	$\text{RV}_{\lambda-1}$
Crude Oil							
$\lvert r_\lambda \rvert$	1.00	−0.02	0.21	0.28	0.39	0.04	0.68
BAS_λ	−0.02	1.00	−0.07	−0.06	0.01	0.14	−0.00
$\text{VPIN}_{\lambda-1}$	0.21	−0.07	1.00	0.73	0.08	−0.05	0.23
$\text{EXPS\_VPIN}_{\lambda-1}$	0.28	−0.06	0.73	1.00	0.17	−0.03	0.23
$\text{NT}_{\lambda-1}$	0.39	0.01	0.08	0.17	1.00	−0.27	0.24
$\text{TS}_{\lambda-1}$	0.04	0.14	−0.05	−0.03	−0.27	1.00	0.03
$\text{RV}_{\lambda-1}$	0.68	−0.00	0.23	0.23	0.24	0.03	1.00
Natural Gas							
$\lvert r_\lambda \rvert$	1.00	0.19	0.07	0.19	0.78	−0.01	0.90
BAS_λ	0.19	1.00	0.13	0.15	0.05	0.17	0.31
$\text{VPIN}_{\lambda-1}$	0.07	0.13	1.00	0.72	0.05	−0.04	0.09
$\text{EXPS\_VPIN}_{\lambda-1}$	0.19	0.15	0.72	1.00	0.19	−0.03	0.19
$\text{NT}_{\lambda-1}$	0.78	0.05	0.05	0.19	1.00	−0.03	0.64
$\text{TS}_{\lambda-1}$	−0.01	0.17	−0.04	−0.03	−0.03	1.00	−0.04
$\text{RV}_{\lambda-1}$	0.90	0.31	0.09	0.19	0.64	−0.04	1.00

Panel B: Confidence Interval

	$\lvert r_\lambda \rvert$	BAS_λ	$\text{VPIN}_{\lambda-1}$	$\text{EXPS\_VPIN}_{\lambda-1}$	$\text{NT}_{\lambda-1}$	$\text{TS}_{\lambda-1}$	$\text{RV}_{\lambda-1}$
Crude Oil							
$\lvert r_\lambda \rvert$	1	[−0.03, −0.01]	[0.20, 0.22]	[0.27, 0.29]	[0.38, 0.39]	[0.03, 0.04]	[0.68, 0.68]
BAS_λ	[−0.03, −0.01]	1	[−0.08, −0.07]	[−0.07, −0.06]	[0.00, 0.02]	[0.13, 0.14]	[−0.01, 0.01]
$\text{VPIN}_{\lambda-1}$	[0.20, 0.22]	[−0.08, −0.07]	1	[0.73, 0.73]	[0.07, 0.08]	[−0.06, −0.05]	[0.23, 0.24]
$\text{EXPS\_VPIN}_{\lambda-1}$	[0.27, 0.29]	[−0.07, −0.06]	[0.73, 0.73]	1	[0.17, 0.18]	[−0.04, −0.03]	[0.22, 0.23]
$\text{NT}_{\lambda-1}$	[0.38, 0.39]	[0.00, 0.02]	[0.07, 0.08]	[0.17, 0.18]	1	[−0.28, −0.26]	[0.24, 0.25]
$\text{TS}_{\lambda-1}$	[0.03, 0.04]	[0.13, 0.14]	[−0.06, −0.05]	[−0.04, −0.03]	[−0.28, −0.26]	1	[0.03, 0.04]
$\text{RV}_{\lambda-1}$	[0.68, 0.68]	[−0.01, 0.01]	[0.23, 0.24]	[0.22, 0.23]	[0.24, 0.25]	[0.03, 0.04]	1
Natural Gas							
$\lvert r_\lambda \rvert$	1	[0.18, 0.19]	[0.06, 0.08]	[0.19, 0.20]	[0.78, 0.78]	[−0.02, −0.01]	[0.89, 0.90]
BAS_λ	[0.18, 0.19]	1	[0.12, 0.14]	[0.14, 0.16]	[0.04, 0.06]	[0.16, 0.17]	[0.30, 0.31]
$\text{VPIN}_{\lambda-1}$	[0.06, 0.08]	[0.12, 0.14]	1	[0.71, 0.72]	[0.04, 0.06]	[−0.05, −0.04]	[0.08, 0.10]
$\text{EXPS\_VPIN}_{\lambda-1}$	[0.19, 0.20]	[0.14, 0.16]	[0.71, 0.72]	1	[0.18, 0.20]	[−0.04, −0.03]	[0.18, 0.19]
$\text{NT}_{\lambda-1}$	[0.78, 0.78]	[0.04, 0.06]	[0.04, 0.06]	[0.18, 0.20]	1	[−0.04, −0.02]	[0.64, 0.64]
$\text{TS}_{\lambda-1}$	[−0.02, −0.01]	[0.16, 0.17]	[−0.05, −0.04]	[−0.04, −0.03]	[−0.04, −0.02]	1	[−0.05, −0.03]
$\text{RV}_{\lambda-1}$	[0.89, 0.90]	[0.30, 0.31]	[0.08, 0.10]	[0.18, 0.19]	[0.64, 0.64]	[−0.05, −0.03]	1

Table 10: Conditional probability distributions of the absolute value of intraday returns, r_λ, given $\mathtt{VPIN}_{\lambda-1}$ (Panel A) and $\mathtt{EXPS\_VPIN}_{\alpha=0.10,\lambda-1}$ (Panel B) for crude oil following Table 4 in ELO (2012). λ denotes the intraday VPIN timings.*

Panel A: Prob $(|r_\lambda|/\text{VPIN}_{\lambda-1})$

	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	2.00%	$\geq$ 2.00%
0.05	95.5	3.9	0.4	0.1	0.0	0.0	0.0	0.0	0.0
0.10	92.0	7.3	0.7	0.0	0.0	0.0	0.0	0.0	0.0
0.15	90.6	8.4	0.9	0.1	0.0	0.0	0.0	0.0	0.0
0.20	89.9	9.3	0.6	0.1	0.1	0.0	0.0	0.0	0.0
0.25	87.8	10.9	1.1	0.1	0.0	0.1	0.0	0.0	0.0
0.30	86.6	11.5	1.4	0.2	0.1	0.1	0.0	0.0	0.0
0.35	84.6	13.5	1.5	0.3	0.1	0.0	0.0	0.0	0.0
0.40	83.8	14.1	1.4	0.5	0.1	0.1	0.0	0.0	0.0
0.45	82.4	14.9	1.7	0.7	0.2	0.1	0.0	0.0	0.0
0.50	82.0	15.6	1.5	0.6	0.3	0.1	0.0	0.0	0.0
0.55	81.1	15.7	2.0	0.7	0.3	0.1	0.1	0.0	0.1
0.60	80.3	16.5	2.1	0.7	0.3	0.1	0.1	0.0	0.0
0.65	79.1	16.5	2.8	0.7	0.6	0.2	0.1	0.0	0.0
0.70	76.9	18.4	3.2	0.9	0.3	0.2	0.1	0.1	0.0
0.75	76.4	18.5	3.3	0.8	0.6	0.2	0.1	0.0	0.1
0.80	74.5	19.7	3.4	1.3	0.6	0.1	0.2	0.1	0.2
0.85	71.6	21.6	4.0	1.6	0.5	0.3	0.2	0.0	0.2
0.90	70.2	22.7	3.8	1.7	0.9	0.3	0.2	0.1	0.1
0.95	69.3	22.8	4.6	1.9	0.8	0.2	0.1	0.1	0.1
1.00	63.9	25.6	6.5	1.8	1.3	0.2	0.2	0.2	0.3

Panel B: Prob $(|r_\lambda|/\text{EXPS\_VPIN}_{\alpha=0.10,\lambda-1})$

	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	2.00%	$\geq$ 2.00%
0.05	96.9	2.9	0.2	0.0	0.0	0.0	0.0	0.0	0.0
0.10	93.9	5.7	0.4	0.0	0.0	0.0	0.0	0.0	0.0
0.15	91.2	8.0	0.7	0.2	0.0	0.0	0.0	0.0	0.0
0.20	89.0	9.8	0.9	0.2	0.0	0.1	0.0	0.0	0.0
0.25	88.3	10.7	0.9	0.2	0.1	0.0	0.0	0.0	0.0
0.30	86.6	11.9	1.3	0.1	0.0	0.0	0.0	0.0	0.0
0.35	85.1	13.6	1.1	0.1	0.1	0.1	0.0	0.0	0.0
0.40	83.9	14.1	1.3	0.5	0.2	0.0	0.0	0.0	0.0
0.45	83.5	14.6	1.6	0.2	0.1	0.0	0.0	0.0	0.0
0.50	82.1	15.8	1.4	0.5	0.2	0.0	0.0	0.0	0.0
0.55	81.4	16.1	1.7	0.4	0.3	0.1	0.1	0.0	0.0
0.60	79.6	17.0	2.5	0.6	0.3	0.0	0.0	0.0	0.0
0.65	78.9	17.1	2.7	0.8	0.3	0.1	0.1	0.0	0.1
0.70	77.0	18.9	2.8	0.8	0.4	0.1	0.0	0.0	0.0
0.75	75.2	20.0	3.4	0.8	0.5	0.1	0.1	0.0	0.0
0.80	73.5	20.5	3.8	1.4	0.6	0.2	0.0	0.0	0.0
0.85	73.9	20.2	4.0	1.2	0.4	0.2	0.1	0.0	0.1
0.90	70.1	22.4	4.5	1.8	0.7	0.3	0.2	0.1	0.0
0.95	66.9	23.7	5.1	2.2	1.1	0.5	0.3	0.0	0.1
1.00	59.2	26.1	7.1	3.3	2.0	0.7	0.4	0.5	0.7

* The row labels (first column) denotes the $\mathtt{VPIN}_{\lambda-1}$ percentiles and the column labels (top row) are the absolute value of intraday returns, r_λ, percentiles.

Table 11: Conditional probability distributions of the absolute value of intraday returns, r_λ, given $\texttt{VPIN}_{\lambda-1}$ (Panel A) and $\texttt{EXPS\_VPIN}_{\alpha=0.10,\lambda-1}$ (Panel B) for natural gas following Table 4 in ELO (2012). λ denotes the intraday VPIN timings.*

Panel A: Prob $(|r_\lambda|/\text{VPIN}_{\lambda-1})$

	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	2.00%	$\geq$ 2.00%
0.05	78.3	18.4	2.4	0.5	0.2	0.1	0.1	0.0	0.0
0.10	75.0	20.0	3.7	0.9	0.3	0.1	0.1	0.0	0.0
0.15	73.6	20.5	4.0	1.2	0.3	0.2	0.1	0.0	0.1
0.20	73.9	20.8	3.5	1.0	0.3	0.2	0.1	0.0	0.2
0.25	70.1	22.5	4.7	1.5	0.6	0.2	0.1	0.1	0.3
0.30	69.0	23.4	5.1	1.5	0.4	0.2	0.0	0.0	0.3
0.35	68.5	22.8	6.1	1.5	0.4	0.4	0.1	0.1	0.1
0.40	64.6	26.3	6.1	1.6	0.8	0.3	0.1	0.1	0.2
0.45	63.7	26.1	6.6	1.9	0.7	0.2	0.1	0.2	0.4
0.50	64.4	25.0	6.9	2.2	0.7	0.2	0.1	0.1	0.3
0.55	63.5	25.4	7.3	2.1	0.6	0.3	0.2	0.1	0.4
0.60	60.7	27.8	7.7	2.5	0.6	0.4	0.1	0.1	0.1
0.65	59.9	27.0	8.5	2.8	0.6	0.5	0.2	0.2	0.3
0.70	60.6	27.8	8.1	2.0	0.6	0.3	0.3	0.1	0.1
0.75	60.7	26.2	8.5	2.6	0.8	0.4	0.3	0.1	0.3
0.80	61.6	25.9	8.0	2.3	1.2	0.4	0.2	0.1	0.3
0.85	58.4	26.1	9.4	3.3	1.4	0.3	0.4	0.3	0.5
0.90	59.1	25.4	9.5	3.2	1.1	0.8	0.4	0.3	0.3
0.95	57.5	26.3	8.7	3.6	1.9	0.7	0.3	0.3	0.7
1.00	58.1	25.6	8.8	3.7	1.5	0.8	0.6	0.2	0.6

Panel B: Prob $(|r_\lambda|/\text{EXPS\_VPIN}_{\alpha=0.10,\lambda-1})$

	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	2.00%	$\geq$ 2.00%
0.05	85.3	13.2	1.3	0.2	0.0	0.0	0.0	0.0	0.0
0.10	80.4	17.1	2.1	0.3	0.1	0.0	0.1	0.0	0.0
0.15	76.0	20.5	2.7	0.6	0.1	0.1	0.0	0.1	0.0
0.20	73.5	21.6	3.8	0.7	0.3	0.0	0.1	0.0	0.0
0.25	71.5	23.0	4.1	1.0	0.3	0.1	0.0	0.0	0.0
0.30	71.2	23.0	4.6	0.8	0.2	0.2	0.1	0.0	0.0
0.35	67.8	25.4	5.3	1.0	0.3	0.1	0.1	0.0	0.0
0.40	65.4	26.3	6.1	1.5	0.3	0.2	0.1	0.1	0.0
0.45	63.2	27.3	7.1	1.6	0.5	0.1	0.1	0.0	0.1
0.50	64.9	25.9	6.9	1.4	0.6	0.2	0.0	0.1	0.1
0.55	63.5	25.9	7.6	1.9	0.6	0.3	0.1	0.1	0.1
0.60	63.8	25.9	6.8	2.4	0.6	0.3	0.1	0.1	0.1
0.65	60.2	27.3	8.9	2.1	0.9	0.2	0.1	0.1	0.1
0.70	61.2	26.2	7.8	3.1	0.9	0.4	0.1	0.1	0.2
0.75	58.2	27.6	9.1	2.9	0.9	0.7	0.1	0.3	0.2
0.80	57.8	26.9	9.7	3.4	1.0	0.7	0.1	0.1	0.2
0.85	54.3	27.9	10.2	4.1	1.5	0.9	0.3	0.2	0.6
0.90	54.4	27.6	10.0	3.9	1.8	0.7	0.5	0.3	0.8
0.95	52.7	26.9	10.5	5.0	2.0	0.8	0.7	0.4	0.9
1.00	47.9	25.7	11.3	5.4	2.7	2.0	1.6	0.8	2.5

* The row labels (first column) denotes the $\texttt{VPIN}_{\lambda-1}$ percentiles and the column labels (top row) are the absolute value of r_λ percentiles.

Table 12: Conditional probability distributions of $\texttt{VPIN}_{\lambda-1}$ (Panel A) and $\texttt{EXPS\_VPIN}_{\alpha=0.10,\lambda-1}$ (Panel B) given the absolute value of r_λ for crude oil following Table 4 in ELO (2012). λ denotes the intraday VPIN timings.*

Panel A: Prob $(\text{VPIN}_{\lambda-1}/|r_\lambda|)$

	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	2.00%	$\geq$ 2.00%
0.05	5.8	1.2	0.8	0.9	0.4	1.1	0.0	0.0	0.0
0.10	5.6	2.3	1.4	0.0	0.0	0.0	0.0	4.0	2.6
0.15	5.5	2.7	1.8	0.7	0.4	0.0	0.0	0.0	0.0
0.20	5.5	3.0	1.4	0.9	0.7	1.1	0.0	0.0	0.0
0.25	5.4	3.5	2.4	0.9	0.4	2.3	0.0	0.0	0.0
0.30	5.3	3.7	3.1	1.6	1.1	3.4	0.0	0.0	0.0
0.35	5.2	4.4	3.1	2.1	0.7	1.1	2.2	4.0	0.0
0.40	5.2	4.6	3.0	3.3	1.5	2.3	0.0	0.0	0.0
0.45	5.1	4.8	3.7	4.7	2.6	4.5	0.0	0.0	0.0
0.50	5.1	5.1	3.2	4.2	3.7	2.3	0.0	0.0	0.0
0.55	5.1	5.2	4.3	4.5	4.4	3.4	6.5	4.0	5.1
0.60	5.0	5.4	4.5	4.5	4.0	4.5	6.5	0.0	0.0
0.65	4.9	5.4	6.0	4.7	8.1	8.0	8.7	4.0	2.6
0.70	4.8	6.0	6.9	5.7	4.8	6.8	4.3	8.0	2.6
0.75	4.8	6.1	7.1	5.7	8.4	6.8	6.5	0.0	5.1
0.80	4.7	6.5	7.2	8.7	8.4	5.7	13.0	8.0	15.4
0.85	4.5	7.1	8.5	11.1	7.0	14.8	13.0	4.0	17.9
0.90	4.4	7.4	8.0	11.8	13.2	13.6	15.2	12.0	10.3
0.95	4.3	7.4	9.8	12.5	11.7	10.2	10.9	16.0	7.7
1.00	3.9	8.2	13.7	11.8	18.7	8.0	13.0	36.0	30.8

Panel B: Prob $(\text{EXPS\_VPIN}_{\alpha=0.10,\lambda-1}/|r_\lambda|)$

	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	2.00%	$\geq$ 2.00%
0.05	6.1	0.9	0.3	0.2	0.4	1.1	0.0	0.0	0.0
0.10	5.8	1.8	0.9	0.2	0.0	0.0	2.2	0.0	0.0
0.15	5.6	2.6	1.4	1.0	0.4	1.1	0.0	0.0	0.0
0.20	5.5	3.2	2.0	1.4	0.0	2.3	0.0	0.0	0.0
0.25	5.5	3.5	1.8	1.0	1.1	0.0	0.0	0.0	0.0
0.30	5.4	3.9	2.7	0.9	0.4	0.0	0.0	0.0	2.6
0.35	5.3	4.4	2.4	0.3	1.1	2.3	0.0	0.0	0.0
0.40	5.2	4.6	2.9	3.1	2.2	1.1	0.0	0.0	0.0
0.45	5.2	4.8	3.4	1.0	1.5	1.1	0.0	4.0	2.6
0.50	5.1	5.2	3.0	3.1	2.6	0.0	0.0	0.0	2.6
0.55	5.1	5.3	3.7	2.6	4.0	2.3	4.3	0.0	0.0
0.60	4.9	5.6	5.3	4.3	3.7	0.0	0.0	0.0	0.0
0.65	4.9	5.6	5.8	5.7	3.7	2.3	4.3	0.0	5.1
0.70	4.7	6.1	5.9	5.2	5.5	3.4	0.0	4.0	0.0
0.75	4.7	6.5	7.3	5.2	6.6	3.4	4.3	0.0	2.6
0.80	4.6	6.7	8.2	9.4	8.1	9.1	2.2	4.0	0.0
0.85	4.5	6.5	8.5	7.8	5.9	6.8	10.9	4.0	5.1
0.90	4.3	7.2	9.5	12.0	10.3	14.8	13.0	8.0	0.0
0.95	4.1	7.6	10.8	14.9	15.8	20.5	26.1	4.0	10.3
1.00	3.5	8.0	14.2	20.6	27.1	28.4	32.6	72.0	69.2

* The row labels (first column) denotes the $\texttt{VPIN}_{\lambda-1}$ percentiles and the column labels (top row) are the absolute value of intraday returns, r_λ, percentiles.

Table 13: Conditional probability distributions of $\texttt{VPIN}_{\lambda-1}$ (Panel A) and $\texttt{EXPS\_VPIN}_{\alpha=0.10,\lambda-1}$ (Panel B) given the absolute value of r_λ for natural gas following Table 4 in ELO (2012). λ denotes the intraday VPIN timings.*

Panel A: Prob $(\text{VPIN}_{\lambda-1}/|r_\lambda|)$

	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	2.00%	$\geq$ 2.00%
0.05	5.9	3.7	1.8	1.1	1.1	1.9	1.4	1.1	0.0
0.10	5.7	4.0	2.7	2.1	1.7	1.1	2.1	1.1	0.0
0.15	5.6	4.1	3.0	2.8	1.8	2.2	1.4	1.1	2.5
0.20	5.7	4.2	2.6	2.4	1.8	3.3	2.8	1.1	3.5
0.25	5.3	4.5	3.4	3.5	3.7	2.6	2.1	4.4	4.5
0.30	5.3	4.7	3.8	3.7	2.9	3.3	0.7	0.0	5.0
0.35	5.3	4.7	4.6	3.6	2.8	5.2	1.4	4.4	2.5
0.40	4.9	5.3	4.5	3.8	5.3	4.5	2.1	2.2	3.5
0.45	4.9	5.3	5.0	4.6	4.8	3.3	3.5	6.7	7.0
0.50	4.9	5.1	5.1	5.2	4.6	3.3	3.5	4.4	5.5
0.55	5.0	5.3	5.5	5.0	4.0	4.5	6.4	5.6	6.5
0.60	4.7	5.7	5.8	6.0	3.9	5.9	2.1	3.3	2.5
0.65	4.6	5.5	6.3	6.7	4.2	6.7	6.4	7.8	5.5
0.70	4.7	5.7	6.1	4.7	4.2	4.5	8.5	5.6	2.5
0.75	4.7	5.4	6.4	6.3	5.7	5.9	7.8	2.2	6.0
0.80	4.8	5.4	6.1	5.5	8.5	4.8	5.0	4.4	5.5
0.85	4.6	5.4	7.1	8.0	9.2	4.5	9.9	12.2	8.5
0.90	4.6	5.3	7.2	7.7	7.2	11.5	9.9	12.2	5.5
0.95	4.4	5.4	6.5	8.6	12.7	10.0	7.1	11.1	12.5
1.00	4.4	5.2	6.5	8.8	9.9	10.8	15.6	8.9	11.0

Panel B: Prob $(\text{EXPS\_VPIN}_{\alpha=0.10,\lambda-1}/|r_\lambda|)$

	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	2.00%	$\geq$ 2.00%
0.05	6.8	2.8	1.0	0.5	0.0	0.0	0.0	0.0	0.0
0.10	6.4	3.6	1.6	0.8	0.6	0.0	1.4	0.0	0.5
0.15	6.1	4.3	2.0	1.4	0.7	0.7	0.7	2.2	0.5
0.20	5.8	4.5	2.9	1.8	1.8	0.4	1.4	0.0	0.5
0.25	5.7	4.9	3.2	2.4	2.0	1.1	0.0	1.1	0.0
0.30	5.6	4.8	3.5	1.9	1.1	2.2	1.4	1.1	0.0
0.35	5.3	5.3	4.0	2.6	1.8	1.5	1.4	1.1	0.5
0.40	5.2	5.5	4.7	3.7	2.4	2.2	2.1	2.2	0.5
0.45	4.9	5.7	5.4	3.9	3.3	1.9	2.1	1.1	2.0
0.50	5.0	5.3	5.2	3.3	4.0	2.6	0.0	2.2	1.5
0.55	5.0	5.4	5.8	4.6	4.4	4.1	2.1	2.2	1.0
0.60	5.0	5.4	5.1	5.8	4.4	3.7	2.1	2.2	2.5
0.65	4.7	5.6	6.7	5.0	5.9	2.6	3.5	2.2	2.5
0.70	4.8	5.4	5.9	7.5	5.9	5.2	2.1	4.4	3.5
0.75	4.5	5.7	6.8	6.9	6.1	9.3	2.8	11.1	3.5
0.80	4.4	5.5	7.2	8.0	6.4	10.0	3.5	4.4	4.5
0.85	4.1	5.6	7.5	9.6	10.1	11.5	8.5	8.9	10.0
0.90	4.1	5.5	7.2	9.0	11.9	8.9	13.5	10.0	14.0
0.95	3.8	5.1	7.3	11.2	12.5	10.0	17.0	15.6	16.0
1.00	3.0	4.2	6.8	10.3	14.5	21.9	34.0	27.8	36.5

* The row labels (first column) denotes the $\texttt{VPIN}_{\lambda-1}$ percentiles and the column labels (top row) are the absolute value of r_λ percentiles.

Table 14: Regression results for the dependent variable bid-ask spread (BAS_λ) and lagged independent variables $\text{VPIN}_{\lambda-1}$, $\text{EXPS\_VPIN}_{\lambda-1}$, number of trades ($\text{NT}_{\lambda-1}$), trade size ($\text{TS}_{\lambda-1}$) and realized volatility ($\text{RV}_{\lambda-1}$). λ denotes the intraday VPIN timings.*

$$\text{BAS}_\lambda = \beta_0 \text{VPIN}_{\lambda-1} + \epsilon \qquad \text{I}$$
$$\text{BAS}_\lambda = \beta_1 \text{EXPSVPIN}_{\lambda-1} + \epsilon \qquad \text{II}$$
$$\text{BAS}_\lambda = \beta_0 \text{VPIN}_{\lambda-1} + \beta_2 \text{NT}_{\lambda-1} + \beta_3 \text{TS}_{\lambda-1} + \epsilon \qquad \text{III}$$
$$\text{BAS}_\lambda = \beta_1 \text{EXPSVPIN}_{\lambda-1} + \beta_2 \text{NT}_{\lambda-1} + \beta_3 \text{TS}_{\lambda-1} + \epsilon \qquad \text{IV}$$
$$\text{BAS}_\lambda = \beta_0 \text{VPIN}_{\lambda-1} + \beta_2 \text{NT}_{\lambda-1} + \beta_3 \text{TS}_{\lambda-1} + \beta_4 \text{RV}_{\lambda-1} + \epsilon \qquad \text{V}$$
$$\text{BAS}_\lambda = \beta_1 \text{EXPSVPIN}_{\lambda-1} + \beta_2 \text{NT}_{\lambda-1} + \beta_3 \text{TS}_{\lambda-1} + \beta_4 \text{RV}_{\lambda-1} + \epsilon \qquad \text{VI}$$

Panel A: Crude Oil Regression

	I	II	III	IV	V	VI
Intercept	0.02	0.01	−0.002	−0.002	−0.001	−0.002
	(188.17)	(226.50)	(−60.93)	(−57.25)	(−63.37)	(−75.51)
$\text{VPIN}_{\lambda-1}$	−0.01		0.008		0.001	
	(−20.41)		(60.09)		(11.96)	
$\text{EXPS\_VPIN}_{\lambda-1}$		−0.009		0.007		0.003
		(−17.78)		(68.40)		(34.82)
$\text{NT}_{\lambda-1}$ ($\times 10^6$)			0.73	0.68	0.39	0.37
			(115.76)	(107.49)	(83.60)	(79.99)
$\text{TS}_{\lambda-1}$ ($\times 10^3$)			0.64	0.60	0.30	0.29
			(54.47)	(51.25)	(34.63)	(34.47)
$\text{RV}_{\lambda-1}$					0.009	0.009
					(270.97)	(271.14)
R^2	0.01	0.00	0.19	0.20	0.58	0.59
F	417	316	6,029	6,449	27,170	27,812

Panel B: Natural Gas Regression

	I	II	III	IV	V	VI
Intercept	0.001	0.001	−0.001	−0.001	−0.003	−0.003
	(270.31)	(348.16)	(−22.39)	(−27.59)	(−87.74)	(−113.66)
$\text{VPIN}_{\lambda-1}$	0.0005		0.003		−0.0003	
	(35.04)		(16.03)		(−2.97)	
$\text{EXPS\_VPIN}_{\lambda-1}$		0.0004		0.003		0.0005
		(40.67)		(21.94)		(6.50)
$\text{NT}_{\lambda-1}$ ($\times 10^6$)			3.11	3.08	1.41	1.40
			(321.25)	(312.20)	(197.87)	(196.16)
$\text{TS}_{\lambda-1}$ ($\times 10^3$)			0.11	0.11	0.19	0.19
			(8.31)	(8.21)	(24.88)	(25.23)
$\text{RV}_{\lambda-1}$					0.01	0.01
					(385.02)	(384.28)
R^2	0.02	0.02	0.59	0.59	0.86	0.86
F	1,228	1,654	34,749	34,930	115,991	116,053

* The results in parentheses are the t-statistics for associated coefficient estimates.

Table 15: Regression results for the dependent variable absolute return ($|r_\lambda|$) and lagged independent variables $\text{VPIN}_{\lambda-1}$, $\text{EXPS\_VPIN}_{\lambda-1}$, number of trades ($\text{NT}_{\lambda-1}$), trade size ($\text{TS}_{\lambda-1}$) and realized volatility ($\text{RV}_{\lambda-1}$). λ denotes the intraday VPIN timings.*

$$|r_\lambda| = \beta_0 \text{VPIN}_{\lambda-1} + \epsilon \qquad \text{I}$$
$$|r_\lambda| = \beta_1 \text{EXPSVPIN}_{\lambda-1} + \epsilon \qquad \text{II}$$
$$|r_\lambda| = \beta_0 \text{VPIN}_{\lambda-1} + \beta_2 \text{NT}_{\lambda-1} + \beta_3 \text{TS}_{\lambda-1} + \epsilon \qquad \text{III}$$
$$|r_\lambda| = \beta_1 \text{EXPSVPIN}_{\lambda-1} + \beta_2 \text{NT}_{\lambda-1} + \beta_3 \text{TS}_{\lambda-1} + \epsilon \qquad \text{IV}$$
$$|r_\lambda| = \beta_0 \text{VPIN}_{\lambda-1} + \beta_2 \text{NT}_{\lambda-1} + \beta_3 \text{TS}_{\lambda-1} + \beta_4 \text{RV}_{\lambda-1} + \epsilon \qquad \text{V}$$
$$|r_\lambda| = \beta_1 \text{EXPSVPIN}_{\lambda-1} + \beta_2 \text{NT}_{\lambda-1} + \beta_3 \text{TS}_{\lambda-1} + \beta_4 \text{RV}_{\lambda-1} + \epsilon \qquad \text{VI}$$

Panel A: Crude Oil Regression

	I	II	III	IV	V	VI
Intercept	0.0005	0.0005	−0.002	−0.002	−0.001	−0.002
	(27.75)	(34.45)	(−60.93)	(−57.25)	(−63.37)	(−75.51)
$\text{VPIN}_{\lambda-1}$	0.009		0.008		0.001	
	(60.85)		(60.09)		(11.96)	
$\text{EXPS\_VPIN}_{\lambda-1}$		0.009		0.007		0.003
		(80.61)		(68.40)		(34.82)
$\text{NT}_{\lambda-1}$ ($\times 10^6$)			0.73	0.68	0.39	0.37
			(115.76)	(107.49)	(83.60)	(79.99)
$\text{TS}_{\lambda-1}$ ($\times 10^3$)			0.64	0.60	0.30	0.29
			(54.47)	(51.25)	(34.63)	(34.47)
$\text{RV}_{\lambda-1}$					0.009	0.009
					(270.97)	(271.14)
R^2	0.05	0.08	0.19	0.20	0.58	0.59
F	3,703	6,498	6,029	6,449	27,170	27,812

Panel B: Natural Gas Regression

	I	II	III	IV	V	VI
Intercept	0.001	−0.0005	−0.001	−0.001	−0.003	−0.003
	(14.32)	(−8.38)	(−22.39)	(−27.59)	(−87.74)	(−113.66)
$\text{VPIN}_{\lambda-1}$	0.005		0.003		−0.0003	
	(20.54)		(16.03)		(−2.97)	
$\text{EXPS\_VPIN}_{\lambda-1}$		0.01		0.003		0.0005
		(55.95)		(21.94)		(6.50)
$\text{NT}_{\lambda-1}$ ($\times 10^6$)			3.11	3.08	1.41	1.40
			(321.25)	(312.20)	(197.87)	(196.16)
$\text{TS}_{\lambda-1}$ ($\times 10^3$)			0.11	0.11	0.19	0.19
			(8.31)	(8.21)	(24.88)	(25.23)
$\text{RV}_{\lambda-1}$					0.01	0.01
					(385.02)	(384.28)
R^2	0.01	0.04	0.59	0.59	0.86	0.86
F	422	3,131	34,749	34,930	115,991	116,053

* The results in parentheses are the t-statistics for associated coefficient estimates.

Figures

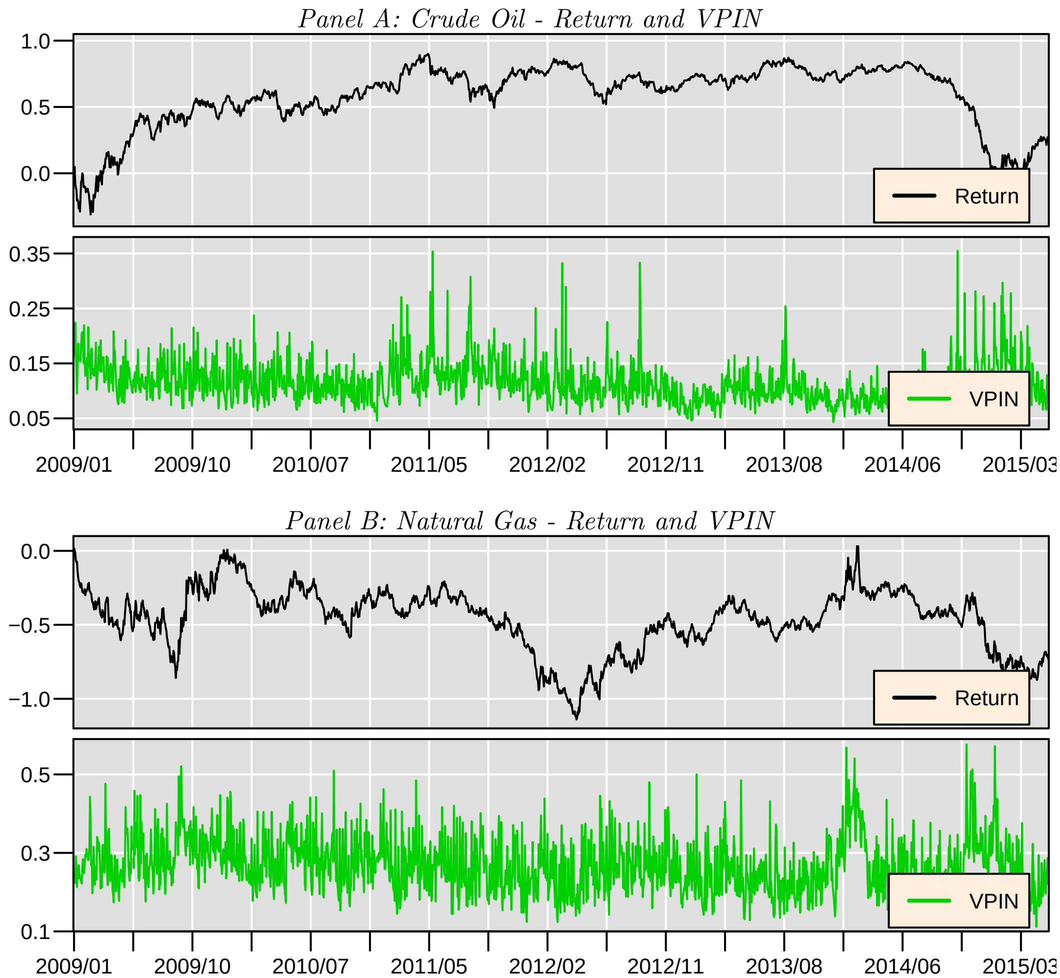

Figure 1: The figure graphs daily returns (blue) and `VPIN` (green) for crude oil (Panel A) and natural gas (Panel B). The daily returns are based on closing prices, $\log(p_t/p_{t-1})$. The daily `VPIN` is represented by the last `VPIN` per day. The `VPIN` calculation is based on one-minute time bars, and averaged over a window with 50 observations.

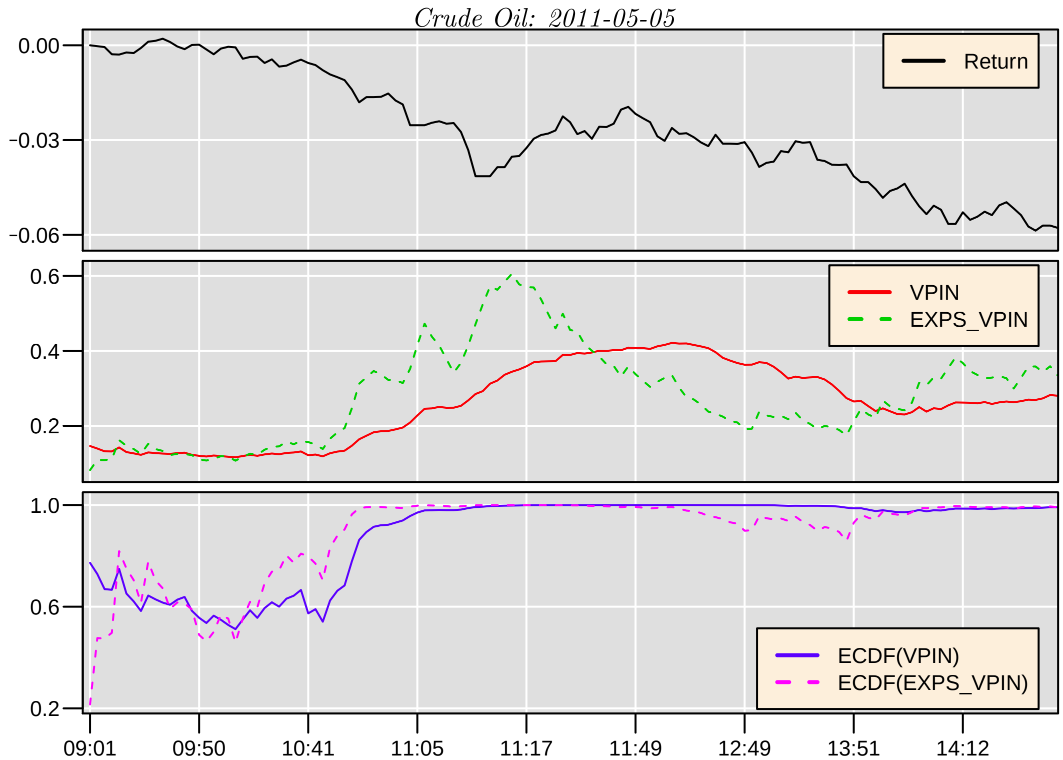

Figure 2: The figure graphs the intraday time series of returns, `VPIN` and $\texttt{EXPS\_VPIN}_{\alpha=0.10}$, and ECDF of the toxicity metrics for crude oil on May 5, 2011. The x-axis is time stamps in minutes. The top panel plots the intraday return, $\log(p_{t_i}/p_{t_{i-1}})$. The second panel plots the `VPIN` (red continuous line) and $\texttt{EXPS\_VPIN}_{\alpha=0.10}$ (green dashed line). The third panel plots the `ECDF`(`VPIN`) (dark blue continuous line) and ECDF(EXPS_VPIN) (pink dashed line). The `VPIN` and $\texttt{EXPS\_VPIN}_{\alpha=0.10}$ calculations are based on one-minute time bars and averaged over a window with 50 observations.

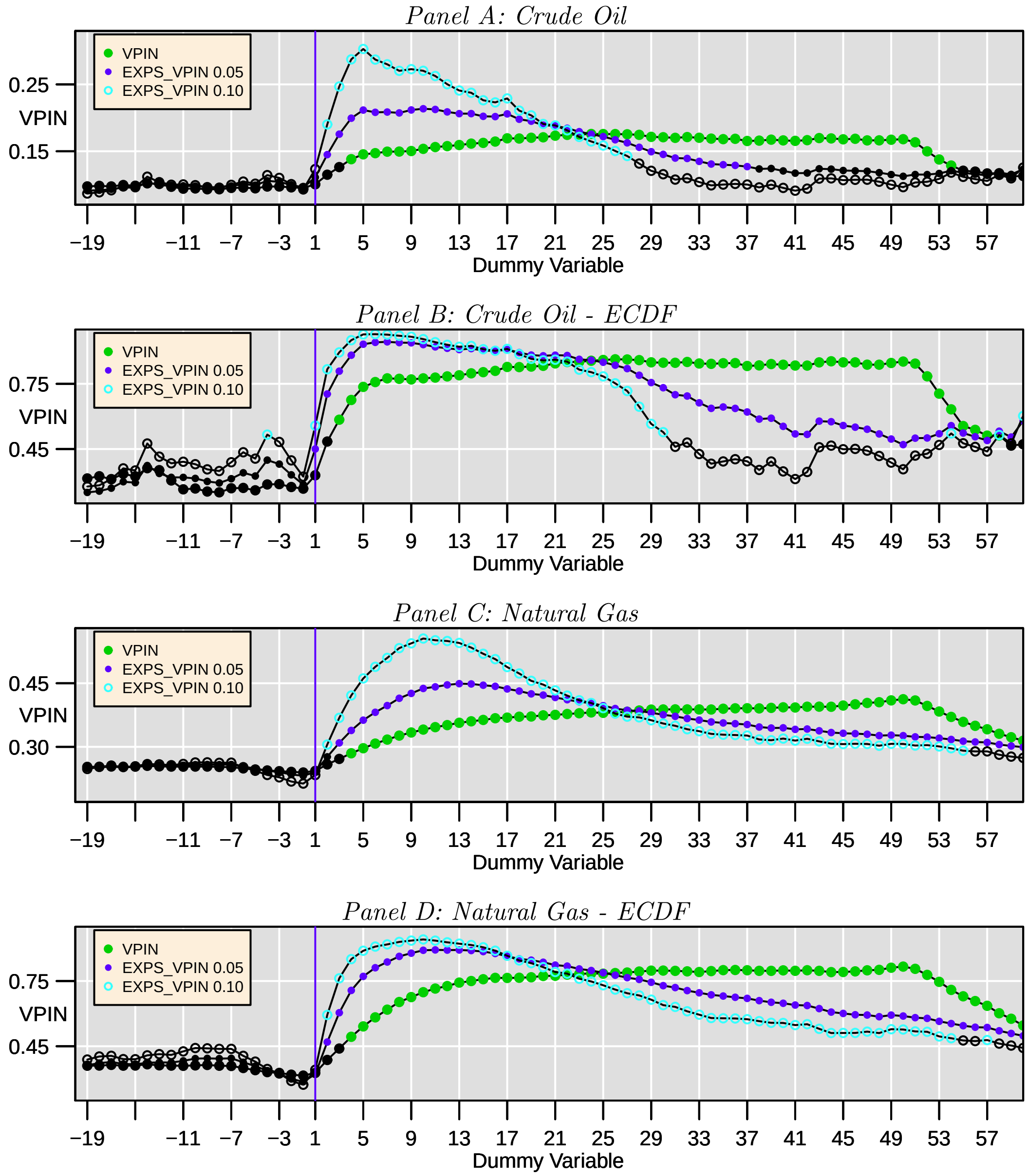

Figure 3: The figure graphs the mean values of `VPIN` (y-axis) at the release of the inventory levels in crude oil (Panel A) and natural gas (Panel C). Values from applying the empirical cumulative density function to the `VPIN` metrics are plotted in Panel B (crude oil) and Panel D (natural gas). Since `VPIN` estimates are asynchronous, they are indexed in order relative to the timing of the inventory release where index -1 and 1 denote the first observations immediately before and after the release, respectively. Only days with a negative and significant jump and a surprise greater than one standard deviation are included in the regression where 20 `VPIN` estimates prior to and 60 following the inventory release are considered. Non-black dots denote estimates which are significantly greater (t-stat greater than 1.68) than the dummy variable D_{-20}, which is used as the basis. The three plotted time series compare results for different `VPIN` calculations. `VPIN` (green dots), $\texttt{EXPS\_VPIN}_{\alpha=0.10}$ (dark blue dots), and $\texttt{EXPS\_VPIN}_{\alpha=0.10}$ (light blue dots). The `VPIN` calculation is based on one minute time bars. All estimates are based on a window with 50 observations.

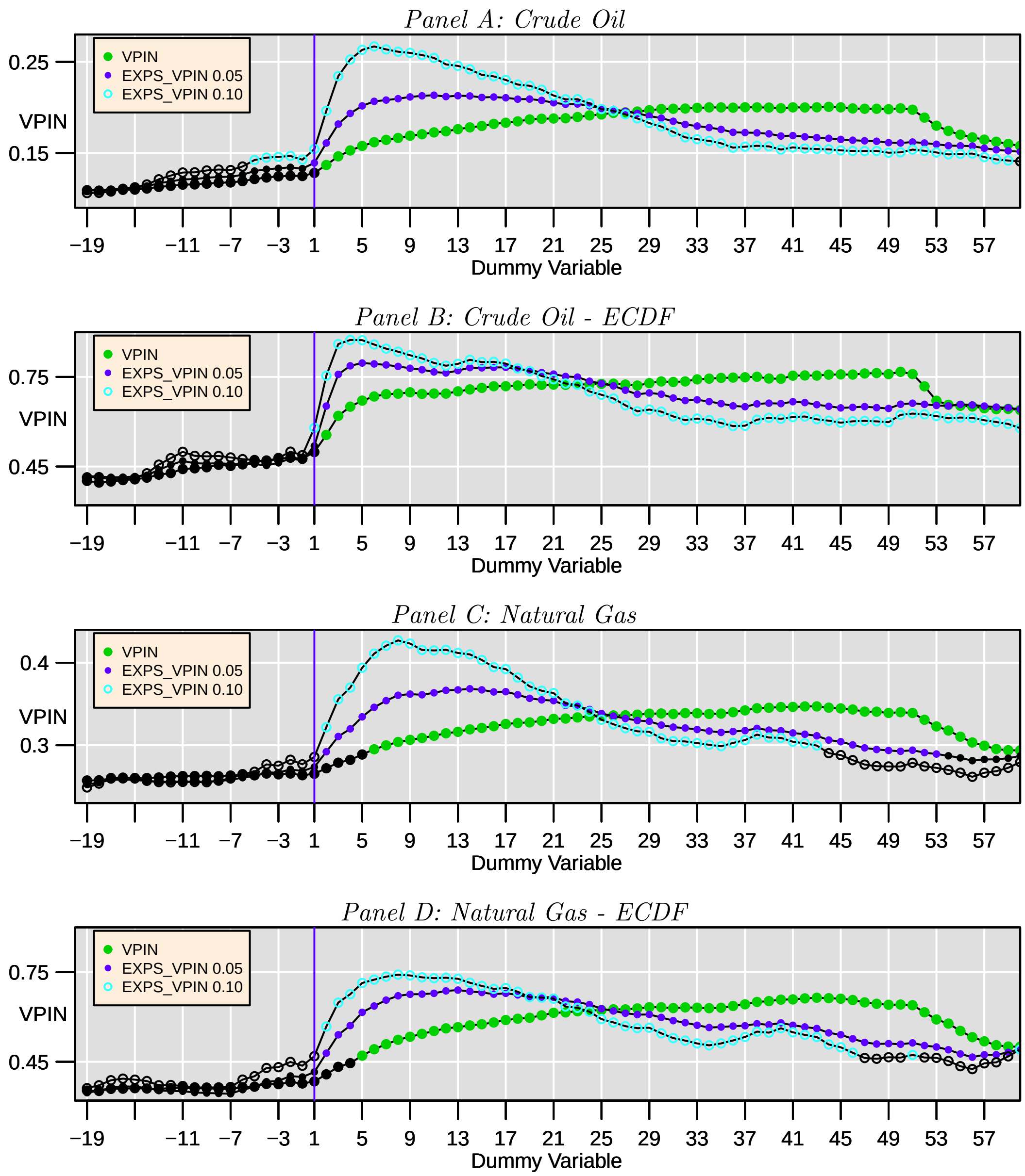

Figure 4: The figure graphs the mean values of `VPIN` (y-axis) at jumps in crude oil (Panel A) and natural gas (Panel C) returns which are not associated with the releases of inventory levels. Values from applying the empirical cumulative density function to the `VPIN` metrics are plotted in Panel B (crude oil) and Panel D (natural gas). Since `VPIN` estimates are asynchronous, they are indexed in order relative to the timing of the inventory release where index -1 and 1 denote the first observations immediately before and after the release, respectively. 20 `VPIN` estimates prior to and 60 following the inventory release are considered. Non-black dots denote estimates which are significantly greater (t-stat greater than 1.68) than the dummy variable D_{20}, which is used as the basis. The three plotted time series compare results for different `VPIN` calculations. `VPIN` (green dots), `EXPS_VPIN`$_{\alpha=0.10}$ (dark blue dots), and `EXPS_VPIN`$_{\alpha=0.10}$ (light blue dots). The `VPIN` calculation is based on one minute time bars. All estimates are based on a window with 50 observations.

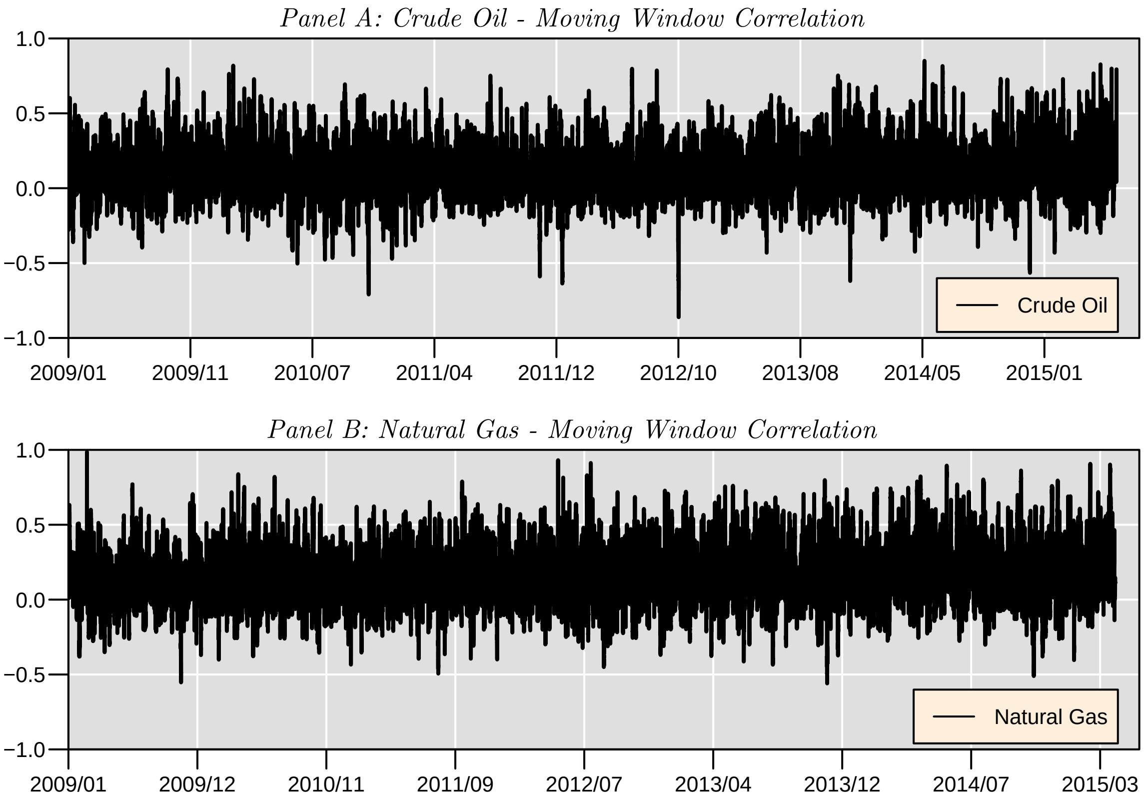

Figure 5: The figures plot the difference between the moving window correlations $\rho_{|r_\lambda|,\texttt{EXPS\_VPIN}_{\lambda-1}} - \rho_{|r_\lambda|,\texttt{VPIN}_{\lambda-1}}$ where $\rho_{|r_\lambda|,\texttt{EXPS\_VPIN}_{\lambda-1}}$ is the correlation between absolute values of intraday returns ($|r_\lambda|$) and $EXPS\_VPIN_{\lambda-1}$; and $\rho_{|r_\lambda|,\texttt{VPIN}_{\lambda-1}}$ is the correlation between $|r_\lambda|$ and $VPIN_{\lambda-1}$. The moving window is based on 50 observations.

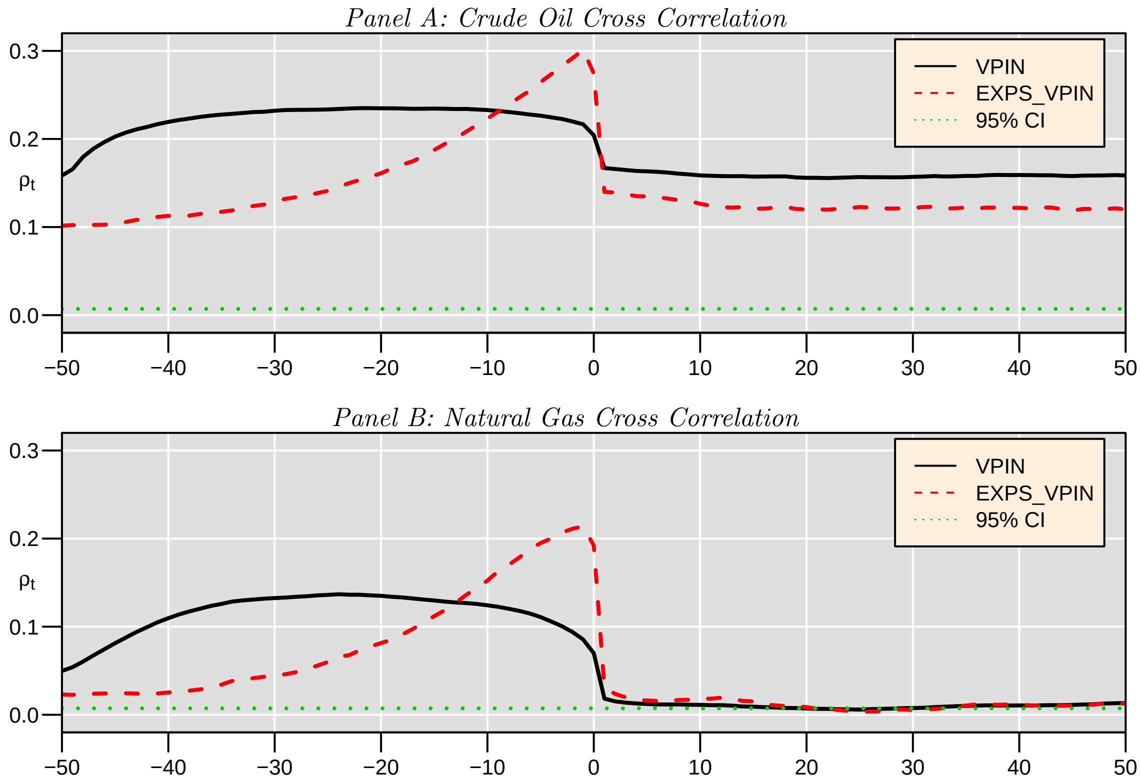

Figure 6: Cross correlation between absolute returns and vpin. The x-axis labels denote lags between the returns and vpin. Negative lags denote that VPIN is lagged relative to absolute returns while positive lags denote that the return is lagged to VPIN.

A Appendix

A.I Contract Specifications

Table A.1: Key features of contract specifications for crude oil, heating oil and natural gas.

Panel A: Light, Sweet Crude Oil Futures
Trading Unit
1000 U.S. barrels (42000 gallons)
Price Quotation
U.S. dollars and cents per barrel
Trading Hours
Open outcry trading is conducted from 9:00 AM until 2:30 PM.
Trading Months
Crude oil futures are listed nine years forward using the following listing schedule: consecutive months are listed for the current year and the next five years; in addition, the June and December contract months are listed beyond the sixth year.
Minimum Price Flucuation
\$0.01 (1¢) per barrel (\$10.00 per contract).
Maximum Daily Price Flucuation
\$10.00 per barrel (\$10,000 per contract) for all months.
Last Trading Day
Trading terminates at the close of business on the third business day prior to the 25th calendar day of the month preceding the delivery month. If the 25th calendar day of the month is a non-business day, trading shall cease on the third business day prior to the business day preceding the 25th calendar day.
Settlement Type
Physical
Delivery
F.O.B. seller's facility, Cushing, Oklahoma, at any pipeline or storage facility with pipeline access to TEPPCO, Cushing storage, or Equilon Pipeline Co., by in-tank transfer, in-line transfer, book-out, or inter-facility transfer (pumpover).
Trading Symbol
CL
Panel B: Henry Hub Natural Gas Futures
Trading Unit
10,000 million British thermal units (mmBtu).
Price Quotation
U.S. dollars and cents per barrel
Trading Hours
Open outcry trading is conducted from 9:00 AM until 2:30 PM.
Trading Months
The current year plus the next twelve years through December 2020. A new calendar year will be added following the termination of trading in the December contract of the current year.
Minimum Price Flucuation
\$0.001 (0.1¢) per mmBtu (\$10.00 per contract).
Maximum Daily Price Flucuation
\$3.00 per barrel (\$30,000 per contract) for all months.
Last Trading Day
Trading terminates three business days prior to the first calendar day of the delivery month.
Settlement Type
Physical
Delivery
The Sabine Pipe Line Co. Henry Hub in Louisiana. Seller is responsible for the movement of the gas through the Hub; the buyer, from the Hub. The Hub fee will be paid by seller.
Trading Symbol
NG