Intraday Crude Oil Volatility: Assessing the Impact of Economic Announcements and Mixed-frequency Data

Igor F. B. Martins[1], **Audronė Virbickaitė**[2], Hoang Nguyen[3] and Hedibert F. Lopes[4]

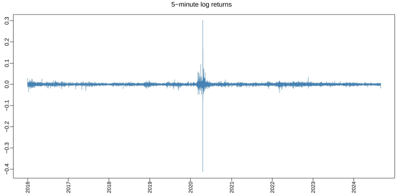
[1]School of Business, Örebro University, Sweden; [2]**Department of Quantitative Methods, CUNEF Universidad, Madrid, Spain**; [3]Department of Management and Engineering, Linköping University, Sweden; [4]Insper, São Paulo, Brazil

Thursday 4th September, 2025

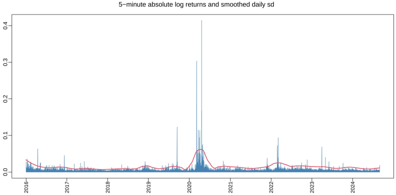
Introduction and motivation

- Risk management for firms and investors.
- Understanding oil volatility helps traders and analysts spot trends, manage liquidity, and design better trading strategies and portfolios.
- The macroeconomic outlook is impacted by fluctuations in oil prices:
  - Oil price volatility helps to forecast aggregate outputs and economic growth in the U.S. (Ferderer, 1996).
  - Option-implied oil price volatility is a strong predictor of economic growth beyond traditional uncertainty measures (Gao et al., 2022).
  - Oil uncertainty can lead to a delay of investment, industrial production and consumption (Elder and Serletis, 2010; Jo, 2014).

First look at the high-frequency intraday oil returns

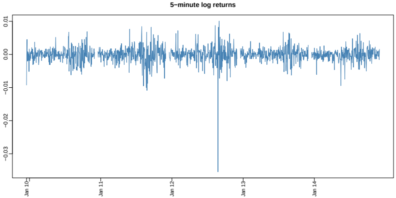


(a) 2016-2024



(b) 2016-2024

(c) January 2016



(d) Second week of January 2016

Contributions

Contributions:

1. A high-frequency model for oil future volatility which takes into account the following effects:
   - Volatility persistence, intraday seasonal effects, macroeconomic announcements and the long-term volatility spillovers.
2. Information release plays a crucial role in forecasting the oil futures market.
   - Geopolitical events, e.g. Noguera-Santaella (2016), Brandt and Gao (2019).
   - Economic data releases and policy announcements:
     - Supplier news: OPEC and U.S. Strategic Petroleum Reserve (SPR) announcements (Demirer and Kutan, 2010; Schmidbauer and Rösch, 2012), EIA weekly crude oil inventory (Bu, 2014; Bjursell et al., 2015), oil supply news shocks (Känzig, 2021);
     - Demand news: Central bank announcements (Yang et al., 2023), scheduled macroeconomic releases (López, 2018; Kilian and Vega, 2011; Chatrath et al., 2012), monetary policy surprises (Basistha and Kurov, 2015).
3. Incorporate exogenous variables into low-frequency volatility via MIDAS (Mixed Data Sampling, Ghysels et al., 2004, 2006).
4. More accurate out-of-sample volatility forecasting.

SV MIDAS models for high frequency intraday returns

Following Stroud and Johannes (2014), let $\mathbf{y} = \{y_t\}_{t=1}^T$ be a series of de-meaned 5-minute log returns, and let $\mathbf{h} = \{h_t\}_{t=1}^T$ be the corresponding log-variance,

$$y_t = \exp\{0.5h_t\}\epsilon_t, \qquad \text{where } \epsilon_t \sim \mathbf{N}(0,1) \text{ for } t = 1,\ldots,T,$$

$$h_t = m_\tau + p_t + s_t + e_t,$$

$$p_t = \beta p_{t-1} + \sigma_\eta \eta_t, \qquad \eta_t \sim \mathbf{N}(0,1), \leftarrow \text{Persistence component}$$

$$s_t = \sum_{k=1}^{K} H_{tk}\beta_k = \mathbf{H}_t^{'}\boldsymbol{\beta}, \qquad \leftarrow \text{Seasonal component}$$

$$e_t = \sum_{i=1}^{N} I_{it}\alpha_i, = \mathbf{I}_t^{'}\boldsymbol{\alpha}, \qquad \leftarrow \text{Announcement component}$$

$$m_\tau = m_0 + \sum_{j=1}^{J}\delta_j\left[\sum_{l=1}^{L_j}\phi_l(w_j)X_{j,\tau-l}\right]. \qquad \leftarrow \text{MIDAS component}$$

HF SV-MIDAS Model

$$m_\tau = m_0 + \sum_{j=1}^{J} \delta_j \left[\sum_{l=1}^{L_j} \phi_l(w_j) X_{j,\tau-l} \right] \qquad \leftarrow \text{MIDAS component}$$

- The unconditional long-term level of the log-variance is m_0 while m_τ is the conditional level shift in the long-term log-variance.
- A set of low-frequency variables X_j for $j = 1, \ldots, J$ is available at the end of the US market trading day.
- The effect of the j^{th} variable can be described through the size of a regression coefficient δ_j.
- The restricted beta weighting function $\phi_l(w_j)$ creates a weighting scheme and regularizes for the effect of the last L_j observations of the explanatory variable X_j.

HF SV-MIDAS Model

The restricted beta weighting function $\phi_l(w_j)$ creates a weighting scheme and regularizes for the effect of the last L_j observations of the explanatory variable X_j.

$$\phi_l(w_j) = \frac{\left[1 - \frac{l}{L_j+1}\right]^{w_j-1}}{\sum_{m=1}^{L_j}\left[1 - \frac{m}{L_j+1}\right]^{w_j-1}}.$$

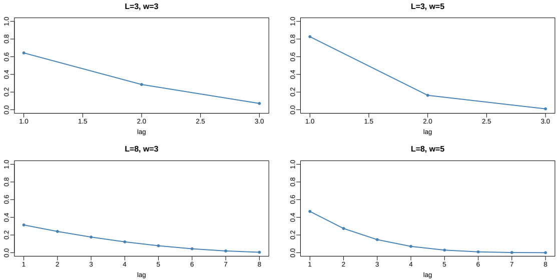


Figure: Restricted beta weighting function for different values of L and w.

Bayesian inference

Posterior $\propto$ **Likelihood** $\times$ **Prior**

- SV model is a non-linear state-space model:

$$
\begin{aligned}
y_t &= \exp\{0.5 h_t\}\epsilon_t, \qquad \text{where } \epsilon_t \sim \mathbf{N}(0,1) \text{ for } t = 1, \ldots, T, \\
h_t &= m_\tau + p_t + s_t + e_t,
\end{aligned}
$$

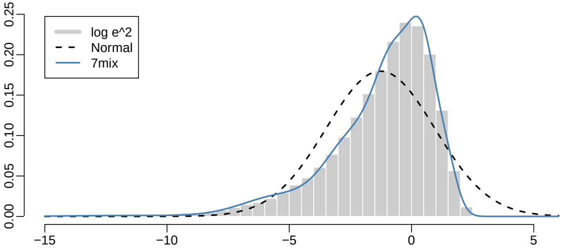
- Apart from model parameters, the state variables h_t need to be estimated as well.
- Rely on particle approximation: infeasible in high-frequency setting.

Bayesian inference

To facilitate the calculation of the likelihood, we rewrite the model as

$$\log y_t^2 = h_t + log\epsilon_t^2$$
$$= m_\tau + p_t + s_t + e_t + u_t$$

- Where $u_t = \log \epsilon_t^2 \sim log\chi^2(1)$, approximate $u_t \sim N(\mu = -1.27, \sigma^2 = 0.5\pi^2)$.
- Better: $log\chi^2(1) \sim$ a mixture of 7 Gaussian components (Kim et al., 1998).



Prior specification: other parameters

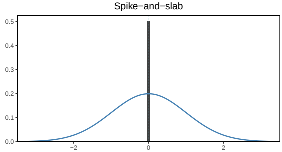
We assume weakly informative priors for the following parameters:

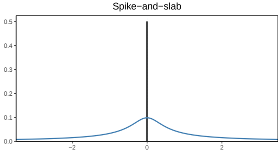
- $\beta \sim N(\beta_0 = 0.95, V_{\beta_0} = 0.5^2)$ ← Persistence
- $\sigma_\eta^2 \sim IG(\alpha_{\sigma_\eta} = 10, \beta_{\sigma_\eta} = 1)$ ← Persistence
- $\{\beta_k\}_{k=1}^{K-1} \sim N(0, 0.5)$ ← Seasonal
- $m_0 \sim N(\bar{y}^*, 2)$ ← MIDAS unconditional mean
- $\{\delta_j\}_{j=1}^{J} \sim N(0, 2)$ ← MIDAS coefficients
- $\{w_j\}_{j=1}^{J} \sim U(1, 20)$ ← MIDAS weights
- Spike-and-slab ← Announcements

Prior specification: announcements

- A spike and slab prior for $\{\alpha_i\}_{i=1}^N$ due to a large number of announcements
- The spike as a Dirac delta on zero and the slab as a Normal distribution with a large variance, similarly to Mitchell and Beauchamp (1988).

$$\alpha_i|\pi_i \sim (1-\pi_i)\delta_0 + \pi_i N(0,\sigma_\alpha^2)$$
$$\pi_i \sim Bernoulli(\gamma)$$
$$\gamma \sim Beta(\alpha_\gamma = 1, \beta_\gamma = 1)$$
$$\sigma_\alpha^2 \sim IG(\alpha_\alpha = 20, \beta_\alpha = 20)$$





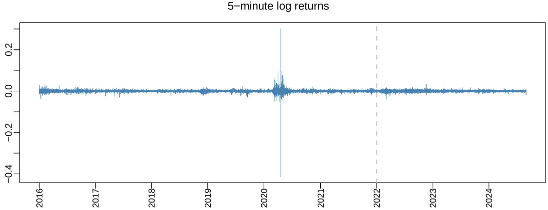
MCMC algorithm

Let $\Theta = (\beta, \sigma_\eta, \{p_t\}_{t=1}^{T}; \{\beta_k\}_{k=1}^{K-1}; \{\alpha_i\}_{i=1}^{N}, \{\pi_i\}_{i=1}^{N}, \sigma_a^2, \gamma; m_0, \{w_j\}_{j=1}^{J}, \{\delta_j\}_{j=1}^{J})$, and let $\mathbf{\Psi}$ be all the parameters in Θ except the ones we sample from. Under the 7-Gaussian approximation, given the mixture assignments, we have a Gaussian DLM.

- Sample $\beta|\mathbf{\Psi}, \mathbf{y}$ from a conjugate Gaussian distribution.
- Sample $\sigma_\eta^2|\mathbf{\Psi}, \mathbf{y}$ from a conjugate inverse Gamma distribution.
- Sample $\{p_t\}_{t=1}^{T}|\mathbf{\Psi}$ using the FFBS algorithm in Carter and Kohn (1994) and Frühwirth-Schnatter (1994).
- Sample $\{\beta_k\}_{k=1}^{K}|\mathbf{\Psi}, \mathbf{y}$ from a conjugate Gaussian distribution.
- Sample $\sigma_a^2|\mathbf{\Psi}, \mathbf{y}$ from an inverse Gamma distribution.
- Sample $\gamma|\mathbf{\Psi}, \mathbf{y}$ from a conjugate Beta distribution.
- Sample $\{\pi_i\}_{i=1}^{N}|\mathbf{\Psi}, \mathbf{y}$ from a conjugate Bernoulli distribution, and sample $\{\alpha_i\}_{k=1}^{K}|\mathbf{\Psi}, \mathbf{y}$ from a conjugate Gaussian distribution for $\pi_i = 1$.
- Sample $\{w_j\}_{j=1}^{J}|\mathbf{\Psi}, \mathbf{y}$ using a Metropolis-Hasting step.
- Sample $m_0, \{\delta_j\}_{j=1}^{J}|\mathbf{\Psi}, \mathbf{y}$ from a conjugate Gaussian distribution.

Data

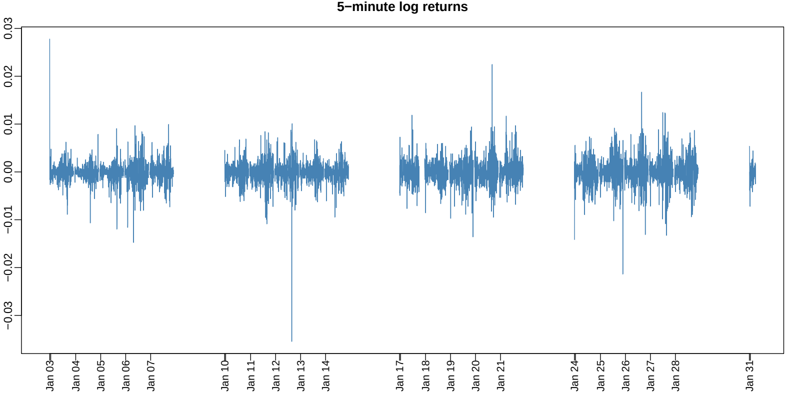
Prices of WTI futures at the New York Mercantile Exchange (NYMEX) from 2016-01-03 to 2024-08-30, with 288 5-minute observations daily on continuous oil futures. Data source: CME Group and Bloomberg (for events).



	2016 Jan - 2021 Dec	2022 Jan - 2024 Aug
mean	0.0000000	0.0000007
stdev	0.0021341	0.0014921
Q1	-0.0005097	-0.0005578
Q2	-0.0000002	-0.0000002
Q3	0.0005198	0.0005748
min	-0.4138839	-0.0400446
max	0.3011773	0.0939433
n	425,127	189,240

Table: Main descriptive statistics for the 5-minute log returns for 2016-01-03 till 2024-08-30.

Data



- Remove holidays, weekends and the 1-hour break in trading between 17:00 and 18:00.
- Plenty of stale prices, especially during the low trading activity hours.

Seasonality: EDA

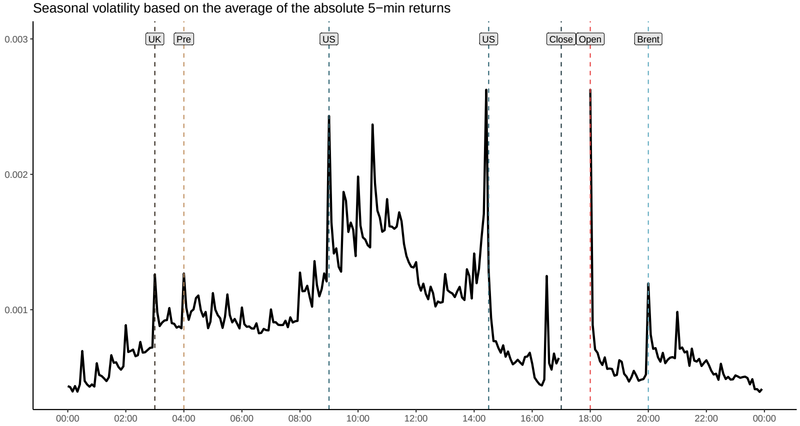


Figure: Intraday pattern of absolute log returns. The series plots the cross-day average return at each 5-minute mark (288 in total per trading day).

Event data

Rigorous data cleaning procedure: after loading the raw **economic and commodity** event data for the **US, Germany and China**, we perform the following:

- Remove all events without hour-minute information.
- Removing duplicates: often there are multiple events for the same data being released.
  - For example, FOMC *Rate Decision (Lower Bound)* and *FOMC Rate Decision (Upper Bound)*.
  - We are keeping this types of announcements as a single event.
- Removing rare events: must repeat **at least** 2 times per year.
- Finally, perform case by case analysis.
  - For example, *CH GDP SA QoQ* and *CH GDP YTD YoY* provide the same information.
  - Not exact duplicate due to the observation in 2016-04-15 21:35:00 vs 2016-04-14 22:00:00; in all other cases they occur at the same time.
  - Keep the earlier occurrence.
- 94 events in total (93 + Sunday Open).

Event data: EDA

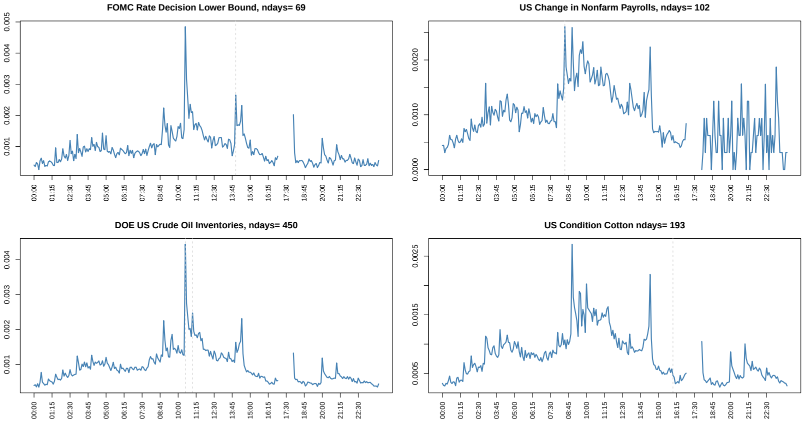
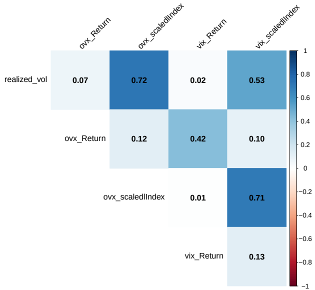


Figure: Average of the absolute value of intraday return for each 5-minute mark on the days of the specific events (US only). Vertical dashed line indicates the time of the event.

MIDAS variables

- Use daily OVX and VIX as explanatory variables:
  - VIX and OVX embed forward-looking volatility expectations from option markets.
  - They capture risk spillovers between equity and crude oil markets.
  - OVX is a direct measure of oil-specific uncertainty, complementing realized volatility.
- In scaled-log-levels and in log-returns.



- For now assume $L_j = 1$ in MIDAS for both variables.

Posterior distributions

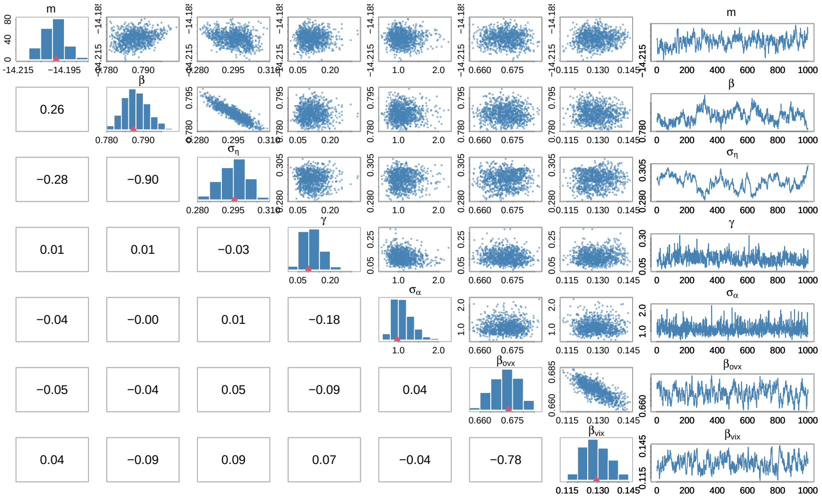
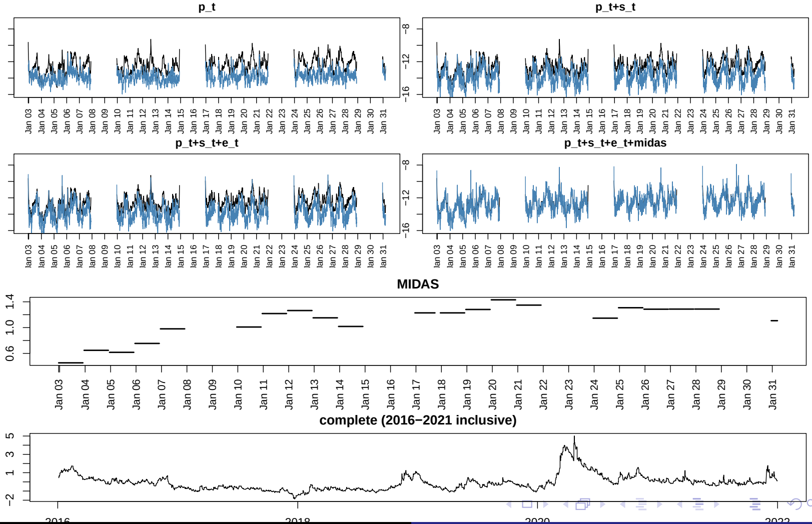


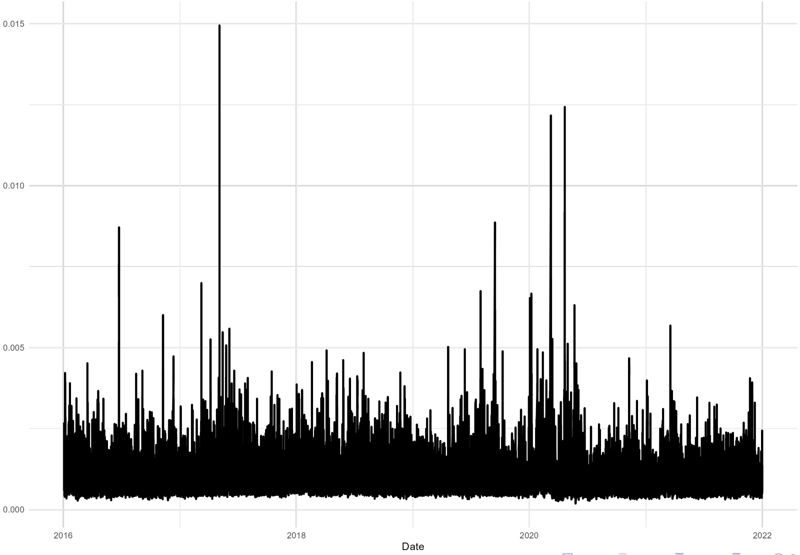
Figure: Posterior distributions of the estimated parameters of the SV MIDAS model with OVX and VIX Indexes (scaled log-levels), 10k iterations, 5k discarded as burn-in, thinned every 5th. Red triangles are posterior MCMC modes.

Volatility components

Benchmark SV estimated using the `stochvol` package of Hosszejni and Kastner (2021).

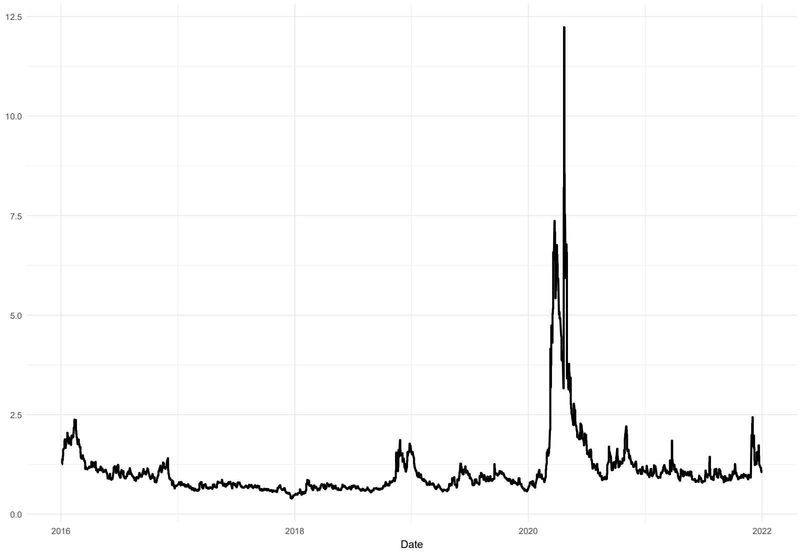


SV component: $exp\left(\frac{level+p_t}{2}\right)$



Seasonal component: $exp\left(\frac{s_t}{2}\right)$

MIDAS component: $exp\left(\frac{m_\tau}{2}\right)$



Estimated events: spike and slab prior at work

The selected OLS events are such that p-values < 0.1 and MCMC events are such that the posterior inclusion probability is > 0.8.

	OLS estimate	OLS p-value	Post.beta	Post.mode	Post.inclusion
GE.GfK.Consumer.Confidence	-1.69	0.01	-0.01	0.00	0.04
GE.HCOB.Germany.Services.PMI	1.13	0.03	0.00	0.00	0.02
GE.ZEW.Survey.Current.Situation	-1.31	0.04	-0.08	0.00	0.18
US.Change.in.Nonfarm.Payrolls	2.01	0.00	1.24	1.11	1.00
US.Condition...Cotton	-1.08	0.09	-0.03	0.00	0.12
US.CPI.MoM	0.92	0.17	0.58	-0.00	0.90
US.DOE.U.S..Crude.Oil.Inventories	2.83	0.00	2.30	2.27	1.00
US.FOMC.Rate.Decision..Lower.Bound.	1.53	0.05	1.19	1.27	1.00
US.ISM.Manufacturing	1.23	0.06	0.07	0.00	0.16
US.U..of.Mich..Sentiment	-0.17	0.72	-0.43	-0.00	0.84
SundayOpen	2.75	0.00	2.13	2.09	1.00

Table: OLS and MCMC estimation results for the events: OLS betas, p-values, MCMC posterior betas (median and mode) and posterior inclusion probabilities.

Posterior distributions

The selected OLS events are such that p-values < 0.1, and MCMC events are such that the posterior inclusion probability is > 0.8.

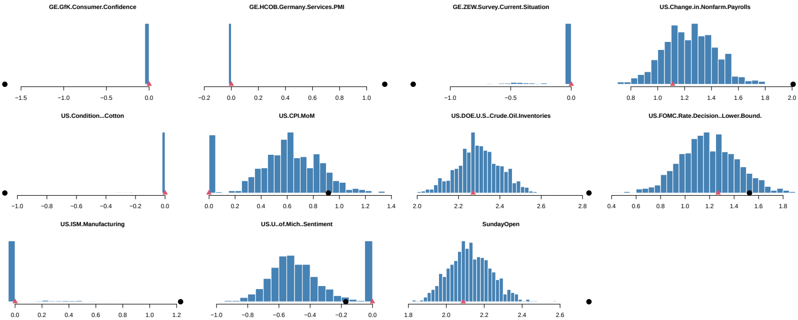


Figure: Posterior distributions of the estimated MIDAS parameters, 10k iterations, 5k discarded as burn-in, thinned every 5th. Black dots are OLS, red triangles are posterior MCMC modes.

Forecasting performance

- Produce 1-s-a forecasts for the out-of-sample period Jan 2022 - Aug 2024 (189k observations).
- Fix parameter estimates at the posterior medians and forecast the point volatility $\widehat{Vol}_{t+1|t}$.
- For every 5-minute mark calculate realized volatilities based on 1-minute log returns as a volatility proxy $RVol_t$.
- Competing models:
  - SSVA (no MIDAS), SSVAg (Gaussian prior), SSV (no announcements), SV (plain)
  - GARCH (Bollerslev, 1987), GJR-GARCH (Glosten et al., 1993).
  - AR1-RV, a version of HAR model (Corsi, 2009).
- Employ Mincer-Zarnowitz regressions and Diebold-Mariano test (MSE and MAE loss functions).
- Calculate volatility-managed portfolios (Moreira and Muir, 2017).

Mincer-Zarnowitz regression

$$RVol_{t+1} = b_0 + b_1 \widehat{Vol}_{t+1|t} + \varepsilon_t$$

	Mincer–Zarnowitz			
	b_0	b_1	$MSE \times 10^{-3}$	R^2
SSVA MIDAS^{ovx,vix}	0.009	0.933	5.313	0.473
SSVA MIDAS^{ovx}	0.009	0.944	5.302	0.474
SSVA MIDAS^{vix}	0.012	0.881	5.329	0.471
SSVA MIDAS^{rovx,rvix}	0.013	0.889	5.326	0.471
SSVA MIDAS^{rovx}	0.013	0.887	5.329	0.471
SSVA MIDAS^{rvix}	0.013	0.886	5.336	0.470
SSVA	0.013	0.888	5.336	0.470
SSVAg	0.013	0.885	5.341	0.470
SSV	0.013	0.884	5.499	0.454
SV	0.013	0.896	6.157	0.389
GARCH	0.018	0.779	6.490	0.356
AR1-RV	0.017	0.849	7.017	0.304

Diebold - Mariano test

	Squared		Absolute	
	DM	$p-value$	*DM*	$p-value$
SSVA MIDAS^{ovx,vix}				
SSVA MIDAS^{ovx}	3.03	1	18.2	1
SSVA MIDAS^{vix}	-6.06	$p < 0.01$	-18.21	$p < 0.01$
SSVA MIDAS^{rovx,rvix}	-3.71	$p < 0.01$	-9.89	$p < 0.01$
SSVA MIDAS^{rovx}	-3.83	$p < 0.01$	-10.3	$p < 0.01$
SSVA MIDAS^{rvix}	-3.73	$p < 0.01$	-10.05	$p < 0.01$
SSVA	-3.68	$p < 0.01$	-9.78	$p < 0.01$
SSVAg	-4.05	$p < 0.01$	-10.79	$p < 0.01$
SSV	-2.87	$p < 0.01$	-12.92	$p < 0.01$
SV	-6.88	$p < 0.01$	-51.21	$p < 0.01$
GARCH	-10.81	$p < 0.01$	-68.2	$p < 0.01$
AR1-RV	-7.44	$p < 0.01$	-72.67	$p < 0.01$

Volatility-managed portfolios

Moreira and Muir (2017) strategy: exploit the fact that the mean-variance trade-off weakens in periods of high volatility.

	Mean	*Sharpe*	ΔUMV
WTI	0.053	0.132	
SSVA MIDAS^{ovx,vix}	0.213	0.529	14.945
SSVA MIDAS^{ovx}	0.208	0.517	14.262
SSVA MIDAS^{vix}	0.216	0.536	15.396
SSVA MIDAS^{rovx,rvix}	0.194	0.483	12.321
SSVA MIDAS^{rovx}	0.192	0.479	12.057
SSVA MIDAS^{rvix}	0.196	0.487	12.530
SSVA	0.195	0.486	12.455
SSVAg	0.185	0.461	11.098
SSV	0.208	0.519	14.337
SV	0.103	0.255	2.712
GARCH	0.075	0.188	1.011
AR1-RV	0.101	0.252	2.614

Note: ΔUMV is the the percentage utility gain measured as $(SR^2_{managed} - SR^2_{buyandhold})/SR^2_{buyandhold}$.

Conclusions

1. Through the proposed SV-MIDAS model, we found announcement events that shifted the oil volatility, such as:
    - US Change in Non-farm payrolls (a proxy for U.S. labor market strength and broader economic activity)
    - US DOE Crude oil inventories (a direct indicator of supply-demand balance in the oil market)
    - US FOMC Rate decision (a driver of interest rate policy, U.S. dollar strength, and global liquidity)
    - Sunday Open (reflects weekend liquidity gaps at market reopening)
2. Inclusion of MIDAS regressions improves the model performance.
3. More accurate out-of-sample volatility forecasting; favorable economic gains.

Thank you

A. Basistha and A. Kurov. The impact of monetary policy surprises on energy prices. *Journal of Futures Markets*, 35(1):87–103, 2015.

J. Bjursell, J. E. Gentle, and G. H. Wang. Inventory announcements, jump dynamics, volatility and trading volume in US energy futures markets. *Energy Economics*, 48: 336–349, 2015.

T. Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, pages 542–547, 1987.

M. W. Brandt and L. Gao. Macro fundamentals or geopolitical events? a textual analysis of news events for crude oil. *Journal of Empirical Finance*, 51:64–94, 2019.

H. Bu. Effect of inventory announcements on crude oil price volatility. *Energy Economics*, 46:485–494, 2014.

C. K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81 (3):541–553, 1994.

A. Chatrath, H. Miao, and S. Ramchander. Does the price of crude oil respond to macroeconomic news? *Journal of Futures Markets*, 32(6):536–559, 2012.

F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196, 2009.

R. Demirer and A. M. Kutan. The behavior of crude oil spot and futures prices around OPEC and SPR announcements: An event study perspective. *Energy Economics*, 32(6):1467–1476, 2010.

J. Elder and A. Serletis. Oil price uncertainty. *Journal of money, credit and banking*, 42(6):1137–1159, 2010.

J. P. Ferderer. Oil price volatility and the macroeconomy. *Journal of macroeconomics*, 18(1):1–26, 1996.

S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202, 1994.

L. Gao, S. Hitzemann, I. Shaliastovich, and L. Xu. Oil volatility risk. *Journal of Financial Economics*, 144(2):456–491, 2022.

E. Ghysels, P. Santa-Clara, and R. Valkanov. The midas touch: Mixed data sampling regression models. *UCLA Working Paper*, 2004.

E. Ghysels, P. Santa-Clara, and R. Valkanov. Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131 (1-2):59–95, 2006.

L. R. Glosten, R. Jagannathan, and D. E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801, 1993.

D. Hosszejni and G. Kastner. Modeling univariate and multivariate stochastic volatility in R with stochvol and factorstochvol. *Journal of Statistical Software*, 100(12): 1–34, 2021. doi: 10.18637/jss.v100.i12.

S. Jo. The effects of oil price uncertainty on global real economic activity. *Journal of Money, Credit and Banking*, 46(6):1113–1135, 2014.

D. R. Känzig. The macroeconomic effects of oil supply news: Evidence from OPEC announcements. *American Economic Review*, 111(4):1092–1125, 2021.

L. Kilian and C. Vega. Do energy prices respond to us macroeconomic news? a test of the hypothesis of predetermined energy prices. *Review of Economics and Statistics*, 93(2):660–671, 2011.

S. Kim, N. Shephard, and S. Chib. Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies*, 65(3):361–393, 1998.

R. López. The behaviour of energy-related volatility indices around scheduled news announcements: Implications for variance swap investments. *Energy Economics*, 72: 356–364, 2018.

T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression. *Journal of the american statistical association*, 83(404):1023–1032, 1988.

A. Moreira and T. Muir. Volatility-managed portfolios. *The Journal of Finance*, 72(4): 1611–1644, 2017.

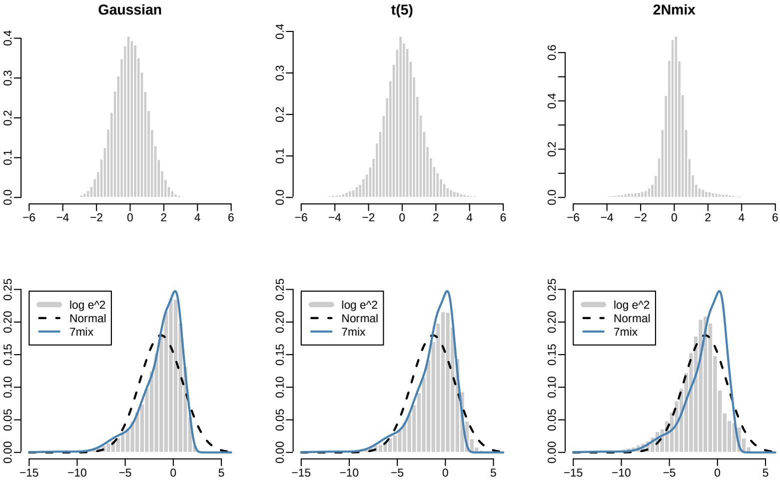
J. Noguera-Santaella. Geopolitics and the oil price. *Economic Modelling*, 52:301–309, 2016.

H. Schmidbauer and A. Rösch. OPEC news announcements: Effects on oil price expectation and volatility. *Energy Economics*, 34(5):1656–1663, 2012.

J. R. Stroud and M. S. Johannes. Bayesian modeling and forecasting of 24-hour high-frequency volatility. *Journal of the American Statistical Association*, 109(508): 1368–1384, 2014.

Y. Yang, J. Zhang, and S. Chen. Information effects of monetary policy announcements on oil price. *Journal of Commodity Markets*, 30:100268, 2023.

Bayesian inference



- Even better: non-parametric DPM errors (for the future).

No-MIDAS model: spike-and-slab vs Gaussian prior

The selected OLS events are such that p-values < 0.1, and MCMC events are such that the posterior inclusion probability is > 0.8.

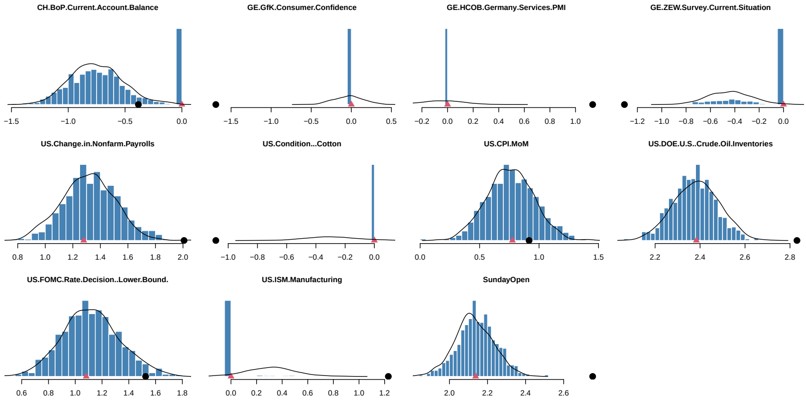


Figure: Posterior distributions of the estimated event parameters of the SV-no-MIDAS model, 10k iterations, 5k discarded as burn-in, thinned every 5th. Black dots are OLS, red triangles are posterior MCMC modes, black lines are Gaussian prior-based posterior distributions.