ISSN 1936-5330

kc FED Research Working Papers

Big Data Meets the Turbulent Oil Market: Supplementary Appendix

Charles W. Calomiris, Nida Çakır Melek, and Harry Mamaysky

November 2022
RWP 20-20
http://doi.org/10.18651/RWP2020-20

Federal Reserve Bank of Kansas City

Big Data Meets the Turbulent Oil Market: Supplementary Appendix

Charles W. Calomiris, Nida Çakır Melek, and Harry Mamaysky[1]

November 2022

This Online Appendix contains details on our timing convention, futures return calculation, energy word sources and corresponding reference list, bootstrap methodology, as well as additional appendix figures and tables referred to in the main body of the paper.

1. Timing convention for variable construction

Oil spot and futures prices are available at 2:30pm ET on each trading day and the oil majors' stock prices are available after 4pm ET. One of the oil majors we consider, Royal Dutch Shell (RDS), trades in Europe and its Friday close occurs in the morning ET. So, for RDS, we calculate its forward return from Monday of week $t+1$ to Monday of week $t+9$. We measure the price-based dependent variables on Fridays (or the next Monday for RDS) to make sure there is no overlap with the physical data releases in week t. For example, in some weeks physical data releases can be delayed by one or two days, and may come out on either Thursday or Friday at 10:30am ET. Therefore, starting our dependent variable measurements from Friday at 2:30pm ET or later is conservative and ensures no overlap between the dependent price-based series and the physical oil forecasting variables. Our oil volatility measure, *DOilVol*, is the change in the trailing 30 trading-day realized volatility of WTI prices from the Friday of week t to the Friday of week $t+8$. Hence, this dependent variable has no overlap with any week t explanatory variable. However, we cannot run four-week ahead forecasting regressions with this variable due to overlap with the four-week lagged explanatory variables. Given the data release schedule for oil inventories and production, our physical (Wednesday) and price-based (Friday or Monday) timing convention for dependent variables is the natural choice for any analysis of U.S. oil markets that involves both physical and price-based data.

[1]Columbia Business School and NBER, cc374@gsb.columbia.edu, Federal Reserve Bank of Kansas City, nida.cakirmelek@kc.frb.org, and Columbia Business School, hm2646@columbia.edu, respectively. Authors are listed in alphabetical order. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City, or the Federal Reserve System.

Since we have two timing conventions for our dependent variables, we need to have two versions of all explanatory variables as well: those that are used for the physical forecasting regressions, and those that are used for the price-based forecasting regressions. For the physical regressions, we use values for variables that are the most recently available ones on the Tuesday of week *t*. This ensures no overlap with the physical dependent variables which are released on Wednesday of week *t* at the earliest. For the price-based regressions, we can use the latest value of a forecasting variable that is not after 2:30pm ET on the Friday of week *t*, otherwise there would be overlap with the forward-looking oil spot and oil futures returns. For example, the week *t* Friday S&P 500 return would overlap with the oil spot and futures dependent variable for that week, because the S&P 500 return is measured as of 4pm ET on Friday. In all such cases, we use the Thursday end of period (EOP) value of the forecasting variable. In cases where the right-hand side variable does not overlap with the Friday dependent variables, such as lagged futures returns, we use the Friday value of the forecasting variable.

Occasionally, a dependent or explanatory variable observation is not available in a given week *t*, often due to holidays. In this case, we would be unable to calculate either the forward-looking week *t* dependent variable or the backward-looking week *t* forecasting variable. When a week *t* price-based dependent variable is missing its Friday level, we calculate forward returns using either the Monday or Tuesday (if Monday is missing) price point of week *t+1* (for RDS, a missing Monday observation would be replaced with either the Tuesday or Wednesday one). For the price-based dependent variable regressions, when a right-hand side variable is missing its week *t* observation and is measured as of Thursday, we use either the week *t* Wednesday or Tuesday value; if the right-hand side variable is measured as of Friday, we use either the week *t* Thursday or Wednesday value. For the physical regressions, when we are missing a week *t* Tuesday value for an explanatory variable, we use the week *t* Monday or week *t-1* Friday value instead. These rules allow us to avoid missing a dependent or independent variable due to a single missing observation, while avoid overlaps between explanatory and dependent variables.

1.1. Details of timing convention for variable construction

Unless explicitly noted, all days refer to week *t*. For example, when used as explanatory variables, BP (*bpRet*) and ExxonMobil (*xomRet*) are the Thursday-to-Thursday returns, and RDS (*rdsaRet*) is Friday-to-Friday (because it trades on Euronext, its Friday close happens

before 2:30pm ET) for the price-based regressions; and all are measured Tuesday-to-Tuesday for the physical regressions.

Variables	Timing Convention	
	Prices Regressions	Physical Regressions
Dependent Variables		
FutRet	Friday	-
DSpot	Friday	-
DOilVol	Friday	-
xomRet	Friday	-
bpRet	Friday	-
rdsaRet	Monday of week *t+1*	-
DInv	-	Wednesday
DProd	-	Wednesday
Independent Variables		
FutRet	Friday	Tuesday
DSpot	Friday	Tuesday
DOilVol	Thursday	Tuesday
xomRet	Thursday	Tuesday
bpRet	Thursday	Tuesday
rdsaRet	Friday	Tuesday
StkIdx	Thursday & Friday	Tuesday
DInv	Wednesday	Wednesday
DProd	Wednesday	Wednesday
OilVol	Thursday	Tuesday
VIX	Thursday	Tuesday
vix_diff	Thursday	Tuesday
ovx_diff	Thursday	Tuesday
sdf_fullSample	Thursday	Tuesday
DFX	Thursday	Tuesday
tnote_10y	Thursday	Tuesday
sp500Ret	Thursday	Tuesday
basis	Friday	Tuesday

WIPI	Monthly	Monthly
trend	Monthly	Monthly
BE/ME	Monthly	Monthly
Mom	Monthly	Monthly
BasMom	Monthly	Monthly
InflaBeta	Monthly	Monthly
DolBeta	Monthly	Monthly
HedgPres	Friday of week *t-1*	Friday of week *t-1*
OpenInt	Thursday	Tuesday
liquidity	Thursday	Tuesday
f[Topic]	Friday	Tuesday
s[Topic]	Friday	Tuesday
artcount	Friday	Tuesday
entropy	Friday	Tuesday
PCAsent	Friday	Tuesday
PCAfreq	Friday	Tuesday
PCAall	Friday	Tuesday

The general philosophy is to take a variable from day *t-1* if it's not clear that it won't overlap with a day *t* variable, and to take it from day *t* if we are sure there is no look-ahead bias. For example, we calculate text-based series to end at 2:30pm on each day, so there is no overlap with these at either the 2:30pm energy futures market closes or the 4pm stock market closes. Some of the independent series are from the prior month.

We calculate Momentum (Mom) for month *m* as the cumulative return from months *m-11* to *m-1*, to be consistent with standard practice. After construction, we use month *m*'s Mom value as a regressor for future four- or eight-week outcomes that start in month *m+1*.

The physical regressions use levels of industrial production and inventories that are reported on Wednesday at 10am. Therefore, the explanatory variables for these are from Tuesday or earlier. The oil production and inventory variables are occasionally reported on either Thursday or even Friday. For this reason, when we are forecasting price-based series that use the physical oil market variables as forecasting variables, we use Friday and forward market-based observations to make sure there is no overlap with the production or inventory series in that week.

For Oil Volatility (OilVol), it is not certain whether Bloomberg uses futures settlement price or last traded price for calculating OilVol. Therefore, we use its Thursday or Tuesday value as an independent variable. Accordingly, for the level difference of OilVol, i.e. DOilVol, we use its Thursday or Tuesday value as an independent variable and use its Friday value as a dependent variable.

For Exxon Mobil stock returns (xomRet) and British Petrol stock returns (bpRet), we use Thursday or Tuesday values as an independent variable and use their Friday value as a dependent variable. We use BP PLC American Depository Receipts (ADR) Stock Price for calculating bpRet. For Royal Dutch Shell class A stock returns (rdsaRet), we use its Friday or Tuesday value as an independent variable and use its Monday week t+1 value when it is a dependent variable because it trades in Europe and has a 10 or 11am NY close.

For Open Interest (OpenInt) calculation, we take logarithm of the product of: WTI spot price, quantity of WTI futures contracts outstanding, and WTI futures contract size. The data for open interest generally comes in around 18:00 EST but this may vary depending on the exact time the exchange sends data to Bloomberg. Hence, we use Thursday or Tuesday value as an independent variable.

For Hedging Pressure (HedgPres), we use its Friday value from the week *t-1* as an independent variable for both prices and physical regressions. The Commodity Futures Trading Commission (CFTC) provides long and short hedge positions data for "CRUDE OIL, LIGHT SWEET - NEW YORK MERCANTILE EXCHANGE" on a weekly basis. The data for week *t* are as of Tuesday, and are released on Friday of week *t* at 3:30pm EST.

2. Construction of cumulative weekly futures returns

If on a given Friday, the front-month futures contract will not expire before the next Friday, the weekly return is defined as

$$R(t+1) = \frac{F1(t+1) - F1(t)}{F1(t)}$$

where F1(t+1) is the price of the front-month futures in one week's time and F1(t) is the price of the same futures today. If on a given Friday, the front-month futures will expire before the next Friday, we switch to the second-month futures contract and define the weekly return as

$$R(t+1) = \frac{F1(t+1) - F2(t)}{F2(t)}$$

where F2(t) is the price of the second-month futures contract today. Then, we obtain n-week ahead weekly cumulative returns as

$$\big(1 + R(t+1)\big) \times \big(1 + R(t+2)\big) \times \cdots \times \big(1 + R(t+n+1)\big) \times 100.$$

Constructing weekly returns series

We use weekly (Friday, end-of-period (EOP)) values for the front-month and second-month contracts. To figure out when the front-month expires and the second-month becomes the front-month, we follow the following steps.

1. *Find range of dates for Fridays with a new contract*

The CME Group describes termination of trading for these contracts as: "Trading terminates 3 business day prior to the twenty-fifth calendar day of the month prior to the contract month. If the twenty-fifth calendar day is not a business day, trading terminates 3 business days prior to the business day preceding the twenty-fifth calendar day."[2] Based on this information, we can find Fridays signaling the start of a new contract.

Assume 25th falls on a	Trading ends	Trading ends (assuming 1-day holiday)	Friday with new contract	Friday with new contract (if 1-day holiday)
Saturday	21st (Tues.)	20th (Mon.)	24th	-
Sunday	20th (Tues.)	19th (Mon.)	23rd	-
Monday	20th (Wed.)	19th (Tues.)	22nd	-
Tuesday	20th (Thurs.)	19th (Wed.)	21st	-
Wednesday	20th (Fri.)	19th (Thurs.)	27th	20th
Thursday	22nd (Mon.)	19th (Fri.)	26th	-
Friday	22nd (Tues.)	21st (Mon.)	25th	-

2. *Identify market holidays that might fall on Friday 20th - Wednesday 25th of a month*

Based on the first step, the only time a holiday would change the Friday on which the front-month contract rolls over is when the 25th falls on a Wednesday. Thus, we need to inspect observations falling on the 27th to see if there's a holiday that would lead to the 20th being the Friday with the new contract.

[2] For more information, see https://www.cmegroup.com/trading/energy/crude-oil/light-sweet-crude_contract_specifications.html.

We want to see if a holiday fell on Friday 20th – Wednesday 25th of a month for 1998-2020.

Holiday[3]	Check based on timing?	When would it need to fall for the new contract to begin on the 20th as opposed to the 27th?
New Years Day	N	-
Martin Luther King, Jr. Day	Y	Always on a Monday, would need to fall on Monday 23rd, but the latest it's fallen on is Monday 21st Friday 27th still when new contract begins
Washington's Birthday	Y	Always on a Monday, would need to fall on Monday 23rd, but the latest it's fallen on is Monday 21st Friday 27th still when new contract begins
Good Friday	Y	Always on a Friday, would need to fall on Friday 20th, which it doesn't for 1998-2019 Friday 27th still when new contract begins
Memorial Day	Y	Always on a Monday, would need to fall on Monday 23rd, but the earliest it's fallen on is Monday 25th Friday 27th still when new contract begins
Independence Day	N	-
Labor Day	N	-
Thanksgiving Day	Y	Always on a Thursday, more than 3 business days before the 25th of a month, therefore we do not need to consider this holiday's effect if the 25th fell on a Wednesday
Christmas	Y	Always on the 25th, would need to fall on a Wednesday for the new contract to start on Friday 20th as opposed to Friday 27th. This was the case in 2002 and 2013, and will be the case in 2019. Thus, in December of 2002 and 2013, the new contract will begin on Friday 20th instead of the 27th

3. *Construct dummy for when to calculate returns using F2, based on steps 1 and 2*

Using step one, create a dummy that equals to 1 if the day is between 21 and 27, inclusive; zero otherwise. Using step two:

If day = 27, month = 12, and (year = 2002 or 2013), change dummy back to 0

If day = 20, month = 12, and (year = 2002 or 2013), change dummy to 1

Using this dummy, we can compute returns using F1 and F2 such that, if dummy = 0,

$$R_t = \frac{F1_t - F1_{t-1}}{F1_{t-1}};$$

and if dummy = 1,

$$R_t = \frac{F2_t - F2_{t-1}}{F2_{t-1}},$$

[3] Refers to NYSE holidays: https://www.nyse.com/markets/hours-calendars.

all in one series. Here, F1 refers only to the front-month series (CL1), and F2 only to the second-month series (CL2), unlike in the definition section where we change notation when referring to the same series in equation above.

To fill in missing Friday observations, we rely on Thursday prices. If Thursday prices are missing, we use Wednesday prices. We only have to go as far back as Wednesday to eliminate all missing observations. We adapt the methodology above to Thursday and Wednesday prices.

3. Sources used to derive energy word lists

Our full list of sources are as follows: Dahl (2004), Downey (2009), Geman (2005), Griffin and Steele (1986), Griffin and Teece (1982), Raymond and Leffler (2006), Yergin (1992), Yergin (2011), as well as Deutsche Bank (2013), Devold (2013), the IEA Oil Market Report glossary, and the Platts energy industry glossary, covering common terms and abbreviations from the oil, power, petrochemicals, nuclear, gas, coal and metals markets. See the References section at the end of this document.

4. Four-week ahead forecasting regressions

Table VIII of this Online Appendix shows the results of the main in-sample forecasting regressions for four-week ahead returns/changes. These are analogous to the eight-week ahead results in the main paper, except that four-week ahead realized oil volatility is not shown; this is because the oil volatility variable looks back 30 trading days, and would thus overlap with the explanatory variables if used as a four-week ahead dependent variable.

5. Bootstrapping methodology

We need to control for two deviations from standard assumptions. First, we likely have serial correlation in the residuals of our time-series regressions because of overlapping observations. Second, we use forward selection for choosing a parsimonious set of in-sample regressors. Both of these considerations may introduce upward bias in the R-squareds, and downward bias in the standard errors. To control for both of these sources of finite sample bias, we bootstrap the data and construct bootstrapped distributions for our t-statistics and R-squareds.

We first detrend all dependent and forecasting variables. We then residualize each dependent variable by regressing out its lagged four-week version. Our in-sample analysis assumes the following specification for the detrended and residualized series:

$$y_{t:t+h} = X_t^{(M)}\beta + \epsilon_{t:t+h}, \quad (1)$$

where $X_t^{(M)}, \beta \in R^M$, M is the number of chosen explanatory variables, and the time index t is in weeks. We assume the $X_t^{(M)}$s are chosen from the larger set X_t of $N > M$ variables using forward selection. Under the null we assume that $\beta = 0$ for all M variable subsets. To match the empirical properties of the data, we estimate an $AR(K)$ model for the dependent variable:

$$y_{t:t+h} = b_0 + b_1 y_{t-1:t-1+h} + \cdots + b_K y_{t-K:t-K+h} + e_{t:t+h}. \quad (2)$$

We run the analysis with $K = h$, i.e., eight lags for eight-week ahead forecasts, and four lags for four-week ahead forecasts. We use the above model to find the empirically estimated coefficients, $\hat{b}_0, \ldots, \hat{b}_K$.

As pointed out in Stambaugh (1999), $y_{t:t+h}$ may be correlated with regressors at $t+h$ if $y_{t:t+h}$ is a return and time $t+h$ regressors are price-based variables (e.g., book-to-market ratios) which adversely impacts the OLS properties of (2). To account for the Stambaugh bias in our simulations, we first identify the five most correlated independent variables with $e_{t:t+h}$ in (2). Next, we regress $e_{t:t+h}$ on these five independent variables, $x_{t:t+h}^{(1)}$, …, $x_{t:t+h}^{(5)}$:

$$e_{t:t+h} = c_0 + c_1 x_{t:t+h}^{(1)} + \cdots + c_5 x_{t:t+h}^{(5)} + e'_{t:t+h}. \quad (3)$$

From this regression, we estimate $\widehat{e'}_{t:t+h}$ and then measure the innovation variance $var(\widehat{e'})$. As can be seen from the following table, the R-squareds from this projection are all quite small, suggesting that the correlation between residuals in (2) and future regressors, i.e., $x_{t:t+h}^{(i)}$, is low and the impact of the Stambaugh bias is unlikely to be large (the result for 4-week ahead change in oil volatility is not available because we do not include this in our main specification due to data overlap issues as explained in the paper). Nevertheless, we control for the Stambaugh bias in our simulations, as detailed below.

R-squareds from regression (3)								
	FutRet	*Dspot*	*DOilVol*	*xomRet*	*bpRet*	*rdsaRet*	*DInv*	*DProd*
h=4	0.02	0.02	N/A	0.03	0.02	0.02	0.02	0.01
h=8	0.01	0.01	0.04	0.03	0.02	0.02	0.01	0.01

A single simulation run proceeds as follows:

1. Set $y_{0:h} = 0$
2. For the 100-step burn-in period, i.e., $k \in [1,100]$, we draw $e_{k:k+h}$ from a normal distribution with mean zero and variance $var(\widehat{e'})$.
3. Use these innovations to recursively generate elements $y_{k:k+h}$ for $k \in [1,100]$ using (2). At step 101 begin collecting data.
4. We use $\hat{e}_{t:t+h}$ from (3) when simulating *y* variables after the burn-in period. We calculate $\hat{e}_{t:t+h}$ by drawing $e'_{t:t+h}$ from a normal distribution with variance $var(\widehat{e'})$, and then adding to that draw the actual *t:t+h* values of the five selected *x* variables as in (3) using the estimated $c_0, c_1, \dots, c_5$ coefficients.
5. Collect the y variables until we match the number of empirical observations.
6. Run the forward selection algorithm using the simulated y's and the detrended and residualized X's. This selects a subset $X^{(M)}$ of explanatory variables. Note that the X's are not bootstrapped – these are the actual explanatory variables for each simulation run.
7. Keep track of the adjusted R-squared of this simulation run.
8. Keep track of the standard (no adjustments) OLS t-statistic for each of the variables that are selected by the forward selection algorithm. In every simulated path this will result in *M* t-statistics, $\{\hat{t}_1, \dots, \hat{t}_M\}$. Here $\hat{t}_1$ correspond to the t-statistic of the first variable chosen by the forward selection algorithm, $\hat{t}_2$ is the t-statistics of the second chosen variable, and so on. We refer to these as the *ordered t-statistics*.

We repeat this procedure 1,000 times to generate a distribution for the observed R-squareds and the observed t-statistics. The simulated R-squareds and ordered t-statistics adjust for overlapping observations and variable selection under the null hypothesis of no relationship between the dependent and the independent variables.

We evaluate the adjusted R-squared for a given dependent variable via the percentage of simulated adjusted R-squareds that are lower. Since this is a one-sided test, a value of

above 95% indicates significance at the 5% level. For p-values of the coefficient estimates in the actual regression, we compare the t-statistic of the n^{th} chosen variable in the forward selection method to the n^{th} ordered t-statistic distribution under the null hypothesis. We report the outcome of the two-sided test $\min(\hat{p}, 1-\hat{p})$ where $\hat{p}$ is the number of simulated t-statistics for the n^{th} chosen variable that are less than the t-statistic for the actual n^{th} chosen variable. Because this is a two-sided test, a p-value of 5% would indicate significance at the 10% level.

For purposes of this test, all t-statistics are calculated using standard OLS assumptions of independence and homoscedasticity. The simulated t-statistic distributions will reflect all of the OLS biases.

6. References

Deutsche Bank Markets Research. 2013. "Oil & Gas for Beginners: A Guide to the Oil & Gas Industry."

Devold, Havard. 2013. "Oil and gas production handbook: An introduction to oil and gas production, transport, refining and petrochemical industry," ABB Oil & Gas.

Downey, Morgan. 2009. "Oil 101," Los Angeles: Wooden Table Press.

Griffin, James M., and Henry B. Steele. 1986. "Energy Economics and Policy," Orlando: Academic Press.

Griffin, James M., and David J. Teece. 1982. "OPEC Behavior and World Oil Prices," London: George Allen & Unwin.

International Energy Agency. Oil Market Report Glossary available at https://www.iea.org/oilmarketreport/glossary/

Raymond, Martin, and William L. Leffler. 2006. "Oil and Gas Production in Nontechnical Language," Tulsa: PennWell.

S&P Global Platts Glossary available at https://www.platts.com/glossary.

Stambaugh, R. F. 1999. "Predictive regressions," *J. Financial Economics*, 54 (3), 375-421.Yergin, Daniel. 1992. "The Prize: The Epic Quest for Oil, Money & Power," NY: Free Press.

Yergin, Daniel. 2011. "The Quest: Energy, Security, and the Remaking of the Modern World," New York: Penguin Group.

7. Figures and Tables

Figure A.1. Clustered Correlation Plot of all Independent Variable Series. All the series are detrended before covariance calculation. The clustering adopts hierarchy method, implemented by Python scipy.clustering.hierarchy method. This plot shows the covariance clusters of the independent variable series constructed every Friday. The color bar on the right indicates the magnitude and sign of the value.

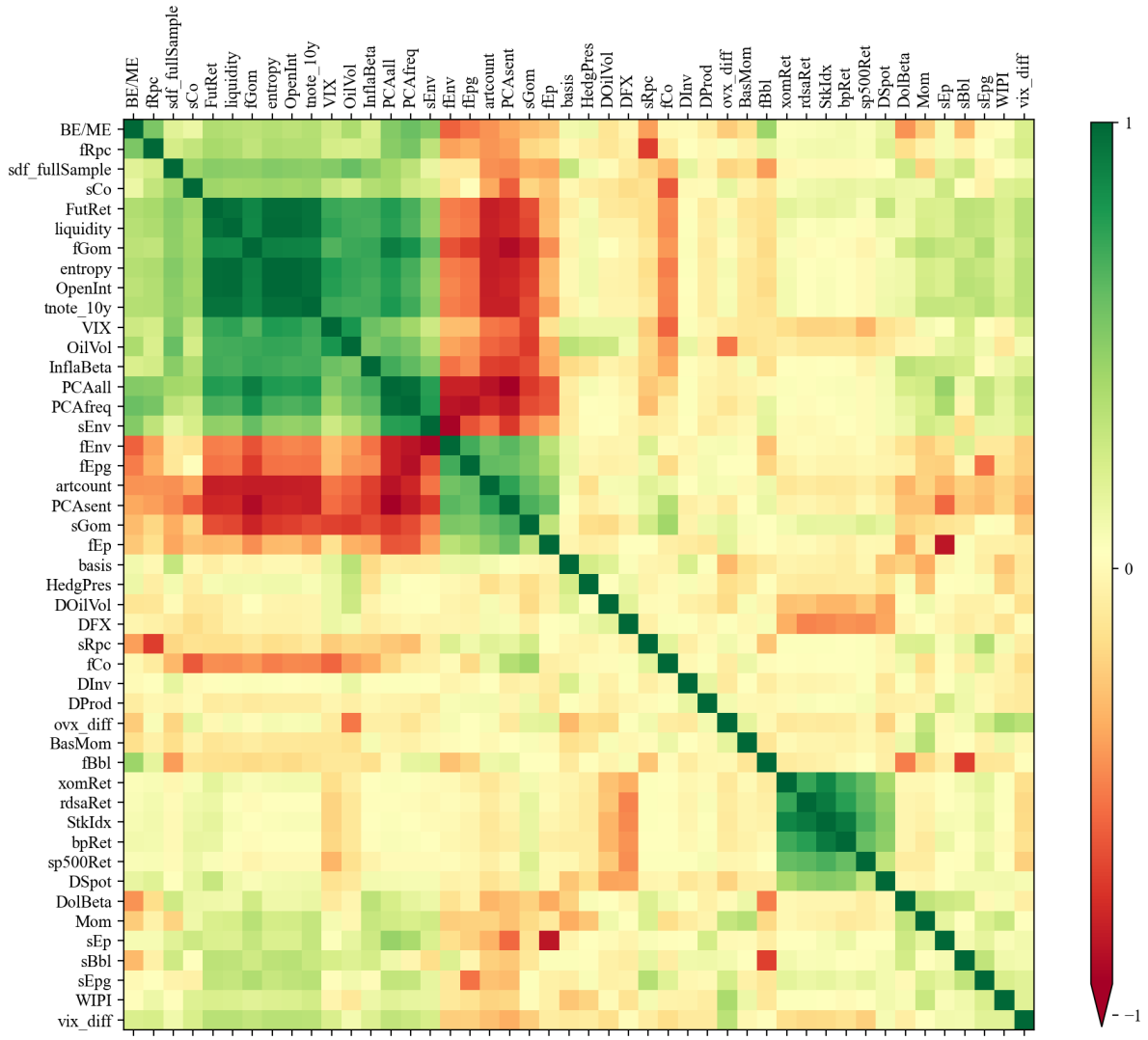

Figure A.2a. Time series plots of selected non-text variables in OOS 2-0 Lasso Model. Selected non-text variables in the out-of-sample R^2 based 5-year lasso updating model. This figure shows the time variation of the non-text variables composing the selected prediction model for *FutRet, xomRet, bpRet, and rdsaRet* in the whole data period. The y-axis lists the variables that at least enter the prediction model once during the whole selection process; the x-axis denotes the time of each forecast point. A blue block indicates the corresponding variable is selected in that prediction window.

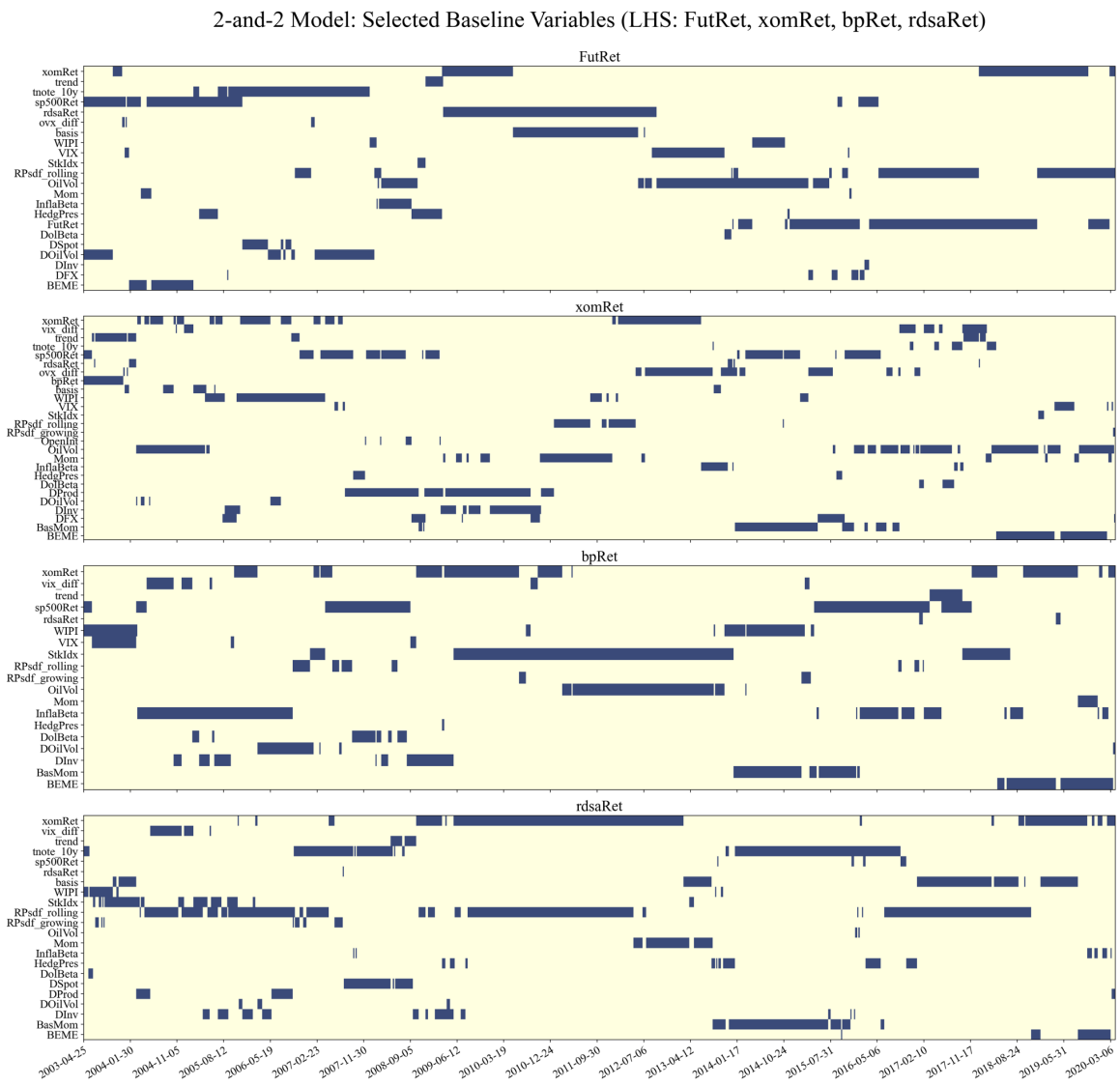

Figure A.2b. Time series plots of selected non-text variables in OOS 2-0 Lasso Model. Selected non-text variables in the out-of-sample R^2 based 5-year lasso updating model. This figure shows the time variation of the non-text variables composing the selected prediction model for *DSpot, DOilVol, DInv, and DProd* in the whole data period. The y-axis lists the variables that at least enter the prediction model once during the whole selection process; the x-axis denotes the time of each forecast point. A blue block indicates the corresponding variable is selected in that prediction window.

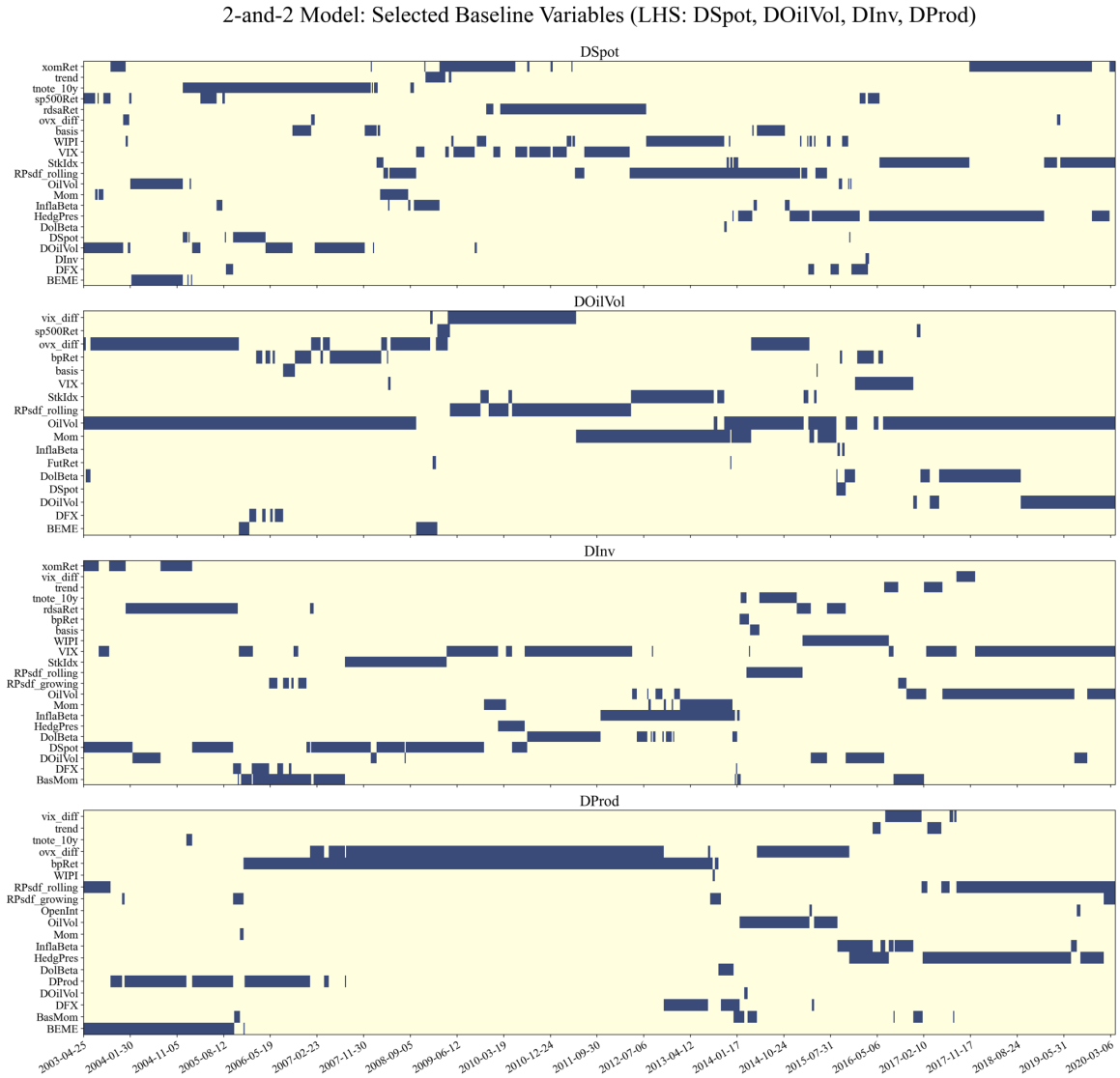

Figure A.2c. Time series plots of selected textual variables in OOS 0-2 Lasso Model. Selected text variables in the out-of-sample R^2 based 5-year lasso updating model. This figure shows the time variation of the text variables composing the selected prediction model for *FutRet, xomRet, bpRet, and rdsaRet* in the whole data period. The y-axis lists the variables that at least enter the prediction model once during the whole selection process; the x-axis denotes the time of each forecast point. A blue block indicates the corresponding variable is selected in that prediction window.

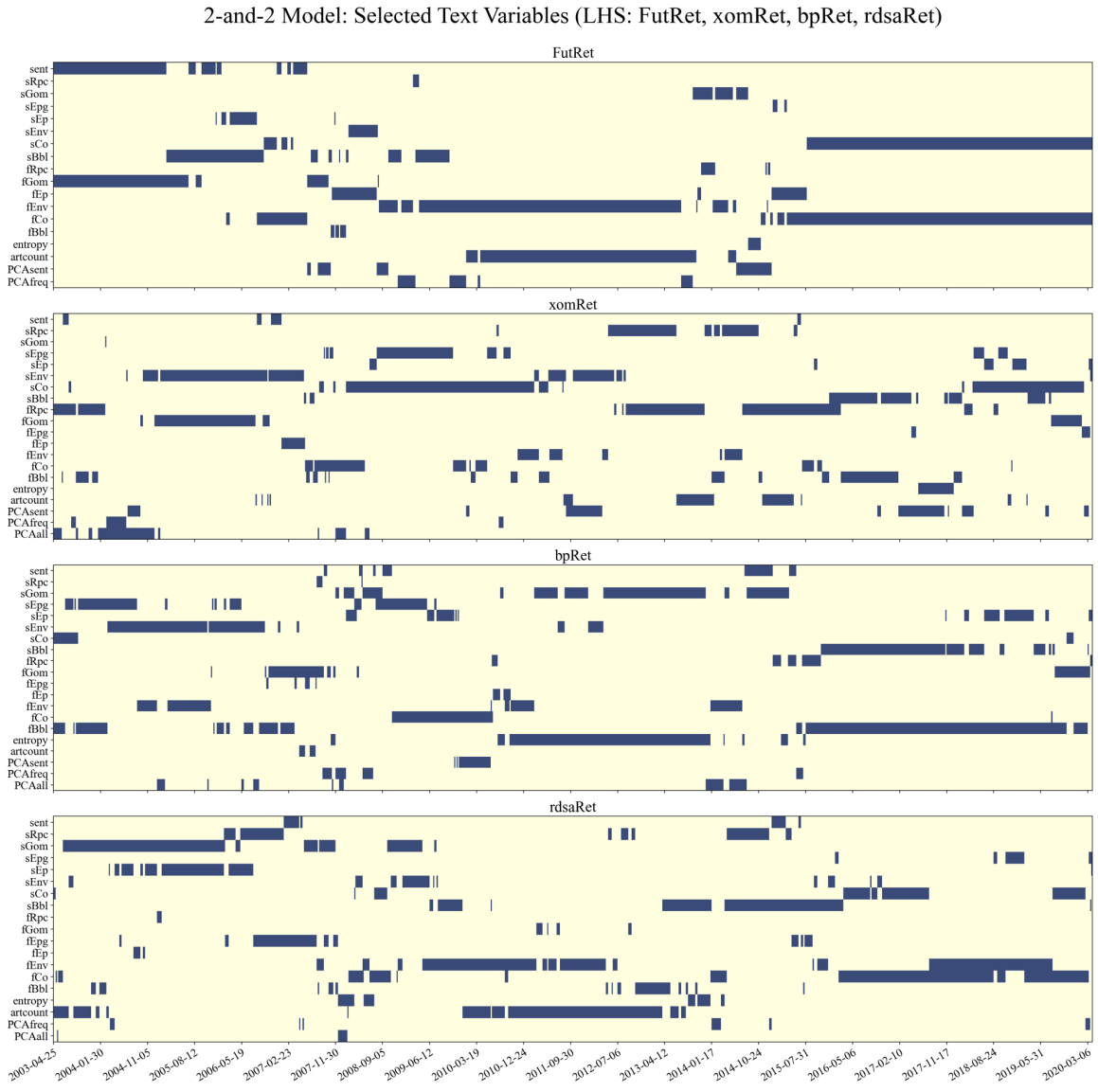

Figure A.2d. Time series plots of selected textual variables in OOS 0-2 Lasso Model. Selected text variables in the out-of-sample R^2 based 5-year lasso updating model. This figure shows the time variation of the text variables composing the selected prediction model for *DSpot, DOilVol, DInv, and DProd* in the whole data period. The y-axis lists the variables that at least enter the prediction model once during the whole selection process; the x-axis denotes the time of each forecast point. A blue block indicates the corresponding variable is selected in that prediction window.

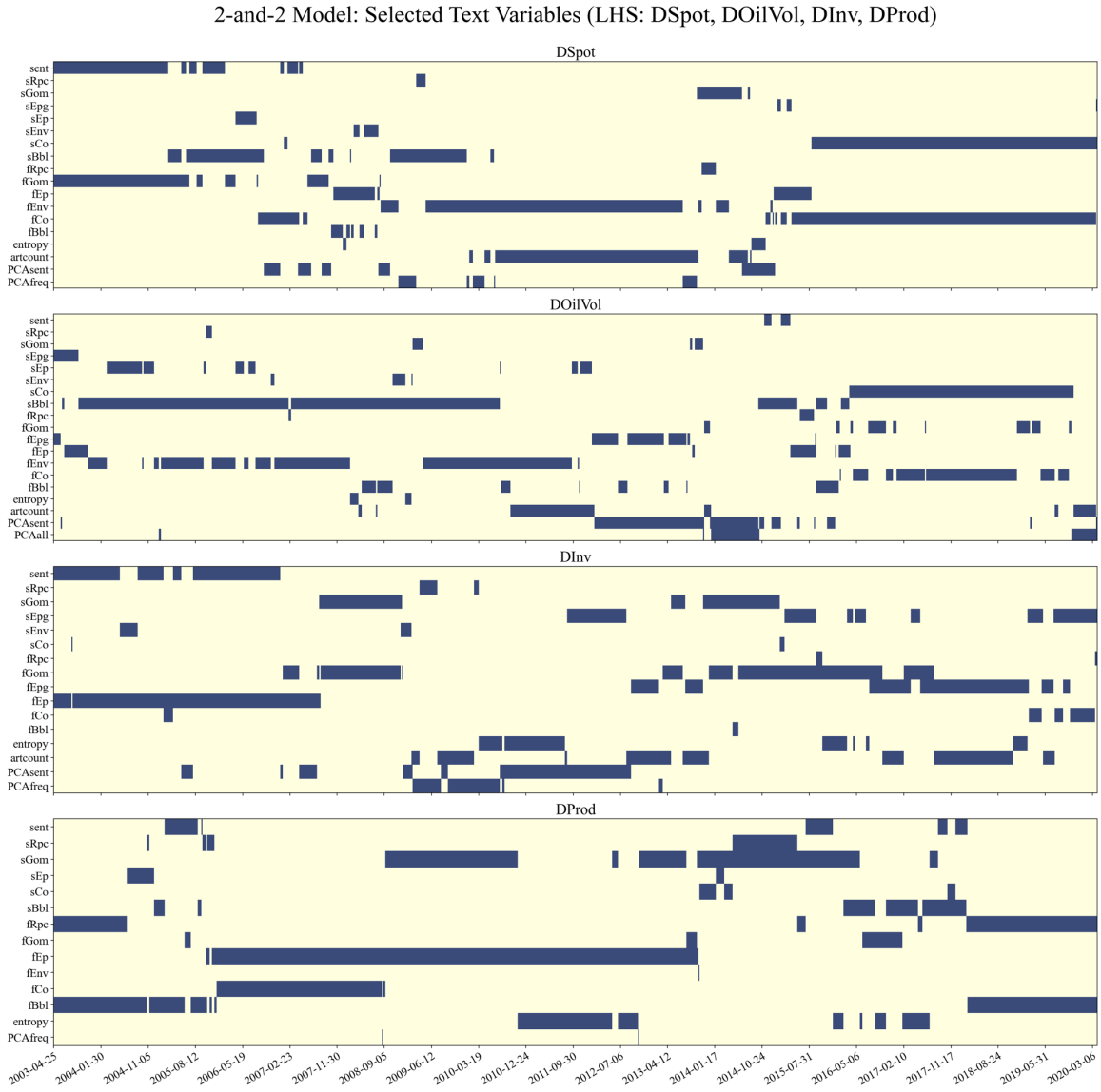

Table A.I

Reuters Topic Codes for Energy Related Article Selection

The table below shows the 98 Reuters Topic Codes we use in the article selection phase in constructing our corpus for textual analysis. We construct the list manually by checking whether the definition of each code captures energy-related news. The article selection is then accomplished by examining if any topic code assigned to a certain article is contained in the following list.

NRG	HOIL	RENE
CO2	JET	RENQ
COA	PWR	RENF
COAPWR	COAPWR	ASIAE
NGS	GASPWR	NASIAE
PETC	NUCPWR	SEASIE
GASPWR	OILPWR	PACE
LNG	RNWPWR	SASIAE
NGL	RNW	AMERSE
SHL	BIOPWR	CAMERE
PLAS	GEOPWR	NAMERE
CRU	HYDPWR	ERCOT1
PROD	SOLPWR	FRCC1
OPEC	TWPWR	MRO1
AMCRU	WINPWR	NPCC1
ASCRU	BIOF	RFC1
EMACRU	BIODSL	SERC1
RUSCRU	BIOETH	SPP1
CANCRU	BIOMS	WECC1
LATCRU	BIOCEL	PADD1
USCRU	REF	PADD3
AUSCRU	PPL	PADD2
NASCRU	PWRL	PADD4
SEACRU	ENER	PADD5
AFRCRU	COAM	SAMERE
MECRU	OILG	EMEAE
NSCRU	OILI	CEEUE
RFO	EXPL	FSUE
MOG	REFI	NWEE
LPG1	ENEQ	MEASTE
DISTLL	DRIL	OAFRE
NAP	OGTR	MEDNAE
BUN	OILQ	

Table A.II: Word List of Louvain Clustered Topics

This table shows the top 20 tokens by frequency in each topical group.

Topic	Word List
global oil market (gom)	oil (4,136,780), barrel (1,226,580), brent (526,719), refin (411,872), crude.oil (409,276), opec (394,754), petroleum (293,525), heat (291,997), diesel (276,319), barg (194,018), ipe (175,841), distil (167,863), tanker (142,160), sulphur (140,039), gallon (136,243), eia (127,622), nwe (70,962), ara (62,293), energi.inform.administr (55,927), bunker (47,736)
energy/power generation (epg)	gas (2,082,748), energi (1,385,165), coal (510,535), outag (409,463), nuclear (381,919), electr (326,305), generat (225,899), equiti (184,324), mine (178,868), lead (165,664), lng (162,184), addit (142,116), reactor (125,164), renew (120,903), solar (101,509), case (91,068), miner (90,722), grid (79,484), hydro (69,220), power.generat (53,787)
company (co)	fuel (1,483,081), bp (369,851), shell (369,655), vitol (237,656), mrpl (220,506), hsfo (158,878), glencor (144,515), exxon (136,651), mop (121,139), hin.leong (113,240), ceypetco (102,883), chevron (96,915), bpcl (95,996), petrochina (93,576), bapco (92,908), essar (90,448), blend (88,597), pertamina (84,403), trafigura (83,198), forti (81,329)
crude oil physical (bbl)	pipelin (409,704), wti (321,512), lls (169,911), wts (117,949), gulf.coast (68,858), cush (53,943), west.texa.intermedi (35,191), bakken (31,987), heavi.louisiana.sweet (22,581), enbridg (18,568), midstream (17,634), permian (13,138), sunoco (12,958), heavi.crude (9,681), lighter (8,541), heavi.oil (8,333), eagl.ford (8,053), suncor.energi (7,419), occident.petroleum (5,411), permian.basin (5,366)
Environment (env)	emiss (189,038), carbon (176,792), climat (105,968), environ (61,429), green (49,666), climat.chang (46,992), pollut (45,532), biofuel (32,514), carbon.dioxid (24,075), epa (22,403), biodiesel (19,407), global.warm (19,067), fossil (18,012), valv (10,182), kyoto.protocol (9,235), environment.protect.agenc (8,036), methan (7,179), emiss.trade.scheme (6,951), alki (6,204), air.pollut (4,723)

exploration & production (ep)	explor (148,206), drill (137,958), offshor (123,543), rig (94,639), shale (58,500), gulf.mexico (52,649), spill (46,891), royal.dutch.shell (37,685), onshor (28,528), pemex (26,894), explor.product (23,701), upstream (23,476), downstream (21,409), baker.hugh (17,968), deepwat (17,860), extract (17,115), halliburton (11,329), texaco (10,093), frack (9,383), transocean (9,373)
refining & petrochemicals (rpc)	reform (110,766), petrochem (88,297), cement (22,637), lpg (20,345), feedstock (18,355), propan (18,259), crude.distil.unit (12,005), netback (7,888), butan (7,407), liquefi.petroleum.gas (6,682), octan (6,045), fluid.catalyt.cracker (5,842), ethan (5,737), visbreak (5,079), olefin (4,370), oxygen (3,418), benzen (2,738), tertiari (2,081), polym (2,075), urea (1,830)

Table A.III: Similarity of Topical Definitions for Louvain Clustering and Latent Dirichlet Allocation

This table displays the comparison of topic modeling on the energy corpus using LDA and Louvain Clustering methods. Panel A displays the document topic distribution of each Louvain topics. The probability distribution of a certain topic is calculated by $\sum_{\mathrm{j}} x_{ij} / \sum_{\mathrm{i}} \sum_{j} x_{ij}$, where $i = 1,2, \ldots, 7$ indicates different topics, $j = 1,2, \ldots, N_i$ labels N_i different words within topic i, and x_{ij} represents the total occurrence of word *ij* in the whole corpus. Boldface indicates probability densities that are greater than 5%. Panel B displays the highest cosine similarity between a Louvain topic and the LDA topics within each trial of the LDA analysis. The cosine similarity is calculated by $(p_i^T \times q_j) / (norm(p_i) \times norm(q_j))$, where p_i, $i = 1,2, \ldots, 7$ is the probability distribution vector of a Louvain topic, and q_j, $j = 1,2, \ldots, 7$ is the probability distribution vector of a LDA topic in a certain trial. Each column corresponds to a Louvain topic, and each row per trial. The variation of cosine similarity results from the indeterminacy of the LDA outcomes, which allocate topics differently across trails. Cosine similarities greater than 80% are in boldface.

Panel III. 1: Document Topic Distribution of Louvain Topics

Topic	Co	Gom	Env	Epg	Bbl	Rpc	Ep
Distribution	**0.1911**	**0.3913**	0.0331	**0.2832**	0.0489	0.0136	0.0388

Panel III. 2: Greatest Cosine Similarities between Louvain & LDA Topics (10 trials)

Trials \ Topic	Co	Gom	Env	Epg	Bbl	Rpc	Ep
Trial 1	0.9037	0.9795	0.7078	0.7967	0.2610	0.1783	0.3354
Trial 2	0.7970	0.9864	0.5505	0.8744	0.2835	0.1346	0.4524
Trial 3	0.8952	0.9934	0.3637	0.8036	0.3697	0.1704	0.2601
Trial 4	0.8756	0.9876	0.1281	0.9299	0.6520	0.2642	0.4190
Trial 5	0.8769	0.9318	0.5228	0.8331	0.5097	0.0731	0.1024
Trial 6	0.9401	0.9719	0.6575	0.8346	0.3739	0.2868	0.1160
Trial 7	0.9019	0.9765	0.3547	0.8148	0.3310	0.0847	0.0815
Trial 8	0.9084	0.9774	0.5595	0.8242	0.3597	0.0506	0.1227
Trial 9	0.9014	0.9919	0.3500	0.8397	0.5550	0.0828	0.3745
Trial 10	0.8808	0.9894	0.3921	0.8042	0.2374	0.1094	0.1218
Average	**0.8881**	**0.9786**	0.4587	**0.8355**	0.3933	0.1435	0.2386

Table A.IV: Count of Selected and Statistically Significant Variables of Forward Selection Model

This table displays in Panel A the count of text, non-text and all predictors that were selected in the stepwise forward selection at the eight-week horizon, and in Panel B the count of text, non-text and all predictors that were selected and statistically significant. The counts are based on the regression results shown in Table IV. The total numbers of text variables and non-text variables considered as regressors in the forward selection model in Table IV are 19 and 22, respectively.

Panel A: Count of selected variables

Specification	FutRet	DSpot	DOilVol	xomRet	bpRet	rdsaRet	DInv	DProd	All
Text	5	4	3	5	3	4	5	3	32
Nontext	2	3	4	2	4	3	2	4	24
All	7	7	7	7	7	7	7	7	56

Panel B: Count of selected and significant Variables

Specification	FutRet	DSpot	DOilVol	xomRet	bpRet	rdsaRet	DInv	DProd	All
Text	4	3	3	0	2	1	3	0	16
Nontext	1	1	4	0	3	1	0	0	10
All	5	4	7	0	5	2	3	0	26

Table A.V: F-test for the Eight-Week Stepwise Forward Selection

The table shows the F-statistics for the eight-week stepwise forward selection model. We test the joint significance of the text variables, the non-text variables, and all the (text + non-text) variables separately by conducting an F-test in the regressions. We regress each dependent variable on eight predictors with Newey-West standard errors allowing for a lag of up to six. Seven of the predictors are selected in the stepwise forward selection process, and the other independent variable is the four-week version of the lagged dependent variable. All predictors are lagged by eight weeks.

	FutRet	DSpot	DOilVol	xomRet	bpRet	rdsaRet	DInv	DProd
text	6.79***	7.54***	7.47***	2.23**	12.42***	5.90***	9.14***	7.21***
non-text	3.81**	5.70***	39.40***	3.28**	5.82***	2.73**	1.05	2.22*
all	5.20***	5.43***	26.19***	2.12**	7.32***	4.73***	7.74***	3.54***

* *p<0.1* ** *p<0.05* *** *p<0.01*

Table A.VI

Stepwise Forward Selection at the Eight-Week Horizon with Risk Premium Measures

The tables below show stepwise forward selection results per dependent variable from Table IV in the Baseline column, with the addition of each risk measure (VIX, vix_diff, ovx_diff, and sdf_fullSample) independently after the 7 variables were selected by forward selection. If one of VIX and vix_diff was picked up by forward selection, we leave the regression as is (which explains why in some cases the Baseline and one of the VIX or vix_diff columns match), and move on to add ovx_diff and sdf_fullSample separately since neither was a candidate variable in forward selection. We report heteroskedasticity and autocorrelation consistent (HAC) standard errors.

Panel VI.1 Predicting eight-weeks ahead Oil Futures Returns (FutRet)

Predictors	*Baseline*	*VIX*	*vix_diff*	*ovx_diff*	*sdf_full Sample*
(Intercept)	0.05	-0.01	-0.01	0.27	-0.19
fRpc	0.92	0.83	0.78	-1.15	1.20
basis	4.68 *	4.41	4.35	6.62 *	6.10 **
entropy	61.40 ***	61.60 ***	61.07 ***	75.52 ***	64.74 ***
sGom	5.89 **	5.79 ***	5.91 ***	10.59 **	5.62 **
PCAall	3.14 ***	3.11 ***	3.18 ***	3.73 **	2.99 ***
Mom	-0.08 *	-0.08 *	-0.08 **	-0.09	-0.09 **
fBbl	-1.65 *	-1.75 *	-1.77 *	-0.84	-1.87 *
VIX		-0.03			
vix_diff			0.05		
ovx_diff				-0.21	
sdf_fullSample					-21.37
Observations	1132	1135	1135	665	1045
R^2 / R^2 adjusted	0.155 / 0.150	0.153 / 0.147	0.153 / 0.147	0.204 / 0.194	0.171 / 0.165

** p<0.1 ** p<0.05 *** p<0.01*

Panel VI.2 Predicting eight-weeks ahead Spot Price Changes (DSpot)

Predictors	*Baseline*	*VIX*	*vix_diff*	*ovx_diff*	*sdf_full Sample*
(Intercept)	0.06	0.06	0.01	0.20	0.00
basis	10.15 ***	10.15 ***	9.78 ***	10.88 ***	11.35 ***
fRpc	0.95	0.95	0.82	-0.87	1.46
entropy	78.17 ***	78.17 ***	78.80 ***	88.34 ***	79.00 ***
VIX	-0.09	-0.09	-0.08	-0.02	-0.11
sGom	6.37 ***	6.37 ***	6.49 ***	11.59 ***	5.89 ***
PCAall	2.97 ***	2.97 ***	3.05 ***	3.64 **	2.58 **
Mom	-0.07	-0.07	-0.07	-0.09	-0.08
vix_diff			-0.05		
ovx_diff				-0.18	
sdf_fullSample					2.92
Observations	1132	1132	1135	665	1045
R^2 / R^2 adjusted	0.176 / 0.171	0.176 / 0.171	0.173 / 0.167	0.224 / 0.215	0.195 / 0.189

** p<0.1 ** p<0.05 *** p<0.01*

Panel VI.3 Predicting eight-weeks ahead Oil Volatility Changes (DOilVol)

Predictors	*Baseline*	*VIX*	*vix_diff*	*ovx_diff*	*sdf_full Sample*
(Intercept)	0.01	0.01	0.10	0.72	0.18
OilVol	-0.67 ***	-0.67 ***	-0.66 ***	-0.40 ***	-0.62 ***
DSpot	-0.35 ***	-0.35 ***	-0.37 ***	-0.42 ***	-0.36 ***
VIX	0.50 ***	0.50 ***	0.49 ***	0.22	0.53 ***
fGom	10.51 ***	10.51 ***	10.09 ***	12.37 **	10.37 ***
entropy	-77.73 **	-77.73 **	-78.61 **	-76.05 **	-81.50 **
WIPI_8wk	-2.99 **	-2.99 **	-3.02 **	-3.85 ***	-2.37 *
fCo	-2.37	-2.37	-2.50	-2.38	-2.50
vix_diff			-0.05		
ovx_diff				0.53 ***	
sdf_fullSample					-32.09
Observations	1132	1132	1135	665	1045
R^2 / R^2 adjusted	0.362 / 0.358	0.362 / 0.358	0.356 / 0.352	0.425 / 0.418	0.354 / 0.349

** p<0.1 ** p<0.05 *** p<0.01*

Panel VI.4 Predicting eight-weeks ahead Exxon Stock Returns (xomRet)

Predictors	*Baseline*	*VIX*	*vix_diff*	*ovx_diff*	*sdf_full Sample*
(Intercept)	-0.01	-0.01	-0.01	-0.21	-0.00
fBbl	-1.21	-1.20	-1.20	-0.75	-1.73 *
DInv	0.75 *	0.75 *	0.70 *	0.60	0.68
sEp	-0.74	-0.68	-0.72	-1.03 *	-0.62
sEpg	0.90 *	1.00 *	1.00 *	0.70	0.79
BE/ME	2.55 *	2.60 *	2.68 *	2.94 **	3.00 **
entropy	17.40 *	17.56	18.64	20.24	15.91 *
sRpc	0.69	0.71	0.67	-0.11	0.50
VIX		0.04			
vix_diff			-0.12		
ovx_diff				-0.01	
sdf_fullSample					-24.09 ***
Observations	1130	1133	1133	665	1043
R^2 / R^2 adjusted	0.072 / 0.067	0.074 / 0.067	0.077 / 0.070	0.080 / 0.068	0.080 / 0.073

** p<0.1 ** p<0.05 *** p<0.01*

Panel VI.5 Predicting eight-weeks ahead BP Stock Returns (bpRet)

Predictors	*Baseline*	*VIX*	*vix_diff*	*ovx_diff*	*sdf_full Sample*
(Intercept)	0.01	0.01	0.01	-0.27	0.02
Mom	-0.09 ***	-0.09 ***	-0.09 ***	-0.12 ***	-0.08 **
sEnv	2.77 ***	2.90 ***	2.77 ***	2.54 ***	2.62 ***
entropy	34.32 **	33.08 **	34.32 **	42.35 **	33.22 **
sGom	3.29 ***	3.45 **	3.29 ***	4.43 *	3.03 **
vix_diff	-0.31 **	-0.29 *	-0.31 **	-0.42 ***	-0.29 *
basis	-4.89 **	-5.23 **	-4.89 **	-6.74 **	-4.56 **
tnote_10y	-2.11 ***	-1.99 ***	-2.11 ***	-1.98 *	-1.97 ***
VIX		0.06			
ovx_diff				-0.10	
sdf_fullSample					9.94
Observations	1130	1133	1130	665	1043
R^2 / R^2 adjusted	0.148 / 0.142	0.150 / 0.144	0.148 / 0.142	0.198 / 0.188	0.127 / 0.120

** p<0.1 ** p<0.05 *** p<0.01*

Panel VI.6 Predicting eight-weeks ahead Royal Dutch Shell Stock Returns (rdsaRet)

Predictors	*Baseline*	*VIX*	*vix_diff*	*ovx_diff*	*sdf_full Sample*
(Intercept)	0.02	0.01	0.00	-0.28	0.24
entropy	67.28 **	71.01 **	69.83 **	68.19 *	67.06 **
sGom	2.79 ***	3.25 ***	2.96 ***	4.98 ***	2.37 **
fCo	1.82	2.47	1.88	1.25	1.85
Mom	-0.04	-0.03	-0.04	-0.07 ***	-0.02
DInv	1.00 **	0.97 **	0.88 **	0.52	0.80
sEnv	1.49	1.80 *	1.61 *	1.98 *	1.61
InflaBeta	-0.34	-0.38	-0.35	-0.60	-0.52 *
VIX		0.17 **			
vix_diff			-0.19 **		
ovx_diff				-0.09	
sdf_fullSample					29.34
Observations	1132	1135	1135	665	1045
R^2 / R^2 adjusted	0.129 / 0.123	0.142 / 0.136	0.135 / 0.129	0.179 / 0.169	0.133 / 0.126

** p<0.1 ** p<0.05 *** p<0.01*

Panel VI.7 Predicting eight-weeks ahead Oil Inventory Changes (DInv)

Predictors	*Baseline*	*VIX*	*vix_diff*	*ovx_diff*	*sdf_full Sample*
(Intercept)	0.01	0.01	0.01	0.04	0.04
entropy	-7.71 ***	-7.72 ***	-7.77 ***	-9.80 ***	-8.73 ***
fRpc	-0.44 ***	-0.43 ***	-0.46 ***	-0.34 **	-0.42 ***
sGom	-0.24	-0.20	-0.25	-0.64 *	-0.27
sEp	-0.61 ***	-0.62 ***	-0.63 ***	-0.55 **	-0.57 ***
fEp	-0.63 ***	-0.66 ***	-0.64 ***	-0.43	-0.53 ***
tnote_10y	-0.31 *	-0.28	-0.30 *	-0.10	-0.21
BasMom	0.04 *	0.05 *	0.04 *	0.04	0.05 *
VIX		0.01			
vix_diff			0.02		
ovx_diff				-0.00	
sdf_fullSample					4.48
Observations	1132	1135	1135	664	1045
R^2 / R^2 adjusted	0.182 / 0.177	0.187 / 0.181	0.186 / 0.180	0.205 / 0.196	0.182 / 0.175

** p<0.1 ** p<0.05 *** p<0.01*

Panel VI.8 Predicting eight-weeks ahead Oil Production Changes (DProd)

Predictors	*Baseline*	*VIX*	*vix_diff*	*ovx_diff*	*sdf_full Sample*
(Intercept)	0.01	0.01	0.01	-0.22	0.01
VIX	0.05	0.05	0.05	0.09 *	0.05
fGom	-1.02	-1.02	-1.02	-3.16 **	-1.04
DSpot	-0.02	-0.02	-0.02	-0.01	-0.02
InflaBeta	0.08 *	0.08 *	0.08 *	0.15	0.06
sRpc	0.20	0.20	0.20	0.01	0.24
fBbl	0.47 ***	0.47 ***	0.47 ***	0.72 ***	0.45 ***
BE/ME	-0.84 **	-0.84 **	-0.84 **	-0.46	-0.72
vix_diff			-0.00		
ovx_diff				-0.02	
sdf_fullSample					-1.54
Observations	1132	1132	1132	664	1045
R^2 / R^2 adjusted	0.067 / 0.061	0.067 / 0.061	0.067 / 0.060	0.185 / 0.175	0.069 / 0.061

** p<0.1 ** p<0.05 *** p<0.01*

Table A.VII

Subperiod stepwise forward selection at the eight-week horizon

The table shows the subperiod regression results for all eight dependent variables at the eight-week horizon using stepwise forward selection to choose 7 of all the variables described in Table I, except ovx_diff and sdf_fullSample. All dependent and independent variables are first detrended and residualized with respect to the lagged four-week dependent variable. Only predictors that were chosen are included in this table and are listed in order of priority in forward selection. Coefficients are standardized using the ratios of the standard deviations of the dependent and predicting variables. There are nine subperiods: 1998-04-01 – 1999-11-30, 1999-12-01 – 2002-07-31, 2002-08-01 – 2005-03-31, 2005-04-01 – 2007-11-30, 2007-12-01 – 2009-06-30, 2009-07-01 – 2012-02-29, 2012-03-01 – 2014-10-31, 2014-11-01 – 2017-06-30, and 2017-07-01 – 2020-03-31, defined as follows. We used NBER recession dating for the period 2007-12-01 to 2009-06-30; then we define post-crisis subperiods of roughly equal (32-month) length. The three pre-crisis subperiods that precede the crisis are also of 32-month length, while the initial (residual) subperiod is 20 months.

Panel VII. 1: Predicting 8-weeks ahead Oil Futures Returns (*FutRet*)

Subperiod 1: 1998-04-01 - 1999-11-30			**Subperiod 2: 1999-12-01 - 2002-07-31**			**Subperiod 3: 2002-08-01 - 2005-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
HedgPres	0.33	0.394	sRpc	-0.28	0.312	PCAfreq	**0.71**	0.007
DFX	0.25	0.498	sEpg	-0.31	0.181	HedgPres	0.11	0.344
BE/ME	0.2	0.455	VIX	-0.34	0.103	fEpg	0.38	0.082
PCAsent	-0.39	0.144	sGom	0.25	0.291	OilVol	0.3	0.077
DolBeta	-0.41	0.11	WIPI_8wk	0.25	0.172	liquidity	-0.19	0.353
DOilVol	-0.28	0.133	InflaBeta	-0.17	0.36	Mom	**-0.45**	0.008
entropy	0.21	0.27	basis	0.16	0.321	InflaBeta	**-0.34**	0.005
Observations	71		Observations	131		Observations	131	
R2 / R2 adj.	0.8/0.78		R2 / R2 adj.	0.59/0.57		R2 / R2 adj.	0.6/0.58	
Mean of sim. Adj. R2	0.702		Mean of sim. Adj. R2	0.482		Mean of sim. Adj. R2	0.422	
CDF (%)	80.1		CDF (%)	83.3		CDF (%)	97.2	

Subperiod 4: 2005-04-01 - 2007-11-30			**Subperiod 5: 2007-12-01 - 2009-06-30**			**Subperiod 6: 2009-07-01 - 2012-02-29**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
sCo	0.31	0.347	BE/ME	**0.59**	0.023	fBbl	0.39	0.144
fEp	0.24	0.451	fBbl	-0.24	0.107	BE/ME	**0.54**	0.024
InflaBeta	0.22	0.461	fEnv	-0.22	0.108	OilVol	0.23	0.3
BE/ME	0.18	0.482	sGom	**0.56**	0	fGom	0.28	0.163
HedgPres	0.33	0.054	sEpg	-0.27	0.071	WIPI_8wk	0.19	0.286
entropy	**0.34**	0.035	tnote_10y	-0.13	0.363	DFX	0.21	0.196
fBbl	**-0.3**	0.011	DOilVol	0.11	0.189	InflaBeta	0.22	0.09
Observations	132		Observations	74		Observations	131	
R2 / R2 adj.	0.65/0.63		R2 / R2 adj.	0.91/0.9		R2 / R2 adj.	0.55/0.53	
Mean of sim. Adj. R2	0.492		Mean of sim. Adj. R2	0.593		Mean of sim. Adj. R2	0.474	
CDF (%)	95.5		CDF (%)	100		CDF (%)	71	

Subperiod 7: 2012-03-01 - 2014-10-31			**Subperiod 8: 2014-11-01 - 2017-06-30**			**Subperiod 9: 2017-07-01 - 2020-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
HedgPres	0.22	0.3	fEnv	**-0.55**	0.014	Mom	-0.2	0.459
sRpc	-0.43	0.137	basis	0.36	0.253	sCo	-0.23	0.408
basis	**0.42**	0.042	sBbl	0.37	0.054	fEpg	0.15	0.435
VIX	0.26	0.227	BasMom	-0.3	0.068	DolBeta	-0.39	0.059
BE/ME	0.39	0.072	sEp	-0.14	0.449	OilVol	**0.59**	0.026
fEpg	0.26	0.153	DInv	0.16	0.309	sRpc	-0.25	0.12
DInv	**-0.25**	0.028	BE/ME	0.17	0.183	DOilVol	**-0.33**	0.037
Observations	132		Observations	131		Observations	135	
R2 / R2 adj.	0.61/0.59		R2 / R2 adj.	0.78/0.77		R2 / R2 adj.	0.7/0.68	
Mean of sim. Adj. R2	0.462		Mean of sim. Adj. R2	0.500		Mean of sim. Adj. R2	0.534	
CDF (%)	94		CDF (%)	100		CDF (%)	96.9	

Panel VII.2: Predicting 8-weeks ahead Spot Price Changes (*DSpot*)

Subperiod 1: 1998-04-01 - 1999-11-30			**Subperiod 2: 1999-12-01 - 2002-07-31**			**Subperiod 3: 2002-08-01 - 2005-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
HedgPres	0.41	0.176	BE/ME	0.3	0.35	fRpc	0.51	0.058
DFX	0.33	0.281	fEpg	**0.83**	0.024	OilVol	0.31	0.172
DProd	-0.35	0.128	VIX	-0.24	0.281	artcount	-0.12	0.493
sBbl	0.28	0.218	PCAfreq	**0.74**	0.04	InflaBeta	**-0.6**	0.02
fEpg	-0.17	0.366	InflaBeta	-0.31	0.098	Mom	-0.45	0.077
OpenInt	-0.17	0.31	sEp	-0.26	0.1	DolBeta	0.24	0.18
WIPI 8wk	0.13	0.362	DOilVol	-0.15	0.257	StkIdx	0.17	0.103
Observations	71		Observations	131		Observations	131	
R2 / R2 adj.	0.8/0.78		R2 / R2 adj.	0.61/0.59		R2 / R2 adj.	0.63/0.61	
Mean of sim. Adj. R2	0.700		Mean of sim. Adj. R2	0.499		Mean of sim. Adj. R2	0.454	
CDF (%)	80.8		CDF (%)	84.3		CDF (%)	96.8	

Subperiod 4: 2005-04-01 - 2007-11-30			**Subperiod 5: 2007-12-01 - 2009-06-30**			**Subperiod 6: 2009-07-01 - 2012-02-29**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
sCo	0.34	0.356	BE/ME	0.37	0.147	fBbl	0.25	0.328
fEp	0.22	0.5	sEp	0.04	0.471	BE/ME	**0.49**	0.038
InflaBeta	0.22	0.499	tnote_10y	**-0.39**	0.002	DOilVol	0.35	0.105
HedgPres	0.35	0.082	sGom	**0.32**	0.012	DFX	0.27	0.146
fBbl	**-0.4**	0.037	fBbl	**-0.27**	0.012	InflaBeta	**0.37**	0.027
entropy	**0.38**	0.013	fEpg	**0.2**	0.039	artcount	-0.38	0.053
BasMom	-0.16	0.22	FutRet	-0.13	0.074	basis	0.16	0.226
Observations	132		Observations	74		Observations	131	
R2 / R2 adj.	0.69/0.67		R2 / R2 adj.	0.91/0.9		R2 / R2 adj.	0.58/0.56	
Mean of sim. Adj. R2	0.509		Mean of sim. Adj. R2	0.563		Mean of sim. Adj. R2	0.482	
CDF (%)	96.8		CDF (%)	100		CDF (%)	79.6	

Subperiod 7: 2012-03-01 - 2014-10-31			**Subperiod 8: 2014-11-01 - 2017-06-30**			**Subperiod 9: 2017-07-01 - 2020-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
HedgPres	0.2	0.348	OilVol	**0.69**	0.004	DolBeta	-0.49	0.087
sRpc	-0.43	0.146	fEnv	**-0.57**	0.002	OilVol	0.63	0.065
basis	0.41	0.071	DOilVol	**-0.39**	0.03	sRpc	-0.31	0.231
BE/ME	0.44	0.064	sBbl	0.29	0.104	DOilVol	-0.39	0.196
VIX	0.25	0.193	FutRet	0.16	0.237	sCo	-0.19	0.387
fEpg	0.27	0.129	vix_diff	0.13	0.352	DFX	0.15	0.464
DInv	**-0.25**	0.025	entropy	0.13	0.238	tnote_10y	-0.11	0.484
Observations	132		Observations	131		Observations	135	
R2 / R2 adj.	0.62/0.6		R2 / R2 adj.	0.82/0.81		R2 / R2 adj.	0.67/0.65	
Mean of sim. Adj. R2	0.478		Mean of sim. Adj. R2	0.500		Mean of sim. Adj. R2	0.544	
CDF (%)	91.8		CDF (%)	100		CDF (%)	90.8	

Panel VII. 3: Predicting 8-weeks ahead Oil Volatility Changes (*DOilVol*)

Subperiod 1: 1998-04-01 - 1999-11-30			**Subperiod 2: 1999-12-01 - 2002-07-31**			**Subperiod 3: 2002-08-01 - 2005-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
OilVol	**-0.95**	0	sEpg	**0.45**	0.037	OilVol	**-0.83**	0
fEpg	0.22	0.198	OilVol	-0.38	0.082	Mom	**0.48**	0.002
VIX	-0.26	0.287	entropy	-0.36	0.056	basis	-0.21	0.154
StkIdx	-0.2	0.293	fRpc	-0.27	0.144	sCo	0.2	0.151
basis	-0.19	0.214	artcount	0.2	0.33	entropy	**0.27**	0.021
vix_diff	0.25	0.026	sCo	0.17	0.384	sEpg	**0.31**	0.004
sp500Ret	0.18	0.07	liquidity	0.12	0.421	PCAfreq	**-0.19**	0.008
Observations	71		Observations	131		Observations	131	
R2 / R2 adj.	0.88/0.86		R2 / R2 adj.	0.63/0.61		R2 / R2 adj.	0.86/0.85	
Mean of sim. Adj. R2	0.677		Mean of sim. Adj. R2	0.466		Mean of sim. Adj. R2	0.552	
CDF (%)	99.8		CDF (%)	96.2		CDF (%)	100	

Subperiod 4: 2005-04-01 - 2007-11-30			**Subperiod 5: 2007-12-01 - 2009-06-30**			**Subperiod 6: 2009-07-01 - 2012-02-29**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
OilVol	**-0.89**	0.008	BE/ME	**-0.82**	0.004	tnote_10y	0.45	0.155
tnote_10y	-0.18	0.48	fCo	-0.35	0.119	OilVol	**-0.8**	0.001
entropy	-0.32	0.304	fEnv	-0.25	0.185	fEpg	**0.46**	0.011
DolBeta	-0.42	0.288	artcount	0.16	0.472	VIX	**0.5**	0.032
sCo	-0.41	0.157	tnote_10y	0.17	0.468	DFX	-0.29	0.066
sBbl	0.24	0.356	basis	0.12	0.434	DInv	**-0.31**	0.007
WIPI_8wk	0.17	0.216	OilVol	-0.09	0.346	sGom	**-0.27**	0.006
Observations	132		Observations	74		Observations	131	
R2 / R2 adj.	0.56/0.53		R2 / R2 adj.	0.87/0.85		R2 / R2 adj.	0.71/0.69	
Mean of sim. Adj. R2	0.495		Mean of sim. Adj. R2	0.730		Mean of sim. Adj. R2	0.452	
CDF (%)	65.2		CDF (%)	88.3		CDF (%)	99.9	

Subperiod 7: 2012-03-01 - 2014-10-31			**Subperiod 8: 2014-11-01 - 2017-06-30**			**Subperiod 9: 2017-07-01 - 2020-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
OilVol	**-0.6**	0.003	OilVol	**-0.9**	0.009	OilVol	**-0.61**	0.019
DSpot	-0.35	0.131	FutRet	-0.21	0.384	DSpot	-0.22	0.414
PCAsent	0.25	0.188	BasMom	0.31	0.151	DolBeta	0.3	0.311
basis	-0.28	0.156	DolBeta	-0.18	0.428	fGom	0.14	0.454
DFX	0.23	0.138	sEnv	0.18	0.399	sCo	0.14	0.444
fBbl	0.25	0.052	fEp	-0.41	0.121	sRpc	0.24	0.326
sEpg	**0.2**	0.045	sEp	-0.31	0.278	HedgPres	0.17	0.397
Observations	132		Observations	131		Observations	135	
R2 / R2 adj.	0.7/0.68		R2 / R2 adj.	0.75/0.73		R2 / R2 adj.	0.64/0.62	
Mean of sim. Adj. R2	0.458		Mean of sim. Adj. R2	0.484		Mean of sim. Adj. R2	0.553	
CDF (%)	99.8		CDF (%)	99.6		CDF (%)	78.3	

Panel VII. 4: Predicting 8-weeks ahead Exxon Stock Returns (xomRet)

Subperiod 1: 1998-04-01 - 1999-11-30			**Subperiod 2: 1999-12-01 - 2002-07-31**			**Subperiod 3: 2002-08-01 - 2005-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
DProd	**-0.55**	0.041	DInv	0.33	0.19	fEpg	0.42	0.176
fBbl	0.2	0.468	fEnv	-0.27	0.335	OilVol	0.45	0.164
PCAfreq	-0.48	0.143	fCo	-0.39	0.072	DFX	0.35	0.195
sEpg	0.32	0.331	fBbl	-0.4	0.132	artcount	-0.26	0.482
fEnv	-0.16	0.466	fEp	**0.49**	0.004	sRpc	0.29	0.266
FutRet	0.16	0.429	vix_diff	0.18	0.304	sBbl	-0.16	0.443
WIPI_8wk	0.13	0.4	BE/ME	0.19	0.201	VIX	-0.16	0.445
Observations	71		Observations	129		Observations	131	
R2 / R2 adj.	0.75/0.72		R2 / R2 adj.	0.52/0.49		R2 / R2 adj.	0.39/0.35	
Mean of sim. Adj. R2	0.648		Mean of sim. Adj. R2	0.414		Mean of sim. Adj. R2	0.465	
CDF (%)	84.2		CDF (%)	80.2		CDF (%)	10.3	

Subperiod 4: 2005-04-01 - 2007-11-30			**Subperiod 5: 2007-12-01 - 2009-06-30**			**Subperiod 6: 2009-07-01 - 2012-02-29**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
sGom	0.52	0.087	VIX	**1.05**	0.011	BE/ME	0.52	0.087
artcount	0.34	0.153	WIPI_8wk	0.12	0.47	tnote_10y	-0.44	0.128
DolBeta	**0.55**	0.017	DInv	0.35	0.263	entropy	-0.23	0.423
tnote_10y	-0.36	0.171	OilVol	**-0.88**	0.043	fRpc	-0.25	0.334
fBbl	-0.34	0.089	DFX	-0.21	0.414	DOilVol	0.19	0.265
DFX	0.22	0.183	DOilVol	**0.53**	0.048	WIPI_8wk	0.2	0.253
entropy	0.17	0.257	BE/ME	0.35	0.083	sp500Ret	0.14	0.351
Observations	132		Observations	74		Observations	131	
R2 / R2 adj.	0.54/0.52		R2 / R2 adj.	0.68/0.64		R2 / R2 adj.	0.61/0.59	
Mean of sim. Adj. R2	0.433		Mean of sim. Adj. R2	0.606		Mean of sim. Adj. R2	0.542	
CDF (%)	84.4		CDF (%)	62.7		CDF (%)	67.2	

Subperiod 7: 2012-03-01 - 2014-10-31			**Subperiod 8: 2014-11-01 - 2017-06-30**			**Subperiod 9: 2017-07-01 - 2020-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
DInv	0.35	0.345	sGom	-0.28	0.342	DolBeta	-0.39	0.278
StkIdx	0.33	0.203	BE/ME	0.34	0.235	OilVol	0.5	0.182
VIX	0.32	0.155	VIX	0.29	0.357	sCo	-0.29	0.297
BasMom	0.43	0.104	artcount	0.24	0.274	sEpg	-0.17	0.493
tnote_10y	-0.44	0.061	InflaBeta	0.33	0.226	DFX	0.24	0.288
entropy	0.23	0.381	sEnv	0.3	0.151	DOilVol	-0.29	0.251
sBbl	0.16	0.254	StkIdx	-0.23	0.054	sRpc	-0.21	0.125
Observations	132		Observations	131		Observations	135	
R2 / R2 adj.	0.56/0.53		R2 / R2 adj.	0.64/0.61		R2 / R2 adj.	0.63/0.61	
Mean of sim. Adj. R2	0.466		Mean of sim. Adj. R2	0.526		Mean of sim. Adj. R2	0.547	
CDF (%)	77.8		CDF (%)	80.4		CDF (%)	75.9	

Panel VII. 5: Predicting 8-weeks ahead BP Stock Returns (bpRet)

Subperiod 1: 1998-04-01 - 1999-11-30			**Subperiod 2: 1999-12-01 - 2002-07-31**			**Subperiod 3: 2002-08-01 - 2005-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
BE/ME	0.31	0.358	HedgPres	0.31	0.494	BE/ME	-0.32	0.421
fEp	0.41	0.167	sGom	0.5	0.089	sGom	-0.52	0.294
HedgPres	0.18	0.448	entropy	**0.52**	0.024	fGom	-0.7	0.16
StkIdx	-0.22	0.234	PCAsent	-0.23	0.414	VIX	-0.4	0.162
sEpg	0.3	0.161	VIX	0.24	0.3	DolBeta	0.1	0.394
DolBeta	-0.28	0.132	BE/ME	0.34	0.121	fCo	-0.36	0.308
FutRet	0.22	0.047	fEp	0.16	0.417	PCAfreq	0.32	0.086
Observations	71		Observations	129		Observations	131	
R2 / R2 adj.	0.78/0.75		R2 / R2 adj.	0.56/0.53		R2 / R2 adj.	0.51/0.49	
Mean of sim. Adj. R2	0.684		Mean of sim. Adj. R2	0.497		Mean of sim. Adj. R2	0.490	
CDF (%)	69.9		CDF (%)	65.6		CDF (%)	47.3	

Subperiod 4: 2005-04-01 - 2007-11-30			**Subperiod 5: 2007-12-01 - 2009-06-30**			**Subperiod 6: 2009-07-01 - 2012-02-29**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
tnote_10y	**-0.68**	0.049	BE/ME	0.12	0.287	tnote_10y	-0.26	0.312
fBbl	-0.42	0.116	VIX	0.67	0.074	fRpc	**-0.57**	0.043
entropy	**0.58**	0.017	sGom	0.87	0.136	sRpc	-0.46	0.091
HedgPres	0.16	0.421	vix_diff	-0.44	0.099	PCAsent	**0.44**	0.043
fGom	-0.41	0.073	FutRet	0.34	0.337	InflaBeta	**0.53**	0.014
DInv	0.28	0.126	DInv	**0.32**	0.048	entropy	**-0.28**	0.035
DFX	**0.22**	0.045	DOilVol	0.23	0.18	DolBeta	**-0.34**	0.011
Observations	132		Observations	74		Observations	131	
R2 / R2 adj.	0.61/0.59		R2 / R2 adj.	0.7/0.67		R2 / R2 adj.	0.65/0.63	
Mean of sim. Adj. R2	0.476		Mean of sim. Adj. R2	0.650		Mean of sim. Adj. R2	0.488	
CDF (%)	92		CDF (%)	55.5		CDF (%)	90.8	

Subperiod 7: 2012-03-01 - 2014-10-31			**Subperiod 8: 2014-11-01 - 2017-06-30**			**Subperiod 9: 2017-07-01 - 2020-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
VIX	0.29	0.451	sGom	-0.43	0.338	DolBeta	-0.31	0.399
sBbl	0.56	0.139	BasMom	-0.3	0.268	OilVol	0.33	0.207
sEpg	-0.18	0.473	DInv	0.26	0.225	fGom	-0.23	0.455
fGom	-0.54	0.071	FutRet	0.24	0.21	BE/ME	0.49	0.083
BasMom	0.22	0.262	PCAsent	-0.35	0.112	sCo	-0.27	0.255
sRpc	-0.22	0.342	PCAfreq	-0.22	0.197	artcount	0.38	0.071
sGom	-0.33	0.226	VIX	0.21	0.192	sRpc	-0.23	0.141
Observations	132		Observations	131		Observations	135	
R2 / R2 adj.	0.59/0.56		R2 / R2 adj.	0.72/0.7		R2 / R2 adj.	0.62/0.6	
Mean of sim. Adj. R2	0.483		Mean of sim. Adj. R2	0.502		Mean of sim. Adj. R2	0.529	
CDF (%)	78.4		CDF (%)	98.4		CDF (%)	78	

Panel VII. 6: Predicting 8-weeks ahead Royal Dutch Shell Stock Returns (rdsaRet)

Subperiod 1: 1998-04-01 - 1999-11-30			**Subperiod 2: 1999-12-01 - 2002-07-31**			**Subperiod 3: 2002-08-01 - 2005-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
BE/ME	0.31	0.284	DInv	0.48	0.145	PCAall	-0.51	0.061
sEpg	**0.63**	0.004	fBbl	-0.34	0.395	entropy	0.48	0.089
fGom	-0.39	0.169	BasMom	-0.22	0.424	BE/ME	-0.31	0.249
HedgPres	0.18	0.376	HedgPres	0.14	0.437	StkIdx	-0.23	0.381
FutRet	**0.34**	0.026	sEp	0.33	0.278	DOilVol	-0.23	0.403
DolBeta	-0.22	0.17	fEp	0.35	0.286	DFX	0.23	0.242
fBbl	-0.2	0.072	BE/ME	0.2	0.358	sp500Ret	0.22	0.088
Observations	71		Observations	131		Observations	131	
R2 / R2 adj.	0.82/0.8		R2 / R2 adj.	0.41/0.38		R2 / R2 adj.	0.53/0.5	
Mean of sim. Adj. R2	0.701		Mean of sim. Adj. R2	0.475		Mean of sim. Adj. R2	0.498	
CDF (%)	86.9		CDF (%)	13.4		CDF (%)	52.4	

Subperiod 4: 2005-04-01 - 2007-11-30			**Subperiod 5: 2007-12-01 - 2009-06-30**			**Subperiod 6: 2009-07-01 - 2012-02-29**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
fRpc	0.37	0.186	DInv	0.38	0.405	tnote_10y	-0.26	0.312
InflaBeta	0.42	0.21	fRpc	0.08	0.399	fEnv	-0.29	0.328
FutRet	-0.22	0.437	BasMom	0.33	0.243	DOilVol	0.2	0.325
DInv	0.35	0.213	FutRet	1.6	0.093	BE/ME	0.18	0.375
VIX	0.4	0.138	DSpot	-1.23	0.178	fRpc	-0.26	0.258
fCo	0.43	0.076	Mom	-0.36	0.091	fEpg	-0.36	0.057
sEp	-0.31	0.057	vix_diff	-0.23	0.292	Mom	**-0.29**	0.023
Observations	132		Observations	74		Observations	131	
R2 / R2 adj.	0.48/0.46		R2 / R2 adj.	0.68/0.64		R2 / R2 adj.	0.6/0.57	
Mean of sim. Adj. R2	0.449		Mean of sim. Adj. R2	0.658		Mean of sim. Adj. R2	0.518	
CDF (%)	53.3		CDF (%)	43.2		CDF (%)	70.7	

Subperiod 7: 2012-03-01 - 2014-10-31			**Subperiod 8: 2014-11-01 - 2017-06-30**			**Subperiod 9: 2017-07-01 - 2020-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
sRpc	-0.39	0.277	fEnv	0.28	0.318	DolBeta	-0.32	0.26
sEnv	0.81	0.154	VIX	0.27	0.298	OilVol	0.14	0.425
artcount	0.27	0.341	sEnv	0.92	0.16	sCo	-0.4	0.075
fEnv	0.49	0.275	HedgPres	0.33	0.088	sEpg	-0.4	0.08
VIX	0.26	0.353	tnote_10y	0.31	0.109	Mom	-0.29	0.334
StkIdx	0.18	0.476	basis	0.2	0.297	sEp	**0.4**	0.037
sGom	-0.21	0.356	sp500Ret	-0.12	0.48	fEp	0.29	0.062
Observations	132		Observations	131		Observations	135	
R2 / R2 adj.	0.54/0.52		R2 / R2 adj.	0.61/0.59		R2 / R2 adj.	0.66/0.64	
Mean of sim. Adj. R2	0.517		Mean of sim. Adj. R2	0.466		Mean of sim. Adj. R2	0.495	
CDF (%)	48.2		CDF (%)	91.2		CDF (%)	94.8	

Panel VII. 7: Predicting 8-weeks ahead Oil Inventory Changes (*DInv*)

Subperiod 1: 1998-04-01 - 1999-11-30			**Subperiod 2: 1999-12-01 - 2002-07-31**			**Subperiod 3: 2002-08-01 - 2005-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
tnote_10y	-0.86	0.117	InflaBeta	0.31	0.37	PCAfreq	-0.92	0.108
entropy	**0.7**	0.046	DolBeta	0.37	0.246	InflaBeta	0.52	0.071
BasMom	-0.3	0.313	vix_diff	0.28	0.36	tnote_10y	-0.24	0.406
StkIdx	-0.31	0.183	fBbl	-0.49	0.112	entropy	-0.34	0.163
PCAfreq	-0.56	0.114	fEp	0.35	0.167	PCAsent	-0.47	0.15
fEnv	-0.26	0.28	sCo	0.22	0.307	sEpg	-0.34	0.201
sEpg	0.17	0.271	sBbl	-0.21	0.284	fEpg	-0.31	0.118
Observations	72		Observations	131		Observations	131	
R2 / R2 adj.	0.78/0.75		R2 / R2 adj.	0.47/0.44		R2 / R2 adj.	0.6/0.58	
Mean of sim. Adj. R2	0.678		Mean of sim. Adj. R2	0.515		Mean of sim. Adj. R2	0.532	
CDF (%)	78.7		CDF (%)	24.1		CDF (%)	67	

Subperiod 4: 2005-04-01 - 2007-11-30			**Subperiod 5: 2007-12-01 - 2009-06-30**			**Subperiod 6: 2009-07-01 - 2012-02-29**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
basis	0.38	0.284	sEpg	-0.72	0.068	fGom	0.4	0.146
PCAsent	-0.02	0.433	Mom	-0.82	0.359	OilVol	-0.23	0.491
artcount	-0.27	0.393	BasMom	-0.34	0.345	fCo	0.39	0.071
sEpg	-0.65	0.092	DFX	-0.21	0.423	tnote_10y	**0.58**	0.013
fEpg	-0.83	0.111	InflaBeta	-0.86	0.271	DProd	0.27	0.177
fGom	-0.51	0.231	fEp	-0.27	0.439	sBbl	-0.32	0.114
BE/ME	0.22	0.167	fEpg	0.2	0.426	BE/ME	**0.26**	0.035
Observations	131		Observations	75		Observations	131	
R2 / R2 adj.	0.51/0.48		R2 / R2 adj.	0.7/0.67		R2 / R2 adj.	0.59/0.57	
Mean of sim. Adj. R2	0.525		Mean of sim. Adj. R2	0.708		Mean of sim. Adj. R2	0.515	
CDF (%)	33.1		CDF (%)	32.2		CDF (%)	67.4	

Subperiod 7: 2012-03-01 - 2014-10-31			**Subperiod 8: 2014-11-01 - 2017-06-30**			**Subperiod 9: 2017-07-01 - 2020-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
HedgPres	-0.15	0.452	PCAall	1.08	0.055	fEnv	-0.43	0.206
fEpg	**-0.7**	0.009	fEpg	-0.9	0.089	PCAall	-0.6	0.159
tnote_10y	0.47	0.074	fCo	**-1.06**	0.046	DFX	-0.35	0.125
InflaBeta	**-0.75**	0.005	HedgPres	**-0.39**	0.031	InflaBeta	-0.4	0.073
BE/ME	0.45	0.082	PCAsent	**0.69**	0.043	fEpg	-0.31	0.342
sEp	-0.31	0.055	fRpc	-0.24	0.063	VIX	0.19	0.393
DolBeta	**0.29**	0.009	fBbl	**0.28**	0.035	HedgPres	-0.18	0.308
Observations	131		Observations	131		Observations	134	
R2 / R2 adj.	0.57/0.55		R2 / R2 adj.	0.71/0.7		R2 / R2 adj.	0.58/0.56	
Mean of sim. Adj. R2	0.449		Mean of sim. Adj. R2	0.504		Mean of sim. Adj. R2	0.529	
CDF (%)	84.2		CDF (%)	97.1		CDF (%)	59.7	

Panel VII. 8: Predicting 8-weeks ahead Oil Production Changes (*DProd*)

Subperiod 1: 1998-04-01 - 1999-11-30			**Subperiod 2: 1999-12-01 - 2002-07-31**			**Subperiod 3: 2002-08-01 - 2005-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
BE/ME	-0.86	0.201	fGom	0.5	0.202	Mom	**0.66**	0.034
HedgPres	-0.25	0.294	InflaBeta	**-0.43**	0.023	DFX	-0.26	0.234
fBbl	0.31	0.248	vix_diff	0.21	0.498	artcount	-0.41	0.09
DolBeta	0.55	0.116	OilVol	0.43	0.055	DInv	0.29	0.163
DFX	-0.31	0.141	sEp	-0.26	0.116	fEp	0.23	0.267
DSpot	-0.22	0.214	sBbl	-0.23	0.096	HedgPres	-0.17	0.307
Mom	-0.5	0.179	DOilVol	**-0.25**	0.042	tnote_10y	-0.13	0.362
Observations	72		Observations	131		Observations	131	
R2 / R2 adj.	0.79/0.76		R2 / R2 adj.	0.64/0.62		R2 / R2 adj.	0.6/0.58	
Mean of sim. Adj. R2	0.658		Mean of sim. Adj. R2	0.488		Mean of sim. Adj. R2	0.487	
CDF (%)	85.6		CDF (%)	92.6		CDF (%)	79	

Subperiod 4: 2005-04-01 - 2007-11-30			**Subperiod 5: 2007-12-01 - 2009-06-30**			**Subperiod 6: 2009-07-01 - 2012-02-29**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
PCAall	-0.32	0.46	VIX	**0.76**	0.003	tnote_10y	-0.32	0.487
sGom	0.46	0.057	FutRet	0.15	0.418	InflaBeta	0.57	0.242
DSpot	-2.25	0.378	BasMom	**0.47**	0.008	VIX	0.38	0.412
DInv	0.27	0.207	PCAsent	**0.44**	0.009	StkIdx	0.36	0.268
BasMom	**0.48**	0.048	BE/ME	**0.59**	0.001	Mom	0.44	0.335
FutRet	1.85	0.19	HedgPres	0.18	0.358	liquidity	-0.17	0.415
DolBeta	-0.28	0.149	fBbl	-0.14	0.188	entropy	0.12	0.476
Observations	131		Observations	75		Observations	131	
R2 / R2 adj.	0.6/0.57		R2 / R2 adj.	0.82/0.81		R2 / R2 adj.	0.42/0.38	
Mean of sim. Adj. R2	0.520		Mean of sim. Adj. R2	0.596		Mean of sim. Adj. R2	0.463	
CDF (%)	72.7		CDF (%)	96.6		CDF (%)	21.3	

Subperiod 7: 2012-03-01 - 2014-10-31			**Subperiod 8: 2014-11-01 - 2017-06-30**			**Subperiod 9: 2017-07-01 - 2020-03-31**		
Predictors	coef	pval	Predictors	coef	pval	Predictors	coef	pval
fGom	-0.46	0.103	HedgPres	-0.33	0.304	sGom	0.56	0.093
OilVol	**-0.52**	0.045	fBbl	0.23	0.38	VIX	**0.51**	0.025
Mom	**0.41**	0.041	sGom	**0.68**	0.026	sRpc	**0.52**	0.039
sBbl	-0.3	0.142	PCAsent	**-0.68**	0.024	fCo	0.24	0.31
fEpg	-0.25	0.194	fEnv	**0.52**	0.025	DolBeta	-0.27	0.152
StkIdx	-0.17	0.356	OilVol	**0.43**	0.007	sBbl	-0.24	0.2
DOilVol	0.2	0.208	Mom	**0.32**	0.003	fRpc	0.23	0.239
Observations	131		Observations	131		Observations	134	
R2 / R2 adj.	0.63/0.61		R2 / R2 adj.	0.57/0.54		R2 / R2 adj.	0.61/0.58	
Mean of sim. Adj. R2	0.491		Mean of sim. Adj. R2	0.455		Mean of sim. Adj. R2	0.509	
CDF (%)	88.7		CDF (%)	80.5		CDF (%)	77.8	

Table A. VIII: Stepwise forward selection at the four-week horizon

The table shows the forecasting regression results for all eight dependent variables at the four-week horizon using stepwise forward selection to choose seven regressors from all the variables described in Table I, except *ovx_diff* (which is only available after 2007) and *sdf_fullSample* (the values of which reflect future data). We also exclude three energy company stock returns (*xomRet, bpRet, rdsaRet*) from our regressors and instead include *StkIdx* (which is the average of the three stock returns). All dependent and independent variables are first detrended and residualized with respect to the lagged four-week dependent variable. Only predictors that were chosen by at least one model are included in this table. Coefficients are standardized using the ratios of the standard deviations of the dependent and predicting variables. Superscripts before coefficients indicate order in forward selection (1=chosen first). The p-values are obtained using Monte Carlo simulations that use an AR(8) process to simulate the LHS variable, as well as forward selection to produce both adjusted R^2 and t-statistic distributions. The p-values refer to the minimum of the fraction of simulated t-statistics less than the empirical t-statistic, and 1 minus the fraction of simulated t-statistics less than the empirical t-statistic, where the comparison is relative to the distribution of the order in which the variables were chosen. The bootstrap was repeated 1,000 times. The table also reports the mean of simulated adjusted R^2 resulting from the same bootstrap, as well as the corresponding CDF percentage, computed as the percent of adjusted R^2 simulations less than the empirical adjusted R^2. Statistically significant results at the 10% level or better are shown in bold.

Predictors	*FutRet*		*Dspot*		*xomRet*		*bpRet*		*rdsaRet*		*DInv*		*DProd*	
	coef	*pval*	*coef*	*pval*	*coef*	*pval*	*coef*	*pval*	*coef*	*pval*	*coef*	*pval*	*coef*	*pval*
FutRet													[4]-0.08	0.37
OilVol	[6]0.12	0.094									[1]0.16	0.105		
DInv									[5]0.07	0.396			[3]0.10	0.394
DProd			[7]0.08	0.099	[5]0.07	0.375								
tnote_10y	**[7]-0.09**	0.032			[6]-0.08	0.245							[6]0.07	0.459
sp500Ret					[4]-0.10	0.198								
basis			[1]0.13	0.345			[2]-0.08	0.469						
HedgPres											**[6]-0.13**	0.016		
BE/ME	**[1]0.26**	0.007	**[4]0.26**	0.001	[2]0.14	0.226			[2]0.15	0.187				
Mom							**[1]-0.24**	0.002						
InflaBeta													[5]0.07	0.498
OpenInt					[3]0.09	0.343	[3]0.10	0.263						
VIX	**[5]-0.18**	0.003	[2]-0.14	0.085					[7]0.08	0.126			[1]0.11	0.324
artcount							**[7]-0.10**	0.033			**[4]-0.18**	0.002		
entropy	**[4]0.20**	0.000	**[3]0.18**	0.008			[6]0.10	0.102	**[4]0.19**	0.026	**[2]-0.17**	0.038		
fCo									[6]0.13	0.13				
sGom	**[3]0.19**	0.005	**[6]0.13**	0.007	[7]0.08	0.107	[5]0.11	0.148	[3]0.14	0.058				
fGom													[2]-0.17	0.147
sEnv							[4]0.09	0.271						
fEpg													[7]-0.07	0.365
fBbl	[2]-0.17	0.052	**[5]-0.17**	0.015	[1]-0.13	0.477			[1]-0.13	0.344				
sRpc											[7]-0.10	0.22		
fRpc											**[3]-0.26**	0.003		
sEp											**[5]-0.13**	0.035		
Observations	1136		1136		1134		1134		1136		1136		1136	
R^2 / R^2 adjusted	0.115 / 0.110		0.127 / 0.122		0.047 / 0.041		0.069 / 0.063		0.068 / 0.062		0.117 / 0.111		0.049 / 0.043	
Mean of sim. Adj. R2	0.0528		0.0532		0.0541		0.0568		0.0549		0.0588		0.0701	
CDF (%)	100		100		20.2		68.6		68.3		99.2		7.2	

End of Online Appendix.