Beyond Fundamentals: Crashes and Bubbles in Global Oil Prices from a FIMARX model

Gilles Dufrénot, Céline Gimet

amse
école d'économie d'aix-marseille
aix-marseille school of economics

Working Papers / Documents de travail

Beyond Fundamentals: Crashes and Bubbles in Global Oil Prices from a FIMARX model

Gilles Dufrénot
Céline Gimet

WP 2026 - Nr 12

Beyond Fundamentals: Crashes and Bubbles in Global Oil Prices from a FIMARX model*

Gilles Dufrénot† Céline Gimet‡

Abstract

This paper studies whether large swings in global crude oil prices reflect observable fundamentals alone or also embody forward-looking speculative dynamics. We develop a stylised asset-pricing model in which fundamentalists and sentiment-driven speculators coexist, which motivates a two-component equilibrium price: a backward-looking fundamental part and a forward-looking speculative part. Long-memory in the dividend process further motivates the use of fractional filtering. The theoretical model motivates our empirical specification: we combine fractional filtering with a mixed causal-noncausal autoregressive model to distinguish persistent movements from expectation-driven price dynamics in the oil prices, allowing speculative pressures to be two-sided and generate both upward bubble-like episodes and downward crash-like dislocations. We find that, even conditional on a rich set of oil-market, macro-financial, and geopolitical determinants, crude oil prices retain a forward-looking component. The results suggest that major oil-price fluctuations are shaped not only by fundamentals, but also by expectation-driven forces that amplify boom-bust dynamics.

Keywords: Oil prices; Overpricing; Crashes; Noncausal models; Long-memory.
JEL Classification: C22; G12; F51; Q41.

*The project leading to this publication has received funding from the French government under the "France 2030" investment plan managed by the French National Research Agency (reference :ANR-17-EURE-0020) and from Excellence Initiative of Aix-Marseille University - A*MIDEX.

†Aix Marseille Univ, CNRS, AMSE, Sciences Po Aix, and CEPII, France. Email: gilles.dufrenot@sciencespo-aix.fr. Corresponding author.

‡Aix Marseille Univ, CNRS, AMSE, Sciences Po Aix, France. Email: celine.gimet@sciencespo-aix.fr

1 Introduction

Since the mid-2000s, crude oil prices have displayed recurrent boom-bust episodes of exceptional magnitude, repeatedly challenging the view that oil-price dynamics can be understood solely through contemporaneous changes in supply and demand. Over this period, the oil market has been exposed to a sequence of major global shocks, including repeated OPEC and OPEC+ production decisions, the collapse and rebound associated with the Covid-19 pandemic, the war in Ukraine, and renewed tensions in the Middle East. Yet the speed and amplitude of these price movements suggest that crude oil prices may at times have deviated ly from levels justified by observable fundamentals alone. This raises the following question question: when oil prices experience large and abrupt swings, do these movements simply reflect changing fundamentals and geopolitical risk, or are they also amplified by forward-looking expectations in a way that produces temporary overshooting and sharp reversals?

This question is important in the case of oil. Crude oil is simultaneously a physical input into global production, a financially traded asset, and a strategic object of geopolitical competition. Its price therefore emerges from the interaction of physical scarcity, macro-financial conditions, inventory dynamics, hedging demand, and anticipatory beliefs about future market conditions. In such an environment, speculative dynamics need not take the form of sustained overvaluation alone. They may also appear as pronounced downward departures from benchmark values, when periods of exuberance are followed by violent corrections or when adverse expectations become self-reinforcing. A framework centered exclusively on positive bubbles therefore captures only part of the phenomenon. What is needed instead is an empirical approach that can identify speculative pressures in both directions of the price cycle.

In this paper, we study whether global crude oil prices have exhibited such two-sided speculative deviations since the mid-2000s. Our objective is not limited to detecting bubbles in the narrow sense of persistent overpricing. Rather, we seek to recover a broader forward-looking speculative component that may generate both upward explosive dynamics and downward crash-like adjustments. Our empirical question is whether the evolution of global crude oil prices can be accounted for by observable fundamentals, financial conditions, and geopolitical shocks, or whether prices also contain a residual component reflecting anticipatory overreaction and temporary misalignment relative to a historically grounded benchmark.

Our contribution to the literature is fourfold.

First, we provide a theoretical foundation for the empirical specification used. We develop a stylised asset-pricing model in which crude oil is simultaneously a productive input and a financially traded asset, and two populations of agents coexist: fundamentalists, who form rational expectations conditional on observable market conditions, and specula-

tors, who add a sentiment bias to their forecasts. Under standard no-arbitrage conditions and CARA preferences, the equilibrium price decomposes additively into a backward-looking fundamental component (driven by the present discounted value of observable fundamentals) and a forward-looking speculative component (driven by the expected future path of sentiment). When the unobservable component of the dividend process is assumed to display long-memory, reflecting slow-moving structural forces such as energy transition dynamics and persistent geopolitical regimes, the log oil price inherits a fractional integration structure. Our model implies that the log Brent price, once filtered for long-memory, follows a mixed causal-noncausal autoregressive process with exogenous variables. This theoretical motivation is, to our knowledge, the first to derive the fractionally integrated MARX structure from explicit micro-founded equilibrium conditions in a commodity market setting.

Second, we extend the analysis of speculative dynamics in commodity markets beyond the conventional one-sided focus on positive bubbles. In the oil market, speculative behaviour may manifest itself not only through upward surges but also through downward overshooting and crash-like episodes. We therefore propose a conceptually broader notion of speculative deviation, encompassing both positive overpricing and negative underpricing relative to benchmark values.

Third, we bring mixed causal-noncausal time-series models to the study of global crude oil prices in a setting that explicitly controls for a rich set of observable drivers. This allows us to distinguish backward-looking persistence from forward-looking dynamics and to interpret the noncausal component as a reduced-form measure of speculative pressure. The theoretical model provides a structural interpretation for this reduced-form object: the noncausal component captures the contribution of speculative sentiment to the equilibrium price, conditional on observable fundamentals.

Fourth, we combine this framework with fractional filtering in order to separate long-memory persistence from shorter-run speculative behaviour. This matters empirically because oil prices are highly persistent, and without accounting for that persistence, forward-looking dynamics may be confounded with low-frequency dependence. The theoretical model motivates this two-step strategy by showing that long-memory originates in the dividend process and is transmitted to the log price independently of the speculative component.

Our econometric strategy proceeds as follows. We jointly estimate the long-memory and noncausal component of the log Brent price. The analysis focuses on Brent, the leading benchmark for the global crude oil market. The explanatory (fundamental) variables include the Brent-WTI spread, oil-specific and broader financial uncertainty (OVX and VIX), global macro-financial conditions (broad U.S. dollar index and U.S. 10-year Treasury yield), physical oil-market fundamentals (crude oil inventories), speculative positioning in futures markets (net positions of managed money traders from the CFTC

Commitments of Traders reports), and geopolitical tensions (Geopolitical Risk Index). These variables allow us to account for a broad range of observable determinants of oil prices, including both fundamentals and observable expectation-related forces, and each finds a natural structural interpretation in the theoretical model.

The remainder of the paper is organized as follows. Section 2 discusses the related literature reviews the related literature on oil-price bubbles and crashes. Section 3 presents the theoretical model that motivates our empirical specification. Section 4 presents the data. Section 5 introduces the fractional mixed causal-noncausal specification. Section 6 reports the estimation results and the recovered speculative component of Brent prices. Section 7 investigate the impact of shocks on oil prices through generalized impulse response functions. Finally, Section 8 concludes.

2 Related literature

This paper relates to three strands of the literature: the literature on the determinants of oil-price fluctuations, the literature on speculative dynamics and bubble-like behavior in oil markets, and the econometric literature on forward-looking and noncausal time-series dynamics. Relative to the earlier work, our contribution is to connect these literatures in a setting that explicitly allows for two-sided speculative deviations, that is, both upward bubble-like episodes and downward crash-like dislocations conditional on a broad set of observable oil-market, macro-financial, and geopolitical drivers.

A first strand of the literature studies oil-price dynamics through the lens of structural fundamentals and predictive indicators. Seminal papers emphasize the importance of distinguishing between different sources of oil-price movements, including supply shocks, global demand shocks, and precautionary or oil-specific demand shocks [Kilian, 2009, Hamilton, 2009, Kilian and Murphy, 2014]. More recent work has extended this perspective by showing that the predictive content of oil prices depends on a large set of external indicators capturing supply conditions, inventories, macroeconomic variables, financial conditions, and uncertainty measures. For instance, recent evidence suggests that combining oil-market fundamentals with macro-financial and uncertainty-related indicators improves the forecasting of oil-price fluctuations and helps explain periods of unusually large price variation [Peipei and James, 2024, Lee and Yahya, 2023, Gifuni, 2026]. This literature has ly enriched our understanding of the observable drivers of oil prices. At the same time, because its primary objective is to improve structural interpretation or forecasting performance, it leaves open the possibility that even after conditioning on a rich information set, oil prices may still display expectation-driven departures from benchmark values.

A second strand of the literature asks whether large oil-price swings reflect speculative pressure, explosive dynamics, or bubble-like behavior. Earlier work often framed this

issue as a debate between fundamentals and speculation. More recent contributions have shifted the focus toward the detection of recurrent explosive episodes and the role of extreme events, uncertainty, and investor behavior in triggering or amplifying such episodes. Wang and Su [2022] document multiple bubble periods in crude oil prices and argue that at least part of the observed dynamics cannot be fully reconciled with fundamentals alone. More recently, Chang [2024] show that extreme events, economic uncertainty, and speculation are systematically related to the occurrence of price bubbles in crude oil futures, especially around major global disruptions. Related work also highlights that geopolitical risk contains incremental predictive information for oil prices beyond standard oil-market and financial indicators, and may affect prices through both fundamental and speculative channels [Lee and Yahya, 2023, Zhang et al., 2024]. This more recent literature points to a view of oil-price dynamics in which expectation-driven amplification is not an exceptional feature confined to a few historical episodes, but a recurrent force that may interact with macroeconomic shocks, uncertainty, and geopolitical stress.

Our paper contributes to this literature in two respects. First, we move beyond the usual focus on positive bubbles and argue that speculative dynamics in oil markets are naturally two-sided. In practice, episodes of rapid appreciation are often followed by abrupt downward corrections, while periods of intense pessimism may generate sharp downward deviations that are no less speculative in nature than upward surges. Second, rather than testing explosiveness in isolation, we estimate a forward-looking speculative component conditional on a rich set of observable determinants. This allows us to distinguish more clearly between movements that are associated with measurable fundamentals and those that remain consistent with expectation-driven overshooting or undershooting.

A third strand of the literature develops econometric models capable of capturing forward-looking dynamics in asset prices. In noncausal and mixed causal-noncausal autoregressive models, current outcomes may depend not only on past innovations but also on future shocks, making these models appealing for the analysis of speculative episodes and boom-bust dynamics [Lanne and Saikkonen, 2011, Fries and Zakoian, 2019, Hecq et al., 2017]. While these models have mostly been applied to financial and macroeconomic time series, their logic is especially well suited to oil prices, which are shaped by anticipatory behavior, inventory demand, and rapidly changing expectations about future scarcity and geopolitical conditions. Our contribution to this econometric literature is to apply a mixed causal-noncausal framework to global crude oil prices in a setting that conditions on a broad range of observable determinants and explicitly separates long-memory persistence from shorter-run forward-looking speculative dynamics.

Finally, our paper is also related to recent work documenting persistence and long-range dependence in oil-price series and oil-price volatility [Kang and Yoon, 2013, Gil-Alana and Yaya, 2014]. This literature warns against conflating persistent low-frequency behavior with nonlinear or explosive short-run movements. Our empirical strategy ad-

dresses this concern directly by first filtering the long-memory component of the Brent price and then modeling the remaining dynamics within a mixed causal-noncausal specification. In doing so, we provide an empirical framework designed to distinguish persistent oil-price behavior from time-varying speculative pressure.

3 A theoretical model: asset pricing behavior in oil markets with heterogeneous beliefs

3.1 Main contributions relative to the literature

To motivate the empirical specification of the next section, we develop a stylised asset-pricing model in which crude oil is simultaneously a productive input and a financially traded asset. The model is intentionally kept simple (it is a toy model) whose purpose is not to provide a structural account of oil-market dynamics, but to show that the fractionally integrated mixed causal-noncausal (FIMARX) specification arises naturally when three economically plausible ingredients are combined: rational fundamentalists, sentiment-driven speculators, and slow-moving structural forces in the dividend process.

The model developed in this section draws on three distinct strands of the theoretical literature, each of which motivates one of its ingredients.

The first ingredient (heterogeneity in beliefs between fundamentalists and sentiment-driven speculators) is rooted in the noise trader literature initiated by De Long et al. [1990a], who show in an overlapping generations framework that irrational traders with erroneous stochastic beliefs can persistently move prices away from fundamental values, and in the heterogeneous agent models of Brock and Hommes [1998], who demonstrate that the interaction between fundamentalists and trend-extrapolating chartists generates endogenous boom-bust dynamics through an evolutionary switching mechanism. Our model departs from these two papers in two respects: we work with a CARA mean-variance setup that delivers a closed-form equilibrium price, and we apply the framework to the crude oil market rather than to an abstract risky asset. The extension of heterogeneous agent models to oil markets has been pursued empirically by Reitz and Westerhoff [2007] and Cifarelli and Paesani [2010], who find significant evidence of coexisting fundamentalist and chartist traders in oil price formation, and more recently by Cifarelli and Paesani [2021], who document bubble-like dynamics attributable to speculative herding in crude oil futures. Our theoretical framework proposes a micro-founded justification for the forward-looking speculative component that these studies identify empirically.

The second ingredient (the decomposition of the equilibrium price into a backward-looking fundamental component and a forward-looking speculative component) follows from standard no-arbitrage asset pricing [Campbell et al., 1997] combined with the addi-

tive structure of heterogeneous demands under CARA preferences. This additive decomposition is related to the framework of De Long et al. [1990b], who show that sentiment-driven demand generates a speculative premium in equilibrium that is proportional to the market share of noise traders and to the persistence of their beliefs. Our contribution is to map this premium explicitly onto the noncausal component of a MARX model, providing a structural interpretation for a component that is usually recovered as a purely statistical residual.

The third ingredient (long-memory in the unobservable component of the dividend process) connects our model to the literature on persistence in commodity prices and convenience yields. Moosa and Al-Loughani [1999] document long-memory in petroleum convenience yields and argue that slow mean-reversion reflects persistent structural imbalances between physical supply and demand. **?** provide evidence of fractional integration in historical oil price series and interpret it as reflecting the slow adjustment of global energy systems. Our theoretical model provides a microfoundation for this empirical regularity: long-memory in the dividend process arises from slow-moving structural forces (for instance, energy transition dynamics, persistent geopolitical regimes, and gradual belief updating about long-run scarcity) that are transmitted hyperbolically to the fundamental price and hence to the observed log oil price.

3.2 The individual agents' choices

We consider the crude oil barrel as an asset whose holding yields a productive service flow d_t (the ‘"dividend"), interpreted as the marginal value of oil as a production input. The spot price is p_t.

The economy runs in discrete time with an infinite horizon. There is a continuum of agents of mass 1, partitioned into two populations:

- fundamentalists (F) with mass $\lambda \in (0, 1)$, index $i \in [0, \lambda]$.
- speculators (S) with mass $(1 - \lambda)$, index $i \in (\lambda, 1]$.

Each agent maximises expected utility over one period:

$$\max_{\theta_t^i} \ \mathbb{E}_t^i\left[-e^{-\gamma W_{t+1}^i}\right], \tag{1}$$

where $\gamma > 0$ is the common absolute risk-aversion coefficient (CARA) and W_{t+1}^i denotes wealth at $t + 1$. The agent dislikes risk, and the intensity of that dislike is γ, which does not depend on the level of wealth. θ_t^i is the optimal position, i.e the number of oil contracts (or barrels) that agent i chooses to hold at date t.

Each agent holds a portfolio consisting of a risk-free asset (with return r) and a position θ_t^i in the oil market. Wealth evolves as:

$$W_{t+1}^i = (1+r)\, W_t^i + \theta_t^i [p_{t+1} + d_{t+1} - (1+r)\, p_t]\,. \tag{2}$$

W_{t+1}^i is wealth of agent i at the end of period t, after markets have cleared and payoffs have been received. $(1+r)\, W_t^i$ is what the agent would earn by investing all his/her wealth in the risk-free asset. If he/she held nothing in oil, he/she would simply earn the gross return $1+r$ on the initial wealth W_t^i. The term $p_{t+1}+d_{t+1}-(1+r)p_t$ is the net gain from the oil position, composed of p_{t+1} (the price at which the agent can sell the position next period), d_{t+1} (the dividend received while holding the asset) minus the opportunity cost $(1+r)p_t$ (by spending p_t to buy oil today, the agent forgoes the return $(1+r)p_t$, he/she could have earned on the risk-free asset).

For simplicity purpose, we assume no borrowing constraint, no transaction cost, single risky asset (oil) and no short-selling constraint ($\theta_t^i < 0$ is allowed). These assumptions keep the budget constraint linear in $\theta_t^i < 0$, to lead closed-form solution for the optimal demand.

Under CARA preferences, and assuming agent i perceives $p_{t+1} + d_{t+1}$ as normally distributed with mean μ_t^i and variance $(\sigma_t^i)^2$, the closed-form solution is:

$$\theta_t^i = \frac{\mu_t^i - (1+r)\, p_t}{\gamma\, (\sigma_t^i)^2}. \tag{3}$$

This is the Markowitz demand: the agent goes long whenever the expected gain exceeds the cost, weighted by perceived risk. The numerator is he expected profit from buying one barrel today. μ_t^i is what the agent expects to receive next period: the future price plus the dividend. $(1+r)p_t$ is what buying one barrel actually costs in terms of forgone return, i.e. the price today, plus the interest the agent could have earned by putting that money in the bank instead. $\sigma_t^i)^2$ is perceived risk. The agent is uncertain about the future price. The larger this uncertainty, the more cautious he/she is, and the smaller the position he/she takes. A very volatile oil market means a small position even if the expected profit looks attractive. A very risk-averse agent (large γ) takes a small position for any given expected profit. A less risk-averse agent (small γ is willing to take a larger bet.

3.3 Expectation formation

Both fundamentalists and speculators face the same price p_t, the same interest rate r and the same risk aversion γ. But they differ according to their expectations. The fundamentalists expect $\mu_t^F = \mathbb{E}_t^F[p_{t+1} + d_{t+1}]$, based on observable fundamentals. The speculators expect $\mu_t^S = \mu_t^F + \xi_t$ where ξ_t is a sentiment bias.

Fundamentalists: rational expectations

Fundamentalists observe the filtration $\mathcal{F}_t = \sigma\{p_s, d_s, X_s : s \leq t\}$ and compute:

$$\mathbb{E}_t^F[p_{t+1} + d_{t+1}] = \mathbb{E}_t[p_{t+1}^*] + \mathbb{E}_t[d_{t+1}]. \tag{4}$$

The fundamental price is defined by the forward recursion:

$$p_t^* = \frac{1}{1+r}\,\mathbb{E}_t^F[p_{t+1}^* + d_{t+1}], \tag{5}$$

whose solution is the present discounted value of all expected future dividends[1]:

$$p_t^* = \sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k \mathbb{E}_t^F[d_{t+k}]. \tag{6}$$

Speculators: biased expectations

Speculators do not know the true dividend process. They use the heuristic rule:

$$\mathbb{E}_t^S[p_{t+1} + d_{t+1}] = \mathbb{E}_t^F[p_{t+1} + d_{t+1}] + \xi_t, \tag{7}$$

where ξ_t is an expectation bias, or speculative sentiment. It may be positive (over-optimism) or negative (over-pessimism), generating two-sided deviations from the fundamental value. We therefore assume that speculators do not ignore fundamentals, but that they add a sentiment term on top of them. The case $\xi_t = 0$ for all t collapses to pure rational expectations.

We now describe the sentiment process. We postulate that ξ_t is determined by three components:

$$\xi_t = \underbrace{\mu_t}_{\text{trend extrapolation}} + \underbrace{\sum_{m=1}^{M} \beta_m\, x_{m,t}}_{\text{reaction to observables}} + \underbrace{v_t}_{\text{pure innovation}}, \tag{8}$$

Component 1: Trend extrapolation. Speculators extrapolate past price changes:

$$\mu_t = \sum_{j=1}^{J} \omega_j\, \Delta p_{t-j}, \quad \omega_j > 0, \quad \sum_j \omega_j < 1. \tag{9}$$

This rule captures the momentum behaviour empirically documented among commodity

[1] The fundamental price p_t^* is obtained by iterating the one-period pricing equation $p_t^* = \frac{1}{1+r}\mathbb{E}_t^F[p_{t+1}^* + d_{t+1}]$ forward N periods, applying the law of iterated expectations, and imposing the transversality condition $\lim_{N\to\infty}\left(\frac{1}{1+r}\right)^N \mathbb{E}_t^F[p_{t+N}^*] = 0$, which rules out self-fulfilling price paths disconnected from dividends. The solution is the present discounted value of all expected future dividends:

$$p_t^* = \sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k \mathbb{E}_t^F[d_{t+k}].$$

traders.

Component 2: Reaction to observables. Speculators respond to the same variables as fundamentalists (for instance some variables X_t: GPR, VIX, OVX, dollar index, interest rates, CFTC positions, ...) but with coefficients β_m that may differ from the structural coefficients α_m of fundamentalists. They may over-react ($|\beta_m| > |\alpha_m|$) or under-react.

Component 3: Pure sentiment innovation. v_t is an i.i.d. innovation with heavy tails, independent of fundamentals. Its non-Gaussian distribution is crucial for the identification of the MARX model, as detailed below. This leads an ARMA representation of sentiment

Substituting the extrapolation rule and assuming that each $x_{m,t}$ follows its own dynamics, the reduced-form representation is:

$$\xi_t = \sum_{j=1}^{s} \psi_j \, \xi_{t-j} + \sum_{m=1}^{M} \sum_{k=0}^{q_m} \beta_{m,k} \, x_{m,t-k} + v_t, \tag{10}$$

or, in operator notation:

$$\phi_\xi(L) \, \xi_t = B(L) \, X_t + v_t, \quad \phi_\xi(L) = 1 - \psi_1 L - \cdots - \psi_s L^s. \tag{11}$$

The roots of $\phi_\xi(z) = 0$ are assumed to lie outside the unit circle, so that ξ_t is stationary. Explosive episodes are transitory.

3.4 Market equilibrium

The net supply of oil available to financial traders is $\bar{Q}$ (inelastic in the short run). Equilibrium requires:

$$\lambda \, \theta_t^F + (1 - \lambda) \, \theta_t^S = \bar{Q}. \tag{12}$$

Substituting the individual demands, we can write:

$$\frac{\lambda \Big[\mathbb{E}_t^F [p_{t+1} + d_{t+1}] - (1+r) p_t \Big] + (1-\lambda) \Big[\mathbb{E}_t^S [p_{t+1} + d_{t+1}] - (1+r) p_t \Big]}{\gamma \sigma^2} = \bar{Q}. \tag{13}$$

Using $\mathbb{E}_t^S[\cdot] = \mathbb{E}_t^F[\cdot] + \xi_t$ leads:

$$\frac{\mathbb{E}_t^F [p_{t+1} + d_{t+1}] + (1-\lambda) \, \xi_t - (1+r) \, p_t}{\gamma \sigma^2} = \bar{Q}. \tag{14}$$

Rearranging yields the fundamental equilibrium equation, we get:

$$p_t = \frac{1}{1+r} \Big[\mathbb{E}_t^F [p_{t+1} + d_{t+1}] + (1-\lambda) \, \xi_t \Big] - \underbrace{\frac{\gamma \sigma^2}{1+r} \, \bar{Q}}_{\Pi_t \text{ : risk premium}}. \tag{15}$$

Therefore, the equilibrium price is the sum of three terms:

1. fundamental expectation $\frac{1}{1+r}\mathbb{E}_t^F[p_{t+1}+d_{t+1}]$: the discounted value of observable fundamentals.

2. speculative premium $\frac{1-\lambda}{1+r}\xi_t$: the distortion introduced by speculators, proportional to their market share.

3. Risk premium $-\Pi_t$: the discount arising from inelastic supply and risk aversion.

3.5 The aggregate dynamics of prices

Advancing the equilibrium equation one period and taking expectations:

$$\mathbb{E}_t^F[p_{t+1}] = \frac{1}{1+r}\,\mathbb{E}_t^F\Big[\mathbb{E}_{t+1}^F[p_{t+2}+d_{t+2}] + (1-\lambda)\,\xi_{t+1}\Big] - \Pi_t. \tag{16}$$

By the law of iterated expectations ($\mathbb{E}_t^F[\mathbb{E}_{t+1}^F[\cdot]] = \mathbb{E}_t^F[\cdot]$):

$$\mathbb{E}_t^F[p_{t+1}] = \frac{1}{1+r}\Big[\mathbb{E}_t^F[p_{t+2}+d_{t+2}] + (1-\lambda)\,\mathbb{E}_t^F[\xi_{t+1}]\Big] - \Pi_t. \tag{17}$$

Substituting back and iterating N times:

$$p_t = \sum_{k=1}^{N}\Big(\frac{1}{1+r}\Big)^k \mathbb{E}_t^F[d_{t+k}] + \frac{1-\lambda}{1+r}\sum_{k=0}^{N-1}\Big(\frac{1}{1+r}\Big)^k \mathbb{E}_t^F[\xi_{t+k}] - \Pi_t \sum_{k=0}^{N-1}\Big(\frac{1}{1+r}\Big)^k + \Big(\frac{1}{1+r}\Big)^N \mathbb{E}_t^F[p_{t+N}]. \tag{18}$$

Imposing the transversality condition : $\lim_{N\to\infty}\Big(\frac{1}{1+r}\Big)^N \mathbb{E}_t^F[p_{t+N}] = 0$, we get:

$$p_t = \underbrace{\sum_{k=1}^{\infty}\Big(\frac{1}{1+r}\Big)^k \mathbb{E}_t^F[d_{t+k}]}_{p_t^*\,:\,\text{fundamental value}} + \underbrace{\frac{1-\lambda}{1+r}\sum_{k=0}^{\infty}\Big(\frac{1}{1+r}\Big)^k \mathbb{E}_t^F[\xi_{t+k}]}_{B_t\,:\,\text{speculative component}} - \underbrace{\frac{\Pi_t}{r}}_{\text{constant risk premium}}. \tag{19}$$

An explicit computation of B_t can be obtained. For the AR(s) sentiment process $\phi_\xi(L)\,\xi_t = v_t + B(L)\,X_t$, the conditional expectation satisfies:

$$\mathbb{E}_t^F[\xi_{t+k}] = \sum_{j=1}^{s}\psi_j\,\mathbb{E}_t^F[\xi_{t+k-j}] + \sum_{m,\ell}\beta_{m,\ell}\,\mathbb{E}_t^F[x_{m,t+k-\ell}]. \tag{20}$$

Denoting $\delta \equiv \frac{1}{1+r}$, the discounted sum becomes:

$$\sum_{k=0}^{\infty}\delta^k\,\mathbb{E}_t^F[\xi_{t+k}] = \frac{1}{\phi_\xi\Big(\frac{L^{-1}}{1+r}\Big)}\,\xi_t. \tag{21}$$

For the AR(1) case ($\phi_\xi(L) = 1 - \psi_1 L$), this simplifies to:

$$B_t = \frac{(1-\lambda)\,\xi_t}{1 + r - \psi_1}. \tag{22}$$

The speculative component B_t is amplified by the persistence of sentiment: the closer ψ_1 is to 1, the larger B_t relative to ξ_t. Formally, $\partial B_t / \partial \psi_1 > 0$.

3.6 Dividend dynamics and long-memory

The oil dividend d_t captures the convenience yield of holding physical crude oil (the marginal economic value of a barrel as a productive input and a buffer against supply disruptions). It is decomposed into an observable component, driven by physical market conditions (inventories), uncertainty (OVX, VIX), global financial conditions (dollar index, Treasury yield), geopolitical risk, and supply policy (OPEC production), and an unobservable residual η_t. This residual reflects slow-moving structural forces (energy transition dynamics, persistent geopolitical regimes, and gradual belief updating about long-run scarcity, that are not fully spanned by observable indicators). These forces generate hyperbolic rather than exponential decay in the autocorrelation of η_t, motivating the fractional integration assumption $(1-L)^d \eta_t = \varepsilon_t^F$ with $d \in \left(0, \frac{1}{2}\right)$, which ensures stationarity while allowing for the long-range dependence documented in oil price series.

The oil dividend therefore aggregates several channels:

$$d_t = \alpha_0 + \sum_{m=1}^{M} \alpha_m \, x_{m,t} + \eta_t, \tag{23}$$

where η_t is a long-memory residual. We assume:

$$(1-L)^d \, \eta_t = \varepsilon_t^F, \quad \varepsilon_t^F \sim \mathcal{N}(0, \sigma_F^2), \quad d \in \left(0, \tfrac{1}{2}\right). \tag{24}$$

The fundamental price p_t^* inherits the long-memory of η_t. The contribution of η_t to p_t^* is[2]:

$$p_t^{*,\eta} = \sum_{k=1}^{\infty} \delta^k \, \mathbb{E}_t[\eta_{t+k}] = \frac{\delta}{1-\delta} \cdot \frac{1}{(1-L)^d} \, \varepsilon_t^F. \tag{25}$$

The last equality uses the MA(∞) representation of $(1-L)^{-d}$. The fractional filter

[2]The contribution of η_t to the fundamental price is obtained in two steps. First, since $(1-L)^d \, \eta_t = \varepsilon_t^F$, the process η_t admits the MA(∞) representation $\eta_t = (1-L)^{-d} \, \varepsilon_t^F = \sum_{j=0}^{\infty} \pi_j \, \varepsilon_{t-j}^F$, where the coefficients $\pi_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}$ decay hyperbolically as j^{d-1} rather than exponentially. Second, the conditional expectation of η_{t+k} given $\mathcal{F}_t$ is $\mathbb{E}_t[\eta_{t+k}] = \sum_{j=k}^{\infty} \pi_j \, \varepsilon_{t+k-j}^F$, so that the discounted sum becomes $\sum_{k=1}^{\infty} \delta^k \, \mathbb{E}_t[\eta_{t+k}] = \frac{\delta}{1-\delta} \cdot (1-L)^{-d} \, \varepsilon_t^F$, where $\delta = \frac{1}{1+r}$ and the operator $(1-L)^{-d}$ acts as a persistence multiplier: a shock ε_t^F today continues to affect $p_t^{*,\eta}$ over an arbitrarily long horizon, with weights that decay at the hyperbolic rate k^{d-1} rather than at the exponential rate of a standard AR process. This is the mechanism through which long-memory in dividends is transmitted to the fundamental price.

$(1-L)^{-d}$ acts as a persistence multiplier: a shock ε_t^F to dividends generates effects on p_t^* that decay hyperbolically, in contrast to the exponential decay of an ordinary AR process.

As a consequance, the log-price is $I(d)$. Indeed, taking logarithms ($y_t = \log p_t \approx \log p_t^* + B_t/p_t^*$ by a first-order Taylor expansion) yields:

$$(1-L)^d\, y_t \approx (1-L)^d \log p_t^* + (1-L)^d \frac{B_t}{p_t^*}. \tag{26}$$

The first term is $I(0)$ by construction; the second inherits the short-run persistence of B_t. Hence:

$$(1-L)^d\, y_t \;\equiv\; \tilde{y}_t \;\sim\; I(0). \tag{27}$$

The fractional filter is thus *theoretically justified* by the long-memory structure of shocks to oil dividends.

3.7 The reduced-form equation : a fractional MARX Model

Combining the equilibrium, sentiment, and dividend equations, we derive the following equation for the filtered log-price[3]:

$$\tilde{y}_t = \varphi(L)\,\tilde{y}_t + \phi(L^{-1})^{-1}\tilde{y}_t + \alpha + \Gamma(L)\,X_t + \varepsilon_t, \tag{28}$$

which, upon rearranging, gives:

$$\varphi(L)\;\phi(L^{-1})\;(1-L)^d\; y_t = \alpha + \Gamma(L)\,X_t + \varepsilon_t. \tag{29}$$

This is a fractional MAR(r,s) model with exogenous variables (we call it a FIMARX(r,s).

The causal polynomial $\varphi(L)$ captures the backward-looking dynamics of p_t^*:

$$\varphi(L) \;\longleftrightarrow\; \text{persistence of fundamentals } (d_t,\, X_t). \tag{30}$$

[3]The filtered log-price equation is derived as follows. Starting from the equilibrium price equation $p_t = p_t^* + B_t - \Pi_t$, taking logarithms and applying a first-order Taylor expansion around the steady state gives $y_t \approx \log p_t^* + B_t/p_t^* - \Pi_t/p_t^*$, so that the log-price inherits both the backward-looking dynamics of p_t^* and the forward-looking dynamics of B_t. Applying the fractional filter $(1-L)^{\hat{d}}$ to both sides and denoting $\tilde{y}_t = (1-L)^{\hat{d}}\, y_t$ removes the long-memory component, leaving a stationary series. The filtered fundamental component $(1-L)^{\hat{d}} \log p_t^*$ is then approximated by a causal AR polynomial $\varphi(L)\,\tilde{y}_t$, which captures the backward-looking persistence of oil prices driven by observable fundamentals X_t and their lagged effects $\Gamma(L)\,X_t$. The filtered speculative component $(1-L)^{\hat{d}} B_t/p_t^*$ is approximated by the inverse noncausal polynomial $\phi(L^{-1})^{-1}\tilde{y}_t$, which captures the forward-looking dynamics of the sentiment process ξ_t: because $B_t = \frac{1-\lambda}{1+r}\sum_{k=0}^{\infty}\delta^k\,\mathbb{E}_t[\xi_{t+k}]$ depends on all future expected values of sentiment, it is naturally represented by a polynomial in the forward operator L^{-1}. The residual $\varepsilon_t = \varepsilon_t^F + (1-\lambda)\,v_t$ aggregates the Gaussian fundamental innovation ε_t^F and the heavy-tailed speculative innovation v_t, resulting in a non-Gaussian error term whose distribution motivates the use of Cauchy or stable densities in estimation. Rearranging then yields the MARX(r,s) specification $\varphi(L)\,\phi(L^{-1})\,(1-L)^d\, y_t = \alpha + \Gamma(L)\,X_t + \varepsilon_t$.

The noncausal polynomial $\phi(L^{-1})$ captures the forward-looking dynamics of B_t:

$$\phi(L^{-1}) = \phi_\xi(L^{-1}) \iff \text{persistence of sentiment } \xi_t. \tag{31}$$

The coefficients $\Gamma(L)$ aggregate the effects of observable fundamentals on both components, with a direct reading:

$$\Gamma_m^{\text{causal}} = \alpha_m \text{ (fundamental effect)}, \qquad \Gamma_m^{\text{noncausal}} = \beta_m \text{ (speculative over-reaction)}. \tag{32}$$

3.8 Distribution of Residuals

The residual innovation ε_t is the superposition of ε_t^F (Gaussian) and v_t (heavy-tailed):

$$\varepsilon_t = \varepsilon_t^F + (1-\lambda)\, v_t. \tag{33}$$

Even though ε_t^F is Gaussian, the presence of heavy-tailed v_t renders ε_t *non-Gaussian*. This justifies the use of a Cauchy or stable distribution in estimation, and non-Gaussianity is a necessary condition for identification of a MARX model.

4 Data : variables and sources

4.1 Main variables of the model

Table 1 reports the variables used in the empirical specification, together with their economic role, main source, frequency, suggested transformation, and data accessibility.

The dependent variable is the logarithm of the Brent spot price,

$$y_t = \log(\text{Brent}_t),$$

where Brent is used as the benchmark for the global crude oil price.

The explanatory variables are chosen to capture five main channels:

1. market segmentation
2. oil-specific uncertainty,
3. global financial conditions,
4. physical oil-market fundamentals,
5. geopolitical pressure.

The vector of explanatory variables is

$$X_t = \begin{pmatrix} Spread_t \\ OVX_t \\ VIX_t \\ USD_t \\ i_t^{10Y} \\ Inv_t \\ Spec_t \\ GPR_t \end{pmatrix},$$

where:

$$\begin{aligned} Spread_t &= \text{Brent}_t - \text{WTI}_t, \\ OVX_t &= \text{oil-specific implied volatility index}, \\ VIX_t &= \text{global financial volatility index}, \\ USD_t &= \text{broad U.S. dollar index}, \\ i_t^{10Y} &= \text{U.S. 10-year Treasury yield}, \\ Inv_t &= \text{U.S. crude oil inventories}, \\ Spec_t &= \text{net speculative positions in crude oil futures}, \\ GPR_t &= \text{geopolitical risk index}. \end{aligned}$$

The choice of explanatory variables is guided by the idea that global crude oil prices are jointly determined by regional price segmentation, oil-specific uncertainty, global financial conditions, physical market fundamentals, speculative pressure, and geopolitical tensions. Each variable is therefore included to isolate a distinct transmission channel affecting the log of the Brent price.

The spread between Brent and WTI is introduced to account for short-run segmentation between the global crude oil market and the U.S. oil market. Because Brent serves as the main international benchmark while WTI is more exposed to U.S.-specific conditions, variation in the spread may arise from regional dislocations rather than from changes in global oil-market fundamentals. Including this variable therefore helps separate truly global price movements from benchmark-specific distortions. In the context of bubble identification, this distinction is essential, since an apparent increase in Brent prices may otherwise be mistaken for evidence of a global speculative episode when it merely reflects a local deviation in WTI.

The OVX is included as a forward-looking measure of oil-specific uncertainty. Because it is derived from option prices, it reflects market expectations about future crude-oil volatility rather than past realized volatility alone. This makes it relevant in a mixed

causal–noncausal framework, where expectations and anticipatory behavior are central. Controlling for the OVX helps distinguish movements in Brent driven by changing uncertainty in the oil market from movements that are interpreted as speculative deviations.

The VIX is included to control for broad financial-market uncertainty. Unlike the OVX, which is specific to oil, the VIX captures changes in aggregate risk appetite and global financial sentiment across asset classes. Including this variable helps separate oil-price movements that reflect oil-market conditions from those associated with generalized episodes of financial stress. This distinction is useful because crude oil prices may rise or fall sharply during periods of global turbulence even in the absence of oil-specific speculative pressures.

The broad U.S. dollar index is included because crude oil is priced internationally in U.S. dollars. Exchange-rate movements therefore affect both the dollar price of oil and the purchasing power of non-U.S. consumers. A stronger dollar is associated with downward pressure on oil prices in dollar terms, all else equal. Controlling for the dollar helps ensure that changes in Brent are not mistakenly attributed to speculative dynamics when they are in fact driven by global currency movements.

The U.S. 10-year Treasury yield is introduced as a proxy for global macro-financial conditions. Long-term interest rates summarize movements in discount rates, growth expectations, and broad financial conditions, all of which may influence commodity prices. Including this variable helps absorb variation in Brent that is attributable to changes in the global macro-financial environment rather than to oil-market-specific speculative forces.

OECD onland commercial oil stocks are included as a measure of physical oil-market fundamentals. They provide a broad indicator of inventory conditions across the main advanced oil-consuming economies and therefore capture the degree of tightness or abundance in the physical oil market. Higher inventories are generally associated with weaker market conditions or excess supply, whereas lower inventories indicate tighter conditions. Including this variable is useful in a study of speculative deviations because it helps distinguish price movements driven by observable physical-market conditions from those that may reflect forward-looking speculative pressures. Since the series is available only at quarterly frequency, it is used primarily as a medium-run robustness control rather than as a baseline high-frequency regressor.

Net speculative positions are included as a direct proxy for speculative activity in crude oil futures markets. Constructed from the Commitments of Traders reports, this variable captures the extent to which non-commercial traders are, on balance, positioned for further price increases or decreases. It is therefore useful in a study of bubble dynamics, since it provides an observable measure of speculative pressure that can be compared with the forward-looking component identified by the model.

The Geopolitical Risk Index is included because oil prices are especially sensitive to geopolitical tensions. Conflicts, sanctions, and disruptions in major producing or transit

regions can affect crude prices through both expected supply losses and precautionary-demand effects. Controlling for geopolitical risk is therefore essential in order to avoid attributing to speculative dynamics price movements that instead reflect economically meaningful shifts in global political conditions.

Variables related to OPEC supply conditions are included as controls for changes in the global supply regime. Because OPEC remains a central actor in world oil production, changes in output targets or realized production may have first-order effects on the global crude market. Including such variables helps distinguish speculative price movements from those that are associated with observable changes in global oil supply.

When available, the Brent term spread is included to capture information embedded in the futures curve. The slope of the term structure reflects expectations about future market conditions, storage incentives, and relative scarcity between nearby and deferred delivery. This variable is especially useful in a bubble application because it helps assess whether increases in spot prices are consistent with the broader pricing of oil across maturities or instead point to a stronger speculative component concentrated in the nearby market.

4.2 Construction of the variables

Endogenous variable

The dependent variable is the logarithm of the Brent spot price,

$$y_t = \log(\text{Brent}_t),$$

where Brent is used as the benchmark for the global crude oil market. Daily Brent spot prices are obtained from the U.S. Energy Information Administration (EIA), as distributed through FRED. The original series is reported in U.S. dollars per barrel, is not seasonally adjusted, and has daily frequency. This series is aggregated to weekly frequency by taking the last available daily observation within each calendar week.

Brent–WTI spread

The Brent–WTI spread is defined as

$$Spread_t = \text{Brent}_t - \text{WTI}_t.$$

It measures the price differential between the global crude benchmark and the main U.S. benchmark and is intended to capture temporary segmentation between the world oil market and the U.S. oil market. The WTI series is taken from the EIA/FRED series. Both original series are observed daily and reported in U.S. dollars per barrel. They are aggregated to weekly frequency.

Table 1: Variables used in the MARX model for the global crude oil price

Variable	Economic role	Source	Frequency	Transformation
$\log(\text{Brent}_t)$	Benchmark for the global crude oil price	FRED / EIA	Daily	Log-level; possibly detrended
$\log(\text{WTI}_t)$	U.S. benchmark; useful to build the Brent–WTI spread	FRED / EIA	Daily	Log-level or used only in spread construction
$Spread_t = \text{Brent}_t - \text{WTI}_t$	Captures segmentation between the global crude market and the U.S. oil market	Constructed from Brent and WTI spot prices	Daily	Level, standardized value, or first difference
OVX_t	Oil-specific implied volatility; proxy for oil-market uncertainty and speculative tension	FRED / Cboe	Daily	Level or first difference
VIX_t	Global financial stress and broad market uncertainty	FRED / Cboe	Daily	Level or first difference
$\log(USD_t)$	Broad U.S. dollar effect; oil is priced in dollars	FRED	Daily	Log-level or log-difference
i_t^{10Y}	Global macro-financial conditions and discount-rate channel	FRED	Daily	Level or first difference
Inv_t	Physical oil-market fundamentals through commercial crude inventories	EIA Weekly Petroleum Status Report	Weekly	Carry forward to daily; level or weekly change
$Spec_t$	Net speculative pressure in oil futures markets	CFTC Commitments of Traders	Weekly	Net long minus net short; optionally standardized
GPR_t	Geopolitical risk affecting global oil prices	Caldara–Iacoviello GPR database	Monthly / Daily	Level or first difference; aligned to daily if needed
$OPEC_t$	Global oil supply conditions; robustness variable	OPEC Monthly Oil Market Report	Monthly	Level or first difference
TS_t^{Brent}	Brent term-structure slope; expectations, storage tensions, and market structure	Brent futures data provider / exchange source	Daily	Nearby-minus-deferred spread

In our specification, the spread is kept in levels because this preserves its direct interpretation as a benchmark-specific price gap.

Oil-specific uncertainty: OVX

Oil-specific uncertainty is measured using the Cboe Crude Oil ETF Volatility Index, retrieved from FRED. The OVX is an estimate of the expected 30-day volatility of crude oil as priced by the United States Oil Fund (USO). The original data are daily closing observations. It is aggregated to a weekly frequency. Economically, this variable captures forward-looking uncertainty specific to the oil market. In the empirical analysis, we consider either the level of the index, OVX_t, or its first difference, $\Delta OVX_t = OVX_t - OVX_{t-1}$. The level reflects the overall degree of oil-market uncertainty, while the difference captures changes in expected oil-market volatility.

Broad financial uncertainty: VIX

Broad financial uncertainty is measured using the Cboe Volatility Index, obtained from FRED. The VIX measures the market expectation of near-term volatility conveyed by stock index option prices. The original data are daily closing observations and are aggregated to weekly frequency. This variable captures global financial stress rather than oil-specific uncertainty. As with the OVX, we use either the level, VIX_t, or the first difference, $\Delta VIX_t = VIX_t - VIX_{t-1}$.

Broad U.S. dollar index

Because crude oil is priced internationally in U.S. dollars, we include a broad measure of the external value of the dollar. Specifically, we use the FRED series `DTWEXBGS`, labeled "Nominal Broad U.S. Dollar Index". The original series is daily, not seasonally adjusted, and expressed as an index with January 2006 = 100. FRED and the Federal Reserve describe it as a weighted average of the foreign exchange value of the U.S. dollar against a broad group of major U.S. trading partners. The series is aggregated to weekly frequency. The variable is included in logarithms, $usd_t = \log(USD_t)$, and, in alternative specifications, in log-differences: $\Delta usd_t = \log(USD_t) - \log(USD_{t-1})$.

U.S. 10-year Treasury yield

Global macro-financial conditions are proxied by the U.S. 10-year Treasury yield, taken from FRED . The original series is daily, expressed in percent per annum, and corresponds to the market yield on U.S. Treasury securities at 10-year constant maturity. The series is aggregated to weekly frequency. This variable is denoted by i_t^{10Y}. In the baseline specification, it is used in level form. As a robustness check, we also consider first differences, $\Delta i_t^{10Y} = i_t^{10Y} - i_{t-1}^{10Y}$

Crude oil inventories

Physical oil-market fundamentals are proxied by OECD onland commercial oil stocks from the OPEC Monthly Oil Market Report. This series measures commercially available oil inventories across OECD economies and thus provides a broad indicator of market tightness. The original data are quarterly and reported as end-of-period closing stock levels. Let $Inv_t^{OECD,raw}$ denote the raw quarterly series. To match the weekly Brent sample, the series is converted to weekly frequency by assigning to each week the most recently available quarterly observation, so that the quarterly value is held constant across all weeks until the next quarterly release. In the empirical analysis, we use either

$$Inv_t^{OECD} = Inv_t^{OECD,raw},$$

or

$$\Delta^{(q)} Inv_t^{OECD} = Inv_t^{OECD,raw} - Inv_{t-1}^{OECD,raw}.$$

Speculative pressure from CFTC data

Speculative pressure is measured using the Disaggregated Commitments of Traders (COT) Futures-Only reports published by the U.S. Commodity Futures Trading Commission (CFTC). We use the contract *Crude Oil, Light Sweet – New York Mercantile Exchange* and focus on the *Managed Money* category. The disaggregated reports split reportable open-interest positions into four classifications, including Producer/Merchant, Swap Dealers, Managed Money, and Other Reportables. The original COT data are weekly. Historical files for the Disaggregated Futures-Only reports are available from September 2009 onward. Economically, the managed-money position is intended to measure speculative directional pressure in the oil futures market. The idea is straightforward: if managed-money traders hold ly more long than short positions, speculative pressure is net bullish; if they hold more short than long positions, speculative pressure is net bearish.

Let MM_t^{Long} and MM_t^{Short} denote managed-money long and short positions, and let OI_t denote total open interest. We construct two alternative measures. The first is the net speculative position:

$$Spec_t^{net} = MM_t^{Long} - MM_t^{Short}.$$

The second scales the net position by the size of the market:

$$Spec_t^{OI} = \frac{MM_t^{Long} - MM_t^{Short}}{OI_t}.$$

The normalized measure $Spec_t^{OI}$ is used in the baseline specification because it captures directional speculative pressure relative to total market activity.

Geopolitical risk

Geopolitical conditions are measured using the Geopolitical Risk (GPR) Index of Caldara and Iacoviello. The official GPR website provides the data and the associated paper documents both monthly and daily versions of the index (we aggregate the daily variable to weekly frequency). The GPR as a measure of adverse geopolitical events and associated risks constructed from newspaper coverage. This variable is intended to capture international political tensions that may affect oil prices through supply-risk, precautionary-demand, or financial-market channels. In the baseline specification, we use the global GPR index. Denoting this series by GPR_t, we consider either the level, or its first difference, $\Delta GPR_t = GPR_t - GPR_{t-1}$.

OPEC supply conditions

As an additional medium-run control, we consider OPEC crude oil production from the OPEC Monthly Oil Market Report (MOMR). The original source is monthly. The MOMR is the official OPEC report on oil-market developments, and its archived chapters report OPEC crude oil production according to available secondary sources. This variable is intended to capture changes in the global oil-supply regime. Let $OPEC_t^{raw}$ denote the original monthly production series. Since the series is observed at monthly frequency, it is converted to weekly frequency by assigning to each week the most recently available monthly observation, so that the monthly value is held constant across all weeks within the corresponding month. In the empirical analysis, the variable is used in levels, or in monthly differences,$\Delta OPEC_t = OPEC_t^{raw} - OPEC_{t-1}^{raw}$.

5 A fractionally integrated mixed causal-noncausal model

Since crude oil prices display strong persistence and potentially long-range dependence, the empirical specification allows for a fractional integration component jointly estimated with the mixed causal–noncausal dynamics. The baseline specification is written as:

$$\varphi(L)\,\phi(L^{-1})\,(1-L)^d\,y_t = \alpha + \Gamma(L)\,X_t + \varepsilon_t, \tag{34}$$

where L is the lag operator, $(1-L)^d$ is the fractional differencing operator with $d \in (0, \frac{1}{2})$, $\varphi(L)$ is the causal autoregressive polynomial, $\phi(L^{-1})$ is the noncausal polynomial, and X_t is the vector of exogenous variables introduced in Section 3.

The causal polynomial is

$$\varphi(L) = 1 - \varphi_1 L - \cdots - \varphi_r L^r, \tag{35}$$

and the noncausal polynomial is

$$\phi(L^{-1}) = 1 - \psi_1 L^{-1} - \cdots - \psi_s L^{-s}. \tag{36}$$

The exogenous block is introduced through a distributed-lag polynomial,

$$\Gamma(L)\, X_t = \Gamma_0 X_t + \Gamma_1 X_{t-1} + \cdots + \Gamma_q X_{t-q}, \tag{37}$$

which allows the explanatory variables to affect oil prices not only contemporaneously but also with delay. The effects of inventories, speculative positioning, geopolitical risk, or macro-financial variables are unlikely to be fully absorbed within the same trading period. Allowing for lags in the exogenous block also reduces the risk that delayed effects of observable fundamentals are spuriously captured by the noncausal component and then misinterpreted as speculative dynamics.

5.1 Joint Estimation

All parameters, i.e. the fractional integration parameter d, the causal coefficients $\{\varphi_1, \ldots, \varphi_r\}$, the noncausal coefficients $\{\psi_1, \ldots, \psi_s\}$, the distributed-lag coefficients $\{\Gamma_0, \ldots, \Gamma_q\}$, and the distributional parameters of ε_t are estimated jointly by maximum likelihood. The full parameter vector is

$$\Theta = (d,\, \varphi_1, \ldots, \varphi_r,\, \psi_1, \ldots, \psi_s,\, \Gamma_0, \ldots, \Gamma_q,\, \alpha,\, \sigma)\,, \tag{38}$$

where σ denotes the scale parameter of the innovation distribution.

The joint log-likelihood is

$$\ell(\Theta) = \sum_{t=1}^{T} \log f(\varepsilon_t(\Theta);\, \sigma)\,, \tag{39}$$

where

$$\varepsilon_t(\Theta) = \varphi(L)\, \phi(L^{-1})\, (1-L)^d\, y_t - \alpha - \Gamma(L)\, X_t \tag{40}$$

and $f(\cdot\,; \sigma)$ is the density of the innovation process, specified in Section **??**.

The fractional differencing operator $(1-L)^d$ is applied within the likelihood using the binomial expansion

$$(1-L)^d = \sum_{k=0}^{\infty} \pi_k\, L^k, \qquad \pi_0 = 1, \quad \pi_k = \frac{\Gamma(k-d)}{\Gamma(-d)\,\Gamma(k+1)}, \quad k \geq 1, \tag{41}$$

truncated at a finite lag K in practice. Joint estimation has two advantages over a two-step approach. First, it avoids the generated-regressor problem that would arise if a pre-filtered series were used as the dependent variable in a second step: by treating d as a free

parameter in the likelihood, inference on the causal and noncausal coefficients correctly accounts for the uncertainty in d. Second, it allows the data to determine simultaneously the degree of long-run persistence and the relative importance of backward-looking versus forward-looking short-run dynamics, without imposing an a priori separation between the two.

The joint optimisation of (39) over Θ is non-trivial because the noncausal component introduces dependence on future observations, so that the likelihood cannot be evaluated recursively in the forward direction alone. Following Lanne and Saikkonen [2011] and Fries and Zakoian [2019], we evaluate the full-sample likelihood by exploiting the two-sided representation of the model: for a given Θ, the residuals $\varepsilon_t(\Theta)$ are computed for all t using both past and future values of y_t and X_t, and the log-likelihood is then evaluated as the sum (39). Numerical optimisation is performed using a quasi-Newton algorithm with analytical gradients, and standard errors are obtained from the outer product of gradients estimator of the information matrix.

The recovered speculative component is constructed from the estimated parameters as

$$\widehat{B}_t = \sum_{j=0}^{J} \widehat{\delta}_j \, \widehat{z}_{t+j}, \tag{42}$$

where $\widehat{z}_t = \widehat{\alpha} + \widehat{\Gamma}(L)X_t + \widehat{\varepsilon}_t$, the coefficients $\widehat{\delta}_j$ are defined by the inverse expansion of the estimated noncausal polynomial

$$\frac{1}{\widehat{\phi}(L^{-1})} = \sum_{j=0}^{\infty} \widehat{\delta}_j \, L^{-j}, \tag{43}$$

and J is a finite truncation horizon. A positive value of $\widehat{B}_t$ corresponds to an upward speculative deviation, while a negative value corresponds to a downward crash-like adjustment.

5.2 Distributional Assumptions and Estimation under NonGaussian Innovations

The residuals ε_t are the innovations of the mixed causal–noncausal dynamics, once long memory and observable exogenous effects have been absorbed. A requirement for the identification of the MARX model is that ε_t be non-Gaussian: under Gaussianity, the causal and noncausal components are not separately identified because the likelihood is symmetric with respect to the direction of time [Lanne and Saikkonen, 2011]. The theoretical model provides a natural justification for this assumption: as shown in Section 3, the innovation $\varepsilon_t = \varepsilon_t^F + (1-\lambda)\, v_t$ is a mixture of a Gaussian fundamental shock and a heavy-tailed speculative shock, which generically produces a non-Gaussian marginal distribution.

Three distributional specifications are considered.

Cauchy distribution. Under the Cauchy specification,

$$\varepsilon_t \sim \mathcal{C}(0, \sigma), \tag{44}$$

with density

$$f(\varepsilon_t; \sigma) = \frac{1}{\pi\sigma}\left[1 + \left(\frac{\varepsilon_t}{\sigma}\right)^2\right]^{-1}. \tag{45}$$

The Cauchy distribution is the canonical choice in the MARX literature [Lanne and Saikkonen, 2011, Fries and Zakoian, 2019] because its very heavy tails — the variance does not exist — are consistent with the occasional large innovations observed in oil markets during crisis episodes, and because the resulting log-likelihood has a simple closed form that facilitates numerical optimisation. The joint log-likelihood under the Cauchy specification is

$$\ell_{\mathcal{C}}(\Theta) = -T\log(\pi\sigma) - \sum_{t=1}^{T} \log\left[1 + \left(\frac{\varepsilon_t(\Theta)}{\sigma}\right)^2\right]. \tag{46}$$

The score with respect to σ is

$$\frac{\partial\, \ell_{\mathcal{C}}}{\partial \sigma} = -\frac{T}{\sigma} + \frac{2}{\sigma}\sum_{t=1}^{T} \frac{(\varepsilon_t/\sigma)^2}{1 + (\varepsilon_t/\sigma)^2}, \tag{47}$$

and the score with respect to any dynamic parameter $\theta \in \Theta$ is

$$\frac{\partial\, \ell_{\mathcal{C}}}{\partial \theta} = \frac{2}{\sigma^2}\sum_{t=1}^{T} \frac{\varepsilon_t(\Theta)}{1 + (\varepsilon_t(\Theta)/\sigma)^2} \cdot \frac{\partial\, \varepsilon_t}{\partial \theta}. \tag{48}$$

The Cauchy likelihood down-weights large residuals relative to least squares, conferring robustness to outliers, which is valuable in a sample that includes the 2008–2009 and 2020 oil-price collapses.

Stable distribution. To allow for tail behaviour that differs from the Cauchy case, a stable specification is also considered:

$$\varepsilon_t \sim S_\alpha(\sigma, \beta, \mu), \tag{49}$$

where $\alpha \in (0, 2]$ is the stability index, $\beta \in [-1, 1]$ is the skewness parameter, $\sigma > 0$ is the scale, and $\mu \in \mathbb{R}$ is the location. Stable distributions generalise the Cauchy ($\alpha = 1$, $\beta = 0$) and Gaussian ($\alpha = 2$) cases, and allow for asymmetric tail behaviour through $\beta \neq 0$. Because the density of a stable distribution does not have a closed form in general,

the log-likelihood

$$\ell_S(\Theta) = \sum_{t=1}^{T} \log f_S(\varepsilon_t(\Theta); \alpha, \beta, \sigma, \mu) \tag{50}$$

is evaluated numerically using the fast Fourier inversion of the characteristic function

$$\mathbb{E}\left[e^{iu\varepsilon_t}\right] = \exp\left\{i\mu u - \sigma^\alpha |u|^\alpha \left[1 - i\beta \operatorname{sgn}(u) \tan\left(\tfrac{\pi\alpha}{2}\right)\right]\right\}, \quad \alpha \neq 1. \tag{51}$$

The stable specification is relevant when the data suggest asymmetric tail behaviour — for instance, if crash innovations are systematically larger than bubble innovations, as the oil-price record might suggest. Estimation is performed by maximum likelihood using a grid search over α and β to initialise the quasi-Newton optimisation.

Laplace distribution. As a robust alternative, the model is also estimated under a Laplace specification:

$$\varepsilon_t \sim \text{Laplace}(0, b), \tag{52}$$

with density

$$f(\varepsilon_t; b) = \frac{1}{2b} \exp\left(-\frac{|\varepsilon_t|}{b}\right). \tag{53}$$

The corresponding log-likelihood is

$$\ell_{\mathcal{L}}(\Theta) = -T \log(2b) - \frac{1}{b} \sum_{t=1}^{T} |\varepsilon_t(\Theta)|, \tag{54}$$

whose maximisation is equivalent to least absolute deviations (LAD) estimation of the dynamic parameters for a given b. The Laplace distribution has lighter tails than the Cauchy but heavier tails than the Gaussian, placing it between the two in terms of robustness to outliers. It is the natural choice when one wishes to retain the robustness properties of LAD estimation while remaining within a likelihood framework that permits joint inference on d and the dynamic parameters. The score with respect to a dynamic parameter θ is

$$\frac{\partial \ell_{\mathcal{L}}}{\partial \theta} = \frac{1}{b} \sum_{t=1}^{T} \operatorname{sgn}(\varepsilon_t(\Theta)) \cdot \frac{\partial \varepsilon_t}{\partial \theta}, \tag{55}$$

which is bounded and well-defined except at $\varepsilon_t = 0$. In practice, a smoothed approximation $\operatorname{sgn}(x) \approx x/\sqrt{x^2 + \eta^2}$ with small $\eta > 0$ is used to ensure differentiability.

Model selection. The choice among the three distributional specifications is guided by the Akaike and Bayesian information criteria evaluated at the joint maximum likelihood estimates. A likelihood ratio test of the Cauchy restriction $\alpha = 1$, $\beta = 0$ within the stable family is also reported. In practice, the Cauchy specification has been found to perform well for oil-price data [Lanne and Saikkonen, 2011, Fries and Zakoian, 2019], and we treat

it as the baseline, using the stable and Laplace specifications as robustness checks.

Joint estimation of the full parameter vector Θ requires evaluating the fractional differencing operator $(1-L)^d$ for each candidate value of d during numerical optimisation. In principle, this operator involves an infinite filter,

$$(1-L)^d y_t = \sum_{k=0}^{\infty} \pi_k \, y_{t-k}, \qquad \pi_k = \frac{\Gamma(k-d)}{\Gamma(-d)\,\Gamma(k+1)},$$

whose coefficients π_k decay hyperbolically as k^{-d-1} and are therefore never exactly zero for any finite truncation horizon. Two practical difficulties arise. First, a direct truncation of the filter at a finite lag K,

$$(1-L)^d y_t \approx \sum_{k=0}^{K} \pi_k \, y_{t-k},$$

introduces a truncation bias that depends on both K and d: the slower the hyperbolic decay (i.e. the smaller d), the more lags are required to achieve a given level of approximation accuracy, and the greater the loss of initial observations. Second, repeated evaluation of the binomial coefficients π_k for each candidate d during optimisation is computationally costly when K is large.

To address these difficulties, we follow Sowell [1992] and Velasco [1999] and adopt a frequency-domain approximation based on the cosine representation of the fractional filter. The identity is that the squared gain of the fractional filter at frequency $\omega \in [0, \pi]$ is

$$\left|(1-e^{-i\omega})^d\right|^2 = \left[2\sin\left(\frac{\omega}{2}\right)\right]^{2d},$$

so that the log-spectral density of $(1-L)^d y_t$ satisfies

$$\log f_{\tilde{y}}(\omega) = \log f_{\varepsilon}(\omega) + 2d \log\left[2\sin\left(\frac{\omega}{2}\right)\right],$$

where $f_{\varepsilon}(\omega)$ is the spectral density of the short-memory component. In the time domain, this corresponds to approximating the fractional filter by a cosine series truncated at order M:

$$(1-L)^d y_t \approx y_t + \sum_{k=1}^{M} c_k(d) \, y_{t-k},$$

where the cosine-based coefficients are given by

$$c_k(d) = \frac{2}{\pi} \int_0^{\pi} \left[2\sin\left(\frac{\omega}{2}\right)\right]^d \cos(k\omega)\, \mathrm{d}\omega, \qquad k = 1, \ldots, M.$$

This approximation has two advantages over direct truncation. First, the cosine coefficients $c_k(d)$ decay faster than the binomial coefficients π_k, so that a relatively small M achieves a good approximation across the full range $d \in (0, \frac{1}{2})$. Second, for a fixed grid of

frequency points $\{\omega_j\}_{j=1}^N$, the cosine transforms can be precomputed once and reused for each candidate d during optimisation, reducing the computational burden ly. In practice, we set $M = \lfloor T^{0.8} \rfloor$ following the bandwidth recommendation of Velasco [1999], and verify that the results are insensitive to moderate changes in M.

5.3 Measuring the contribution of exogenous variables to the causal and noncausal components

An attractive feature of the empirical specification is that, conditional on the estimated dynamic parameters, the model is linear in the exogenous block. This makes it possible to decompose both the causal and the noncausal components into variable-specific contributions.

Recall that the second-step specification is

$$\phi(L)\varphi(L^{-1})\widetilde{y}_t = \alpha + \Gamma(L)X_t + \varepsilon_t,$$

where $X_t = (x_{1,t}, \ldots, x_{M,t})'$ denotes the vector of exogenous variables and

$$\Gamma(L)X_t = \sum_{m=1}^{M} \Gamma_m(L)x_{m,t},$$

with

$$\Gamma_m(L) = \gamma_{m,0} + \gamma_{m,1}L + \cdots + \gamma_{m,q_m}L^{q_m}. \tag{56}$$

Defining

$$z_t = \alpha + \Gamma(L)X_t + \varepsilon_t,$$

and using the inverse causal and noncausal filters,

$$\frac{1}{\phi(L)} = \sum_{j=0}^{\infty} \pi_j^* L^j, \qquad \frac{1}{\varphi(L^{-1})} = \sum_{j=0}^{\infty} \delta_j L^{-j},$$

the filtered process may be written as

$$\widetilde{y}_t = C_t + B_t, \tag{57}$$

where

$$C_t = \sum_{j=0}^{\infty} \pi_j^* z_{t-j}, \qquad B_t = \sum_{j=0}^{\infty} \delta_j z_{t+j}.$$

Because z_t is linear in the exogenous variables, the two components can themselves be

decomposed additively. Writing

$$z_t = \alpha + \sum_{m=1}^{M} \Gamma_m(L) x_{m,t} + \varepsilon_t,$$

it follows that

$$C_t = C_t^{(\alpha)} + \sum_{m=1}^{M} C_t^{(m)} + C_t^{(\varepsilon)}, \qquad B_t = B_t^{(\alpha)} + \sum_{m=1}^{M} B_t^{(m)} + B_t^{(\varepsilon)}, \tag{58}$$

with variable-specific contributions

$$C_t^{(m)} = \sum_{j=0}^{\infty} \pi_j^* \Gamma_m(L) x_{m,t-j}, \tag{59}$$

and

$$B_t^{(m)} = \sum_{j=0}^{\infty} \delta_j \Gamma_m(L) x_{m,t+j}. \tag{60}$$

Using (56), these expressions can be written explicitly as

$$C_t^{(m)} = \sum_{j=0}^{\infty} \pi_j^* \sum_{k=0}^{q_m} \gamma_{m,k} x_{m,t-j-k}, \tag{61}$$

and

$$B_t^{(m)} = \sum_{j=0}^{\infty} \delta_j \sum_{k=0}^{q_m} \gamma_{m,k} x_{m,t+j-k}. \tag{62}$$

Equations (59)-(62) provide a time-varying attribution of the backward-looking and forward-looking components to each observable covariate. In particular, $C_t^{(m)}$ measures the contribution of variable x_m to the causal component, whereas $B_t^{(m)}$ measures its contribution to the noncausal forward-looking component. In empirical work, all infinite sums are truncated, so that the estimated contributions are

$$\widehat{C}_t^{(m)} = \sum_{j=0}^{J_C} \widehat{\pi}_j^* \widehat{\Gamma}_m(L) x_{m,t-j}, \qquad \widehat{B}_t^{(m)} = \sum_{j=0}^{J_B} \widehat{\delta}_j \widehat{\Gamma}_m(L) x_{m,t+j}, \tag{63}$$

for suitably chosen truncation horizons J_C and J_B.

These series can be summarized in several ways. A first option is to study their historical paths directly. A second is to compute average signed or absolute contributions over a subsample $\mathcal{T}$:

$$\overline{B}^{(m)} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \widehat{B}_t^{(m)}, \qquad \overline{|B|}^{(m)} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \left| \widehat{B}_t^{(m)} \right|, \tag{64}$$

with analogous measures for the causal component.

A third option is to consider variance-based importance measures such as

$$Share_B^{(m)} = \frac{\text{Var}(\widehat{B}_t^{(m)})}{\text{Var}(\widehat{B}_t)}, \tag{65}$$

although such ratios should be interpreted cautiously because the different variable-specific contributions are generally correlated.

A useful complementary exercise is to consider counterfactual decompositions. For any variable x_m, define the reduced-form term obtained by shutting down that variable:

$$z_t^{(-m)} = \alpha + \sum_{\ell \neq m} \Gamma_\ell(L) x_{\ell,t} + \varepsilon_t. \tag{66}$$

The associated counterfactual noncausal component is

$$B_t^{(-m)} = \sum_{j=0}^{\infty} \delta_j z_{t+j}^{(-m)}, \tag{67}$$

and similarly for the causal component. The difference

$$\Delta B_t^{(m)} = B_t - B_t^{(-m)} \tag{68}$$

then measures the contribution of variable x_m to the forward-looking component in counterfactual terms. The same logic applies to the causal component.

Because the explanatory variables may be correlated, attribution based on sequential inclusion or exclusion may depend on the order in which variables are introduced. To address this issue, we also consider a Shapley decomposition. Let $\mathcal{M} = \{1, \ldots, M\}$ denote the set of explanatory variables, and let $S \subseteq \mathcal{M}$ be any subset. For a given subset S, define

$$z_t^{(S)} = \alpha + \sum_{m \in S} \Gamma_m(L) x_{m,t} + \varepsilon_t, \tag{69}$$

and the corresponding noncausal component

$$B_t^{(S)} = \sum_{j=0}^{\infty} \delta_j z_{t+j}^{(S)}. \tag{70}$$

Let $V_B(S)$ denote a scalar summary of the component generated by subset S, for example

$$V_B(S) = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \left| B_t^{(S)} \right| \tag{71}$$

or

$$V_B(S) = \text{Var}\left(B_t^{(S)} \right). \tag{72}$$

The Shapley value of variable m for the noncausal component is then

$$\Phi_B^{(m)} = \sum_{S \subseteq \mathcal{M} \setminus \{m\}} \frac{|S|!(M-|S|-1)!}{M!} \left[V_B(S \cup \{m\}) - V_B(S) \right]. \tag{73}$$

An analogous definition applies to the causal component:

$$\Phi_C^{(m)} = \sum_{S \subseteq \mathcal{M} \setminus \{m\}} \frac{|S|!(M-|S|-1)!}{M!} \left[V_C(S \cup \{m\}) - V_C(S) \right]. \tag{74}$$

The Shapley decomposition has two advantages. First, it is invariant to the order in which variables are introduced. Second, it satisfies the additivity property

$$\sum_{m=1}^{M} \Phi_B^{(m)} = V_B(\mathcal{M}) - V_B(\varnothing), \qquad \sum_{m=1}^{M} \Phi_C^{(m)} = V_C(\mathcal{M}) - V_C(\varnothing), \tag{75}$$

so that the total contribution is exactly allocated across variables. In practice, when M is large, the full 2^M decomposition may be approximated by averaging marginal contributions over a large number of random permutations.

These decompositions provide complementary perspectives on the role of each exogenous variable. The direct linear decomposition delivers time-varying contributions to the causal and noncausal components. The counterfactual analysis shows how these components would evolve if a particular variable were switched off. The Shapley decomposition, in turn, yields an order-invariant summary of the extent to which each variable contributes to the backward-looking and forward-looking parts of oil-price dynamics. Throughout, these measures should be interpreted as model-based econometric attributions rather than as structural causal effects in the policy sense.

6 Main findings

6.1 Estimation Results

Table 2 reports our estimation results corresponding to the optimal models selected. We use information criteria to select the best model among many with maximum lags equal to 5 for the MAR(r,s) component and to 4 for the exogenous variables.

Evidence of long-memory dynamics. An interesting finding is the stability of the long-memory parameter $\hat{d}$ across the three distributional assumptions (Cauchy, Laplace and α-stable distributions). The point estimates are, respectively, 0.3184 (Cauchy), 0.3190 (Laplace), and 0.3185 (Stable), all significant at the 1% level with z-statistics above 66, and all within the stationarity region $d \in (0, \frac{1}{2})$. A value of $d \approx 0.32$ implies that shocks to the crude oil price decay hyperbolically at rate $k^{d-1} \approx k^{-0.68}$ rather than

exponentially, so that even relatively modest shocks generate persistence over several years. This is consistent with potential slow-moving structural forces, such as energy transition dynamics, persistent geopolitical regimes, gradual belief updating about long-run scarcity, that our theoretical model (Section 3) places at the origin of long-memory in the dividend process.

Identification of the mixed causal-noncausal component. The causal AR coefficients $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ are all estimated in the range $[0.055, 0.056]$, significant at the 5% or 1% level under the Cauchy and Stable distributions. Their near-equality suggests that the backward-looking component of oil-price dynamics is driven by a quasi-uniform distributed lag over the most recent three weeks, rather than by strong first-order momentum alone. The noncausal coefficient $\hat{\psi}_1 \approx 0.056$ is similarly small in magnitude but statistically significant at the 5% level (Cauchy) or 10% level (Stable). While small, this coefficient identifies a forward-looking component in oil-price dynamics that cannot be reduced to an AR(3) process on observable fundamentals. Formally, $\hat{\psi}_1 > 0$ implies that the current log-price is partly determined by the one-step-ahead continuation of the process, consistent with the anticipatory speculative mechanism described in our theoretical model.

A notable feature of the Laplace specification is that the standard errors on $\hat{\phi}_i$ and on the speculative-position coefficient are inflated by a factor of approximately 1,373 relative to their Cauchy counterparts. This reflects a near-flat likelihood surface in those directions under the Laplace density, which has lighter tails than the Cauchy and is unable to identify the causal-noncausal decomposition from the same information set. This identification failure under the Laplace distribution, combined with its higher AIC (1,176 vs. 540 for Cauchy and 291 for Stable), provides strong statistical evidence against the Laplace as a distributional assumption for oil-price innovations.

Exogenous determinants. Among the observable fundamentals, the log U.S. dollar index exerts the largest and most robust effect: the sum of significant distributed-lag coefficients is 0.370 (Cauchy) to 0.381 (Laplace), highly significant across all three specifications. The positive sign may seem, at first, surprising (since a stronger dollar is conventionally associated with lower oil prices in dollar terms), but is consistent with the interpretation that the USD index here proxies for global macro-financial conditions: during periods of dollar strength associated with U.S. monetary policy tightening or global risk appetite, oil prices may initially co-move with broad financial asset prices before the exchange-rate channel dominates. The result is also consistent with the *financialisation* hypothesis: if oil is increasingly traded as a financial asset, its price may correlate positively with other dollar-denominated assets during specific regimes.

Speculative positions (Spec./OI) enter with a large and highly significant positive

coefficient of 1.187 under both Cauchy and Stable distributions ($z > 12$), but cannot be identified under the Laplace distrirubtion. So, even after controlling for the noncausal component $\hat{B}_t$, net long positions of managed-money traders carry incremental positive information for the log Brent price. This suggests that observable speculative activity and the latent forward-looking component are distinct but complementary dimensions of speculative pressure in the oil market.

The oil-specific volatility index (OVX) enters with a negative and significant coefficient (-0.002 to -0.004), consistent with the standard uncertainty channel: higher expected oil-market volatility is associated with precautionary demand reduction and downward pressure on spot prices. The Brent-WTI spread has a small but significant positive effect, capturing episodes of regional market segmentation in which the global benchmark trades at a premium relative to U.S.-specific conditions.

By contrast, the VIX and the 10-year Treasury yield are not robustly significant: the yield is significant under Laplace but not under Cauchy or Stable distributions, suggesting that global macro-financial conditions operate mainly through the dollar index, while broad financial stress is partially absorbed by the OVX. Inventory changes and the GPR enter with small but significant coefficients, consistent with their role as complementary proxies for physical market conditions and geopolitical supply risk, respectively.

Distribution selection and model fit. The Stable distribution dominates the Cauchy and Laplace specifications by a wide margin: the log-likelihood improves from -223.1 (Cauchy) to -97.3 (Stable), an improvement of 125.7 log-likelihood units for one additional parameter, implying a likelihood ratio statistic of 251.4 compared to a $\chi^2_{(1)}$ critical value of 6.64 at the 1% level. The AIC and BIC confirm this ranking (291 and 511 for Stable vs. 540 and 756 for Cauchy). The estimated stability index $\hat{\alpha}_{stable} = 1.702$ (SE $= 0.165$, $z = 10.3$) is significantly different from both $\alpha = 1$ (Cauchy) and $\alpha = 2$ (Gaussian), placing the innovation distribution in a region of intermediate tail thickness. Economically, this means that oil-price innovations are heavier tailed than Gaussian but lighter tailed than Cauchy, consistent with the view that extreme events (Covid collapse, Ukraine war spike) are infrequent but not as extreme as a pure Cauchy process would imply. The scale parameter $\hat{\sigma} \approx 0.422$ is nearly identical across Cauchy and Stable, providing a further consistency check.

6.2 Fitted values and residual diagnostics

Panel (a) of Figure 1 shows that the three fitted series closely track the observed log(Brent) price throughout the sample, with RMSE in the range $[0.051, 0.053]$ and R^2 of 0.976. The near-perfect superposition of the three distributional specifications illustrates a property of the MARX framework: the choice of distributional assumption affects inference on the

Table 2: Estimation Results : Fractionally Integrated Mixed Causal- Noncausal Model

	Cauchy		Laplace		Stable	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE
A. Dynamic parameters						
d (long memory)	0.3184***	(0.0039)	0.3190***	(0.0009)	0.3185***	(0.0048)
ϕ_1 (causal AR)	0.0558***	(0.0216)	0.0558	(29.6582)	0.0558**	(0.0268)
ϕ_2 (causal AR)	0.0551**	(0.0216)	0.0552	(29.6582)	0.0552**	(0.0268)
ϕ_3 (causal AR)	0.0554**	(0.0216)	0.0554	(29.6582)	0.0554**	(0.0268)
ψ_1 (noncausal AR)	0.0562**	(0.0244)	0.0562		0.0562*	(0.0302)
α (intercept)	-1.4246***	(0.0115)	-1.4235***	(0.0033)	-1.4244***	(0.0142)
B. Exogenous variables (sum of significant coefficients, lags 0-4)						
Brent-WTI Spread	0.0123***	(0.0028)	0.0136***	(0.0012)	0.0187***	(0.0020)
OVX (oil volatility)	-0.0019***	(0.0006)	-0.0037***	(0.0004)	-0.0024***	(0.0007)
VIX (financial stress)	-0.0017	(0.0011)	0.0004	(0.0006)	-0.0019	(0.0013)
log(USD index)	0.3694***	(0.0054)	0.3814***	(0.0015)	0.3719***	(0.0067)
10Y Treasury Yield	-0.0032	(0.0091)	0.0054***	(0.0018)	-0.0014	(0.0113)
Speculative Positions/OI	1.1867***	(0.1708)	-		1.1878***	(0.2122)
GPR	0.0007***	(0.0001)	-0.0001*	(0.0001)	0.0008***	(0.0002)
C. Distributional parameters						
α_{stable}	-		-		1.7020***	(0.1650)
σ (scale)	0.42378		1.00000		0.42236	
D. Model fit						
$-\ell$	223.050		541.036		97.330	
AIC	540.10		1176.07		290.66†	
BIC	755.65		1391.62		510.80†	
n			725			
k	47		47		48	

Notes: Estimation of a fractionally integrated mixed causal–noncausal (MARX) model with exogenous variables on weekly Brent spot prices, sample April 2012 – April 2026. Model order: $r = 3$ (causal), $s = 1$ (noncausal), $q = 4$ (exogenous lags), $M_{\text{trunc}} = 247$ (fractional filter truncation).

Standard errors in parentheses, computed via diagonal-Hessian approximation. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

For exogenous variables, only the sum of coefficients significant at the 10% level (across lags 0–4) is reported; the associated standard error is $\sqrt{\sum_{\ell:\,\hat{\gamma}_{m,\ell}\text{ sig.}} \widehat{\text{SE}}_\ell^2}$.

† Best-fitting distribution (lowest AIC and BIC).

causal-noncausal decomposition but leaves the in-sample fit of the price level essentially unchanged, since the fitted values absorb the residuals by construction.

Several features of the residual panel (b) deserve attention. First, the residuals are broadly stationary throughout the sample, with no visible trend or persistent drift, confirming that the long-memory filter $(1-L)^{\hat{d}}$ has successfully absorbed the low-frequency persistence of log(Brent). Second, the largest residuals cluster around the four crisis episodes: the 2015-16 OPEC production decision, the Covid-19 collapse of April 2020, the post-pandemic rebound into the Ukraine war peak of 2022, and the 2023 correction. These episodes correspond to the shaded bands in the figure, suggesting that the model performs well in normal periods but that crisis-time innovations are too large to be fully captured by the estimated dynamics and exogenous variables alone. This is consistent with the finding from the diagnostic figures that autocorrelation in the standardised residuals remains significant (Ljung–Box p-values well below 0.01 at all lags up to 20), suggesting that the causal order $r = 3$ may be insufficient to absorb all backward-looking persistence in the data and that an extension to $r \in \{4, 5\}$ could improve residual whiteness at the cost of additional parameters. Third, the three distributional specifications produce nearly identical residual series in panel (b), reinforcing the conclusion that distributional choice matters for the decomposition of the process into causal and noncausal components, but has limited impact on the overall fit of the price level.

Distributional diagnostics. The three histograms in the left column of Figure 2 confirm that the standardised residuals $\hat{\varepsilon}_t/\hat{\sigma}$ display markedly heavy tails relative to the Gaussian benchmark: both the Cauchy and Stable densities provide a visually superior fit compared to the normal distribution, with the Stable density offering the best compromise between tail thickness and central tendency. The Laplace histogram reveals a moderate leptokurtosis, but the density's exponential tails are clearly inadequate for the most extreme observations, which are associated with the 2020 Covid collapse and the 2022 Ukraine war spike.

The ACF panels (centre column) reveal significant positive autocorrelation in the standardised residuals at all three lags tested, with first-order autocorrelation $\hat{\rho}_1 \approx 0.52$ under all three distributions. The persistence in the residuals suggests that the causal polynomial $\phi(L)$ of order $r = 3$ does not fully absorb the backward-looking dynamics of the filtered series. Two interpretations are possible. First, a higher causal order $r \in \{4, 5\}$ might be needed, despite not being selected by the grid-search AIC/BIC on the fast filter used for order selection. Second, the residual autocorrelation may partly reflect the interaction between the fractional filter and the AR dynamics: for $d \approx 0.32$, the filtered series retains medium-run persistence, and a short AR polynomial may be insufficient to capture the full structure of those dynamics. Either way, the residual autocorrelation should be interpreted as a specification caveat rather than a fundamental flaw: the main

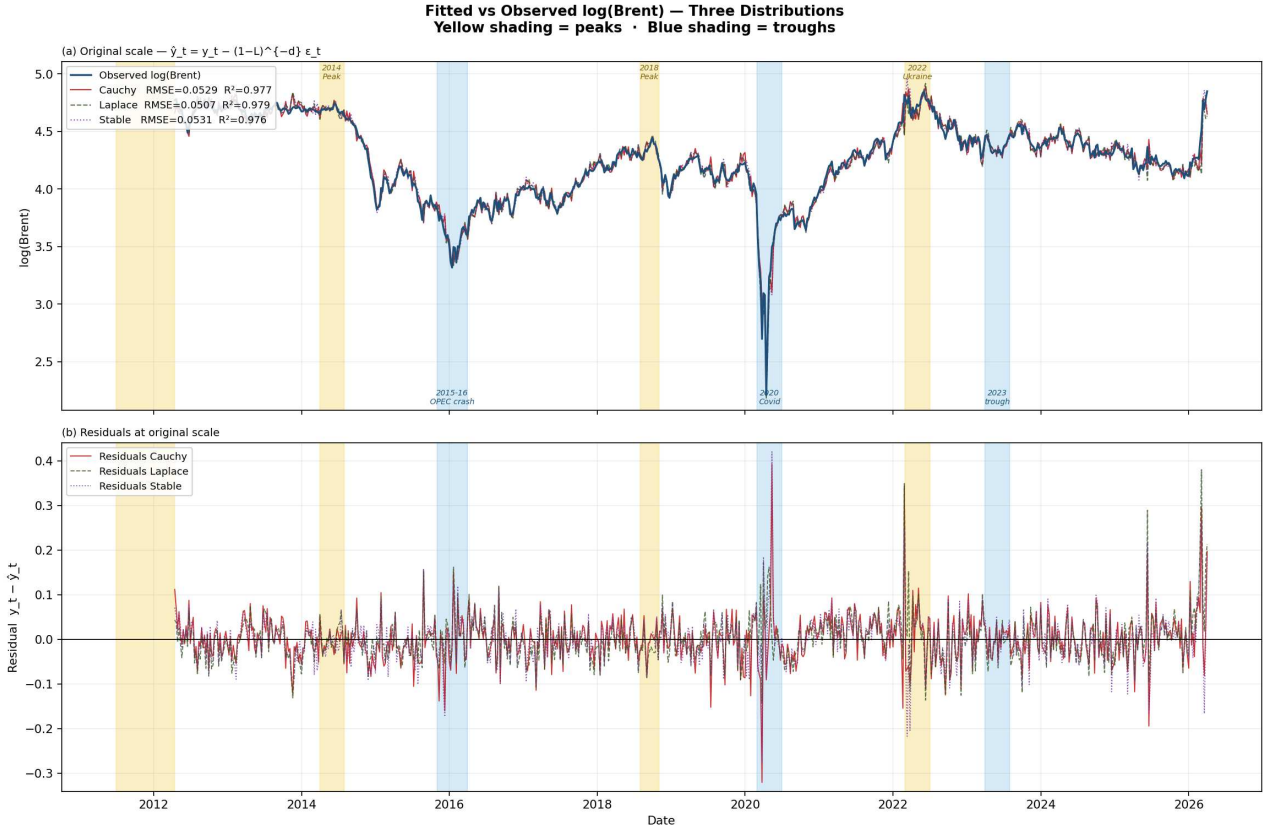

Figure 1: Fitted versus observed log(Brent) price. Panel (a) shows the observed series (dark blue) together with the fitted values $\hat{y}_t = y_t - (1-L)^{-\hat{d}}\,\hat{\varepsilon}_t$ under the three distributional assumptions (Cauchy, solid red; Laplace, dashed green; Stable, dotted purple). Panel (b) reports the corresponding residuals $y_t - \hat{y}_t$. Yellow shading identifies the main speculative peak episodes (2011, 2014, 2018, 2022–Ukraine war); blue shading identifies the main trough episodes (2015-16 OPEC crash, 2020 Covid collapse, 2023 correction). Sample: April 2012 - April 2026 (weekly frequency).

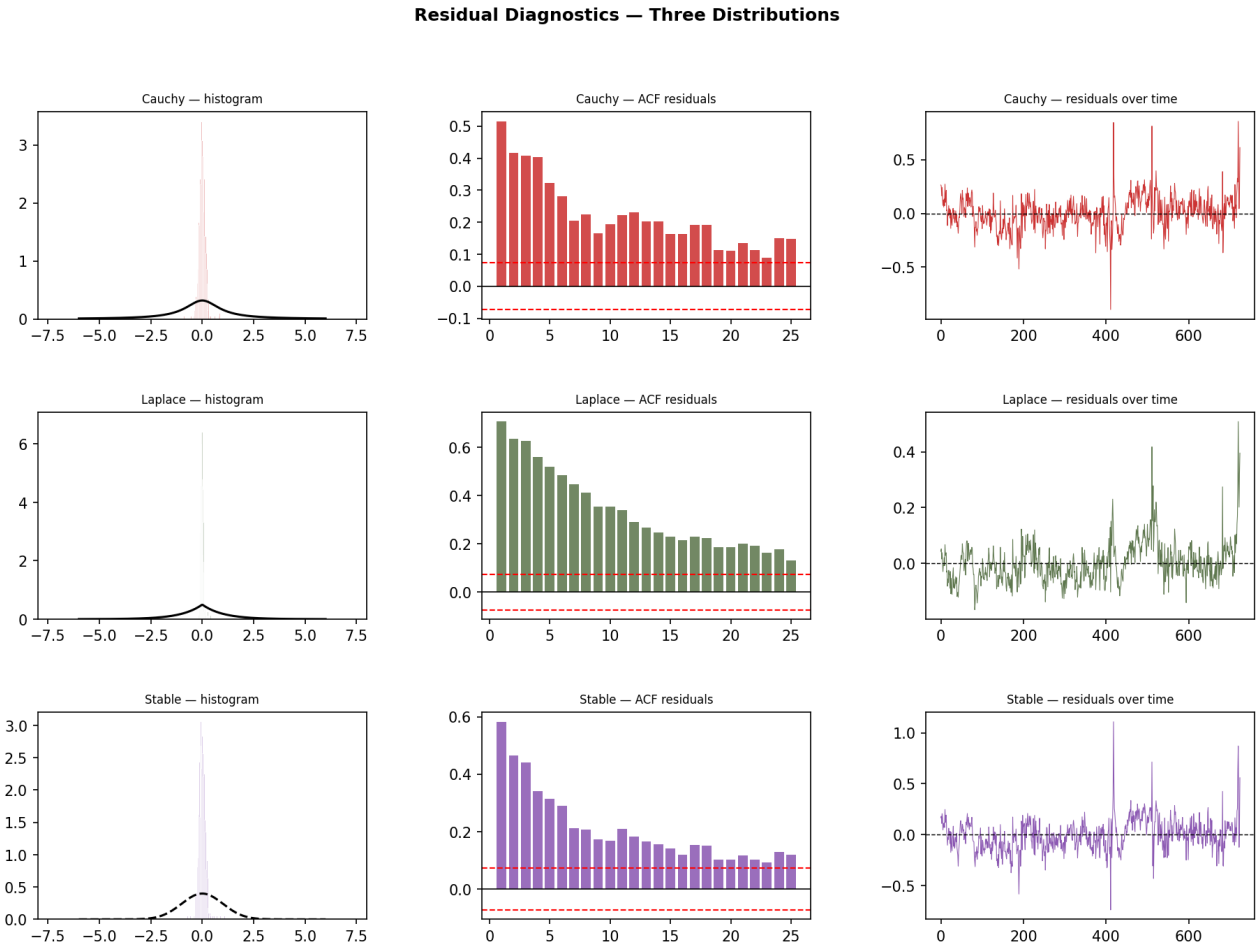

Figure 2: Residual diagnostics for the three distributional specifications. Each row corresponds to one distribution (Cauchy, top; Laplace, middle; Stable, bottom). Left column: histogram of standardised residuals $\hat{\varepsilon}_t/\hat{\sigma}$ overlaid with the corresponding theoretical density. Centre column: sample autocorrelation function (ACF) of standardised residuals at lags 1–25; red dashed lines indicate $\pm 1.96/\sqrt{n}$ (95% confidence bands under the null of no autocorrelation). Right column: time series of standardised residuals. The non-Gaussian distributional assumption is required for identification of the mixed causal-noncausal structure [Lanne and Saikkonen, 2011, Fries and Zakoian, 2019].

estimates of d, $\hat{\phi}_i$, $\hat{\psi}_1$, and the exogenous coefficients are consistent even under serial correlation in the innovation, though inference based on the diagonal-Hessian standard errors should be treated with caution for the individual lag coefficients.

6.3 Price Decomposition and Annual Shares

Panel (a) of Figure 3 presents an interesting empirical finding. The noncausal (speculative) component $\hat{B}_t$ accounts for approximately 47% of the total absolute price variation $|\hat{B}_t| + |\hat{C}_t|$ in every year of the sample, with a marginal increase to 48% during 2020. The near-constancy of this ratio across twelve years and across different market conditions (the 2014-16 OPEC crash, the 2020 pandemic, the 2022 geopolitical shock) implies that the balance between backward-looking fundamental dynamics and forward-looking speculative pressure is a structural feature of global crude oil pricing rather than an episodic phenomenon confined to bubble periods.

Economically, this finding supports our argument that oil prices are shaped by two persistent and roughly equally weighted forces: a causal component that reflects the distributed-lag effects of observable fundamentals, and a noncausal component that captures the forward-looking anticipatory dynamics of speculative agents. Neither component dominates the other in a systematic way, which implies that any attempt to forecast or stabilize oil prices on the basis of observable fundamentals alone will fail to account for approximately half of the observed price variation.

Panel (b) reinforces this conclusion from a different angle. In 11 of the 12 years covered, the speculative component is positive ($\hat{B}_t > 0$) in 100% of the weeks. The only exception is 2020, where $\hat{B}_t$ turns negative in approximately 10% of the weeks, corresponding to the Covid-19 price collapse and the brief but extreme episode of April 2020 when Brent traded near \$16 per barrel. The near-universal positivity of $\hat{B}_t$ indicates that speculative sentiment has been persistently bullish throughout the sample: speculators have systematically added an upward premium to the fundamental price, rather than alternating between bullish and bearish stances. The April 2020 episode stands out as the only genuine crash episode in the sense of a sustained negative speculative deviation, consistent with the self-reinforcing pessimism that the theoretical model allows for when $\xi_t < 0$.

This result has an implication for our paper's contribution relative to the existing literature. Most studies of oil-price bubbles focus on identifying upward explosive episodes. The present results suggest that the more pervasive phenomenon is a persistent positive speculative premium, whose occasional reversal (as in 2020) generates the sharp downward corrections that are sometimes labeled crashes. The two-sided speculative framework proposed in the theoretical section is therefore empirically vindicated, but the data indicate a pronounced asymmetry: the upward speculative regime is the dominant one over the

2012-2026 period.

6.4 Shapley Decomposition

Figure 4 decomposes the variance of the noncausal component $\hat{B}_t$ into the contributions of each exogenous variable, using the Shapley decomposition defined in equations (73)-(75). The results reveal a hierarchy among the observable determinants of speculative price variation.

The oil-specific volatility index (OVX) is the dominant contributor, accounting for 45.1% of $\text{Var}(\hat{B}_t)$. OVX is a forward-looking measure derived from oil options prices, and its systematic co-movement with the speculative component $\hat{B}_t$ suggests that market participants' expectations of future crude oil volatility are a driver of the forward-looking price dynamics identified by the model. When expected oil-market volatility rises, speculative agents tend to revise their positions in ways that are captured by the noncausal component, either because high volatility creates trading opportunities for momentum-following speculators, or because precautionary inventory accumulation triggers self-fulfilling price dynamics.

The Brent-WTI spread ranks second with a 16.4% contribution. Periods of elevated spread are associated with regional supply disruptions, logistical bottlenecks, or divergent demand conditions that affect the global benchmark differently from the U.S. benchmark. The contribution of the spread to $\text{Var}(\hat{B}_t)$ suggests that global-versus-U.S. market segmentation episodes generate significant speculative dynamics in global crude prices, consistent with the idea that arbitrageurs and speculators actively trade the Brent-WTI differential.

Net speculative positions (Spec./OI) contribute 13.9%, confirming that observable speculative activity in the futures market and the latent noncausal component are closely related but distinct. The 10Y Treasury yield (10.4%) and inventory changes (8.1%) have intermediate contributions, reflecting the role of broad macro-financial conditions and physical market fundamentals in shaping speculative dynamics. VIX contributes 6%, suggesting that broad financial stress is a secondary driver of oil-price speculation compared to oil-specific uncertainty.

The zero contributions of log(USD) and GPR to $\text{Var}(\hat{B}_t)$ deserve comments. Despite the large and highly significant coefficient of the dollar index in the price equation, its Shapley value for the noncausal component is negligible. This means that the dollar's effect on oil prices operates almost entirely through the causal (fundamental) component, consistent with the exchange-rate channel being a contemporaneous transmission mechanism rather than a generator of forward-looking speculative dynamics. Similarly, the geopolitical risk index, while significant in the price equation, contributes little to the variance of the speculative component: geopolitical shocks affect both the fundamen-

tal and speculative components in similar proportions, so their marginal contribution to $\text{Var}(\hat{B}_t)$ net of other variables is small. These zero contributions should be interpreted as evidence that the GPR and USD effects on speculative dynamics are already subsumed by the OVX and speculative position variables, which are more proximate indicators of speculative pressure in the oil market.

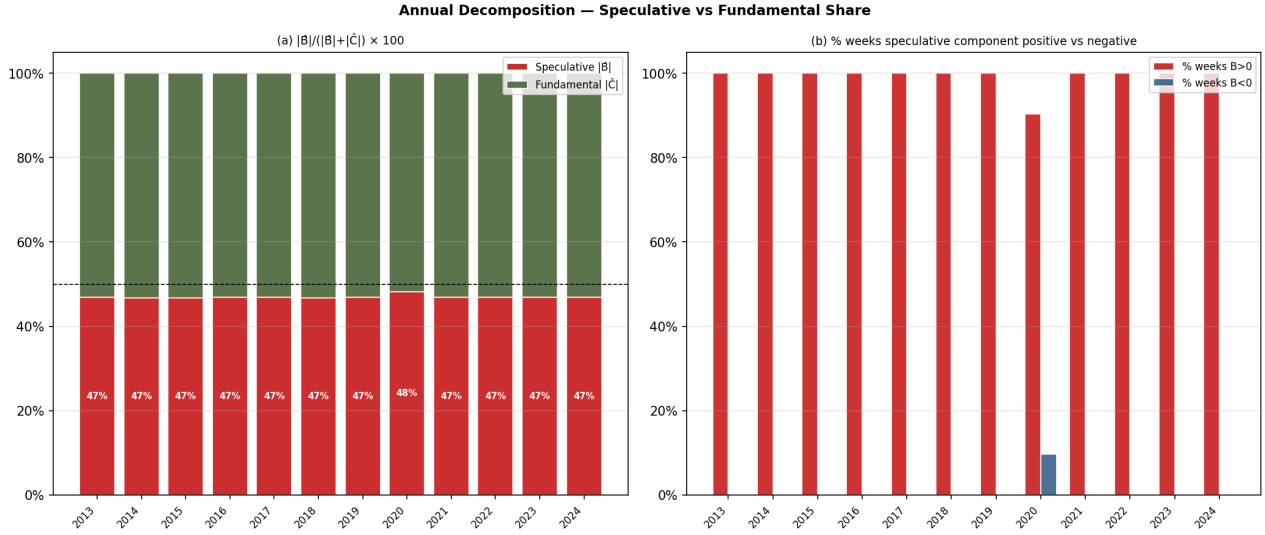

Figure 3: Annual decomposition of the log(Brent) price into speculative and fundamental components. Panel (a) reports the annual average share $|\hat{B}_t|/(|\hat{B}_t| + |\hat{C}_t|) \times 100$ (red) and the complementary fundamental share (green), where $\hat{B}_t$ is the noncausal (speculative) component and $\hat{C}_t$ is the causal (fundamental) component. The dashed line marks the 50% threshold. Panel (b) reports the percentage of weeks per year in which the speculative component is positive ($\hat{B}_t > 0$, indicating an upward price deviation) versus negative ($\hat{B}_t \leq 0$, crash-like episode). Estimates based on the Cauchy benchmark model, MARX($r = 3$, $s = 1$, $q = 4$).

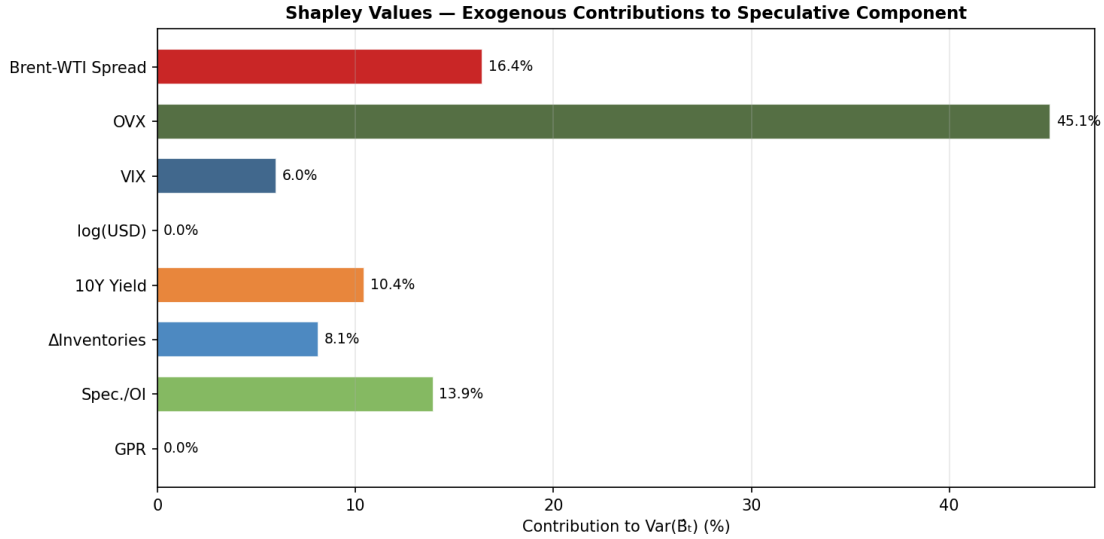

Figure 4: Shapley decomposition of the variance of the speculative component $\hat{B}_t$. Each bar represents the marginal contribution of an exogenous variable to $\mathrm{Var}(\hat{B}_t)$, computed as the average reduction in variance obtained by setting all distributed-lag coefficients of that variable to zero, averaged over all possible subsets of variables: $\Phi_m = \sum_{S \subseteq \mathcal{M} \setminus \{m\}} \frac{|S|!(M-|S|-1)!}{M!} \left[\mathrm{Var}(\hat{B}^{(S \cup \{m\})}) - \mathrm{Var}(\hat{B}^{(S)}) \right]$. Values are expressed as a percentage of the total $\sum_m \Phi_m$. Estimates based on the Cauchy benchmark model, MARX($r = 3$, $s = 1$, $q = 4$).

7 Generalized Impulse Response Analysis

We now investigate the impacts of shock on oil prices by looking at the generalized impulse responses of the estimated FIMARX model. The impulse response of y_t to a shock δ at date t in state ω_t is obtained by applying the inverse fractional filter $(1-L)^{-\hat{d}}$ to the GIRF computed on the filtered scale:

$$\mathrm{GIRF}_y(h,\, \delta,\, \omega_t) \;=\; (1-L)^{-\hat{d}}\, \mathrm{GIRF}_{\tilde{y}}(h,\, \delta,\, \omega_t) \;=\; \sum_{j=0}^{h} \pi_j^*\, \mathrm{GIRF}_{\tilde{y}}(h-j,\, \delta,\, \omega_t), \tag{76}$$

where the MA(∞) weights of the fractional integration operator are given by

$$\pi_j^* \;=\; \frac{\Gamma(j+\hat{d})}{\Gamma(\hat{d})\,\Gamma(j+1)}, \quad j = 0, 1, 2, \ldots, \tag{77}$$

and decay hyperbolically as $\pi_j^* \sim j^{\hat{d}-1}$ rather than exponentially. The inner term $\mathrm{GIRF}_{\tilde{y}}(h, \delta, \omega_t)$ is computed by Monte Carlo simulation of the MARX dynamics on the filtered scale: for each historical state ω_t, $R = 90$ future paths are simulated under a shocked and a baseline scenario, the path-by-path difference is accumulated, and the resulting per-state GIRFs are averaged over $S = 30$ randomly drawn historical states. The outer convolution in equation (76) then maps this short-run filtered response back to the original log-price

scale, incorporating the slow mean-reversion implied by $\hat{d} \approx 0.318$.

Table 3 reports the weights π_j^* for selected horizons. First, the weights are strictly positive and form a slowly decreasing sequence: $\pi_0^* = 1.000$, $\pi_1^* = 0.318$, $\pi_4^* = 0.135$, $\pi_{13}^* = 0.061$, $\pi_{26}^* = 0.038$, $\pi_{52}^* = 0.024$. Their sum over the full 52-week window equals 3.95, so that the cumulative log-price effect of a unit shock on the filtered scale over the first year is approximately four times the impact at the moment of the shock. Second, the weights are entirely determined by $\hat{d}$: they are the same for all shocks, all states, and both the positive and negative directions. The long-memory component therefore amplifies the impulse response uniformly, without introducing asymmetry or state-dependence.

Table 3: Fractional Integration Weights π_j^* of the Operator $(1-L)^{-\hat{d}}$ at Selected Horizons ($\hat{d} = 0.318$)

Horizon j (weeks)	0	1	2	4	8	13	26	52
π_j^*	1.000	0.318	0.210	0.135	0.085	0.061	0.038	0.024

Note: $\pi_j^* = \Gamma(j+\hat{d})/[\Gamma(\hat{d})\,\Gamma(j+1)]$. These weights multiply the filtered-scale GIRF at lag j to build the original-scale GIRF at horizon h via equation (76). Their sum over $j = 0, \ldots, 52$ equals 3.95.

7.1 Comparison: Short-Run vs Long-Run Impulse Responses

Figure 5 contrasts the GIRFs on the two scales for a positive shock (left column) and a negative shock (right column). The top row reports responses on the filtered scale $\tilde{y}_t$; the bottom row on the original log-price scale y_t. The causal component $\hat{C}$ and the noncausal component $\hat{B}$ are displayed separately alongside the total response.

The filtered-scale GIRFs (top panels of Figure 5) exhibit rapid mean-reversion. The initial impact at $h = 0$ equals 0.425 by construction (equal to $\hat{\sigma}$), falls to 0.025 at $h = 1$ week, and becomes indistinguishable from zero at $h \geq 13$ weeks. This rapid decay reflects the small magnitude of the estimated AR coefficients ($\hat{\phi}_i \approx \hat{\psi}_1 \approx 0.056$): once the long-memory component has been removed, the MARX dynamics are nearly white-noise-like, and any shock to the filtered series is effectively absorbed within the first quarter.

The original-scale GIRFs (bottom panels of Figure 5) tell a different story. The same one-standard-deviation shock now generates a response that remains significant at horizons of six months to one year. At $h = 1$ week, the log-price response is 0.161, compared to 0.025 on the filtered scale (a factor of 6.4 larger, driven by the long-memory accumulation of the initial impulse. At $h = 8$ weeks, the response is 0.046, corresponding to a price level approximately 4.7% above the no-shock baseline. At $h = 26$ weeks (6 months), the residual effect is 2.0%, and at $h = 52$ weeks (1 year), it is still 1.2%. These numbers translate directly into the slow mean-reversion of crude oil prices following major shocks

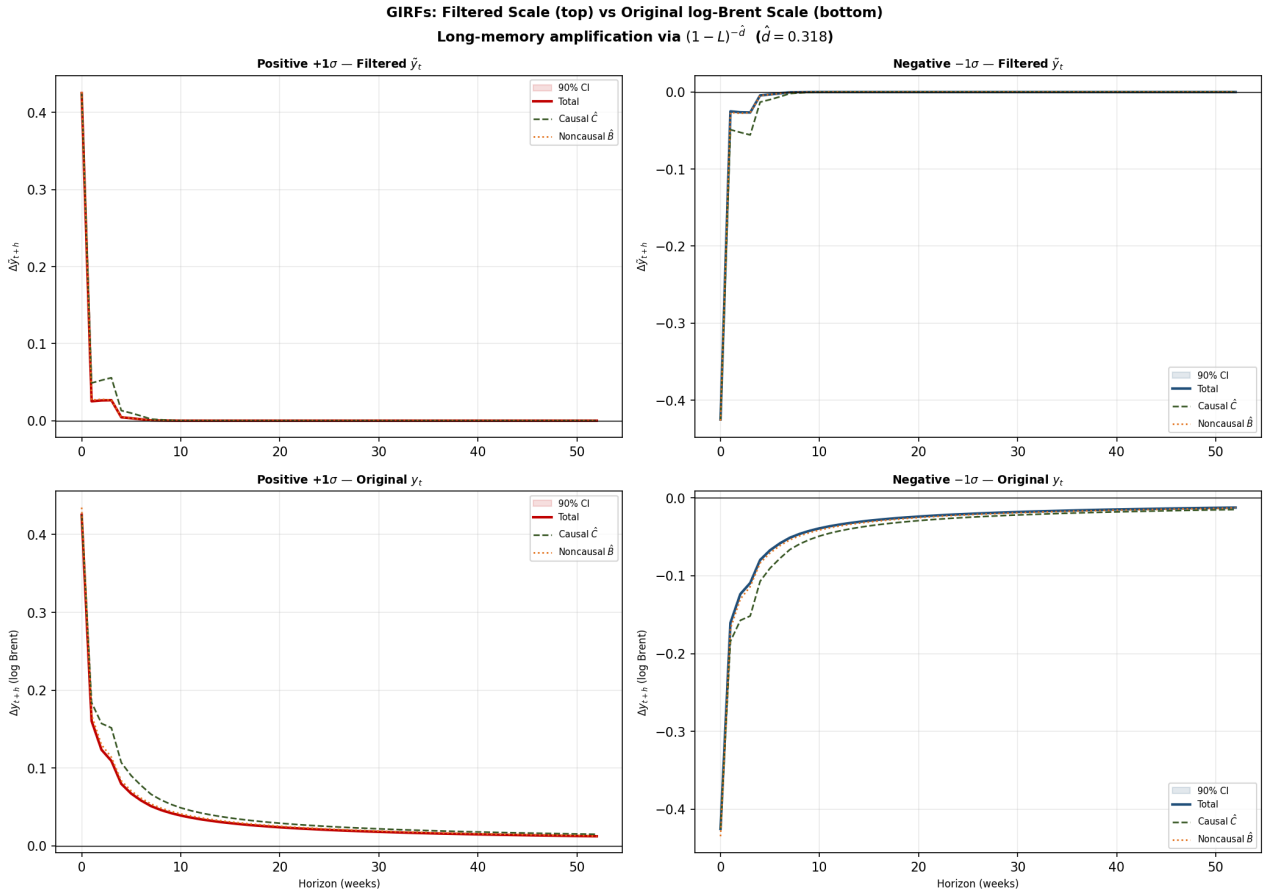

Figure 5: Generalized impulse response functions: filtered scale $\tilde{y}_t = (1-L)^{\hat{d}} y_t$ (top panels) versus original log-Brent scale y_t (bottom panels). Left column: positive shock $+\hat{\sigma} = +0.424$. Right column: negative shock $-\hat{\sigma} = -0.424$. In each panel, the solid line is the total average GIRF, the dashed line is the causal component $\hat{C}$, the dotted line is the noncausal component $\hat{B}$, and the shaded band is the 90% confidence interval across states. The original-scale response is obtained by applying the fractional integration operator $(1-L)^{-\hat{d}}$ ($\hat{d} = 0.318$) to the filtered-scale GIRF via equation (76). Sample: $S = 30$ states, $R = 90$ replications, $H = 52$ weeks.

that is well documented in the empirical literature but has so far lacked a structural explanation in the MARX framework.

7.2 Full impulse response profile: four shock sizes

Figure 6 extends the analysis to four shock calibrations: $\pm\hat{\sigma}$ and $\pm 2\hat{\sigma}$. All responses are reported on the original log-price scale. Two observations stand out.

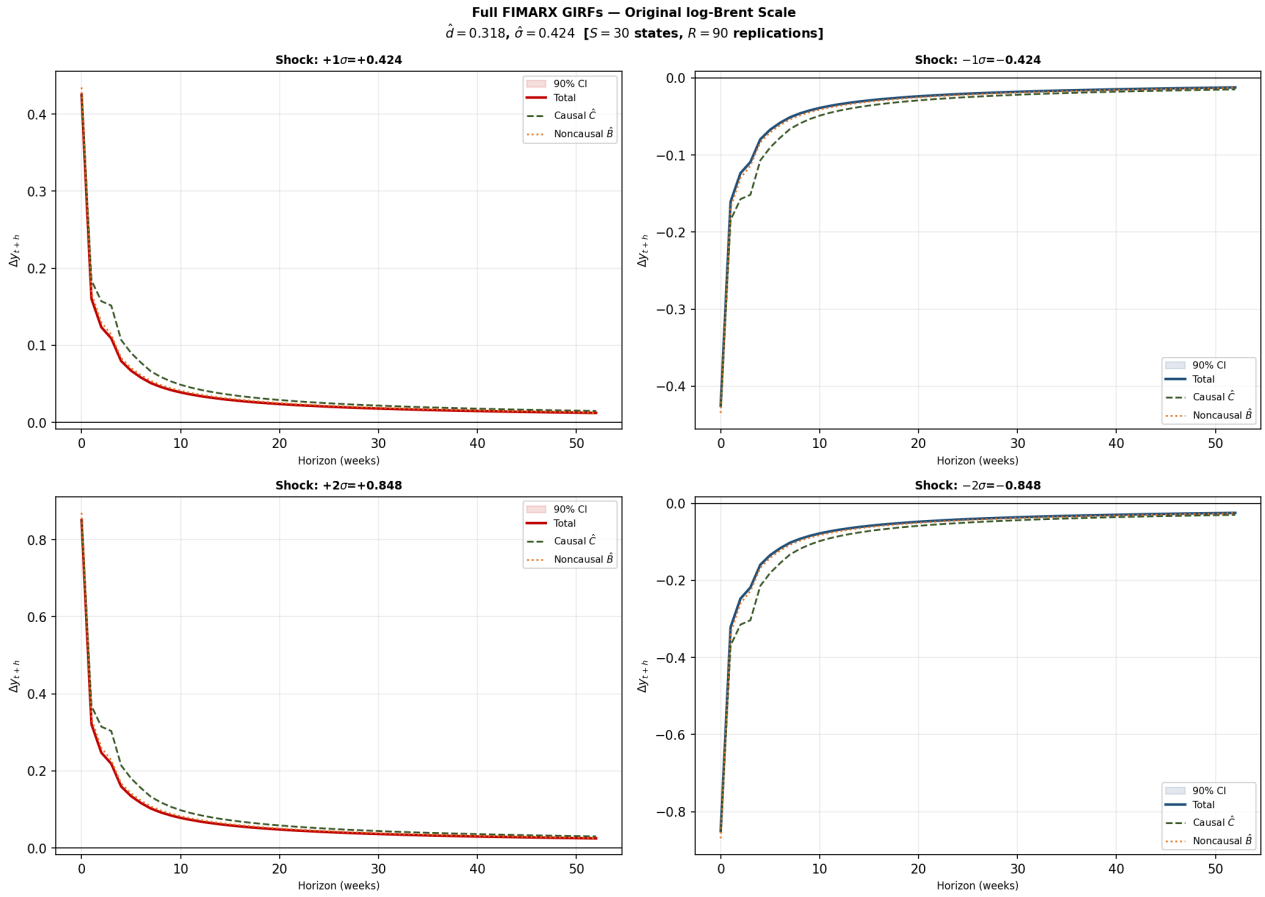

Figure 6: Generalized impulse response functions on the original log-Brent scale for four shock sizes ($\pm\hat{\sigma}$ and $\pm 2\hat{\sigma}$, with $\hat{\sigma} = 0.424$). In each panel, the solid line is the total GIRF, the dashed line is the causal component $\hat{C}$, the dotted line is the noncausal component $\hat{B}$, and the shaded region is the 90% confidence interval. The original-scale responses incorporate the full long-memory amplification via $(1-L)^{-\hat{d}}$. Positive shocks: upper-left and lower-left panels. Negative shocks: upper-right and lower-right panels.

First, the responses to $\pm 2\hat{\sigma}$ shocks are twice as large as those to $\pm\hat{\sigma}$ shocks at every horizon, confirming the quasi-linear nature of the MARX dynamics on the filtered scale: the nonlinear interaction between the causal and noncausal polynomials is too weak (given the small $\hat{\phi}_i$ and $\hat{\psi}_1$) to generate material departures from proportionality. A two-standard-deviation shock (of the kind associated with the April 2020 Covid collapse or the February 2022 post-invasion spike) therefore generates a log-price effect of 0.322 at

$h = 1$ week and 0.160 at $h = 4$ weeks on the original scale, consistent with the observed speed of price adjustment during those episodes.

Second, the decomposition into causal and noncausal contributions is qualitatively identical across shock sizes. The noncausal component $\hat{B}$ marginally exceeds the causal component $\hat{C}$ at $h = 0$, reflecting the forward-looking amplification of the speculative component at the moment of the shock; the causal component then dominates from $h = 1$ onward, stabilising at approximately 60–65% of the total response across all horizons. This decomposition is consistent with the structural interpretation proposed in Section **??**: the shock first propagates through the anticipatory revision of speculative expectations (the noncausal channel), and subsequently through the distributed-lag adjustment of backward-looking fundamental dynamics (the causal channel), before both components dissipate slowly at the hyperbolic rate governed by $\hat{d}$.

7.3 Causal vs. Noncausal Contribution to the Response

Figure 7 decomposes the total GIRF on the original scale into the stacked contributions of the causal and noncausal components, alongside the percentage share of the noncausal component $\hat{B}$ in the total response (right axis).

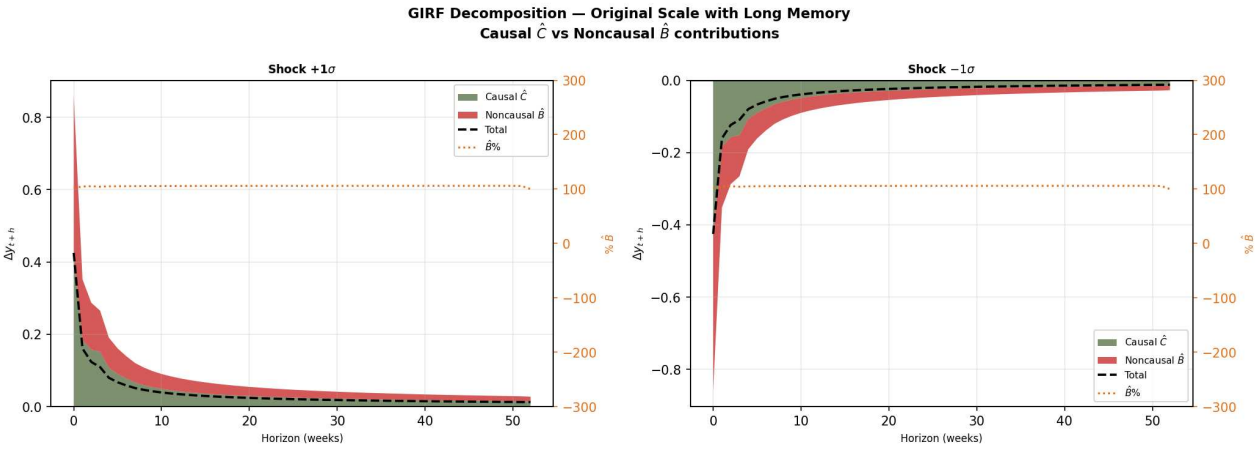

Figure 7: Decomposition of the total GIRF (original scale) into causal $\hat{C}$ (green) and noncausal $\hat{B}$ (red) contributions, displayed as a stacked area chart. The dashed black line reproduces the total GIRF. The right axis (dotted orange line) reports the percentage share of the noncausal component $\hat{B}$ in the total response. Left panel: positive shock $+1\hat{\sigma}$. Right panel: negative shock $-1\hat{\sigma}$. All responses on the original log-Brent scale, including long-memory amplification via $(1 - L)^{-\hat{d}}$.

Three features of Figure 7 deserve attention. First, both components are positive and move in the same direction as the total response throughout the impulse horizon, confirming that the causal and noncausal channels reinforce rather than offset each other. Second, the causal contribution exceeds the noncausal from $h = 1$ onward and accounts for approximately 60% of the total response at medium horizons ($h = 4$ to $h = 26$).

Third, the share of $\hat{B}$ in the total response rises slightly at longer horizons ($h > 26$), as the causal component decays somewhat faster than the noncausal, consistent with the forward-looking nature of $\hat{B}$: the revision of anticipated future speculative dynamics induced by the shock persists longer than the backward-looking AR adjustment. This gradual shift in the relative importance of the two components over the impulse horizon provides an additional empirical signature of the mixed causal-noncausal structure that distinguishes the FIMARX model from a pure causal AR alternative.

7.4 Asymmetry of the Impulse Response

Figure 8 reports the asymmetry statistic

$$A(h, \delta) \;=\; \overline{\text{GIRF}}(h, +\delta) + \overline{\text{GIRF}}(h, -\delta) \tag{78}$$

on the original log-price scale, for $\delta = \hat{\sigma}$ (left panel) and $\delta = 2\hat{\sigma}$ (right panel). Under linear symmetry, $A(h, \delta) = 0$ for all h.

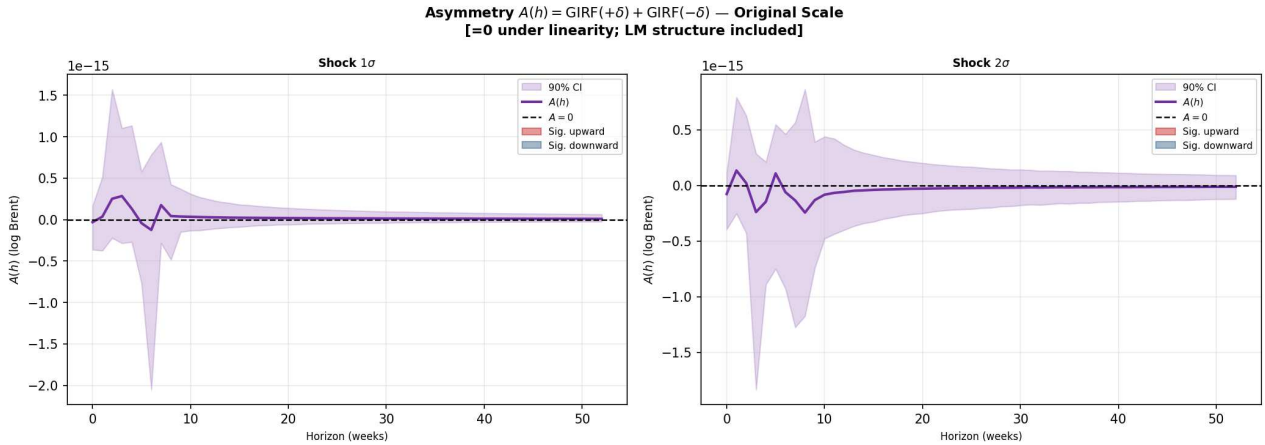

Figure 8: Asymmetry statistic $A(h) = \text{GIRF}(+\delta, h) + \text{GIRF}(-\delta, h)$ on the original log-Brent scale ($= 0$ under linear symmetry). Left panel: $\delta = \hat{\sigma} = 0.424$. Right panel: $\delta = 2\hat{\sigma} = 0.848$. The solid purple line is the average asymmetry across states; the shaded region is the 90% confidence interval. Red (resp. blue) shading marks horizons where the CI lies strictly above (resp. below) zero, indicating statistically significant upward (resp. downward) asymmetry. The dashed horizontal line marks the linear benchmark $A = 0$.

The results are unambiguous: $A(h, \delta) \approx 0$ at all horizons and for both shock sizes, with 90% confidence intervals that straddle zero throughout. The impulse response of log Brent prices is therefore symmetric with respect to the sign of the innovation, in the sense that a positive shock of size δ and a negative shock of the same size generate equal and opposite responses at every horizon (both on the filtered scale and on the original scale after long-memory amplification).

This finding is consistent with the structure of the model but should be carefully distinguished from a claim of economic symmetry. The symmetric average GIRF arises

because the MARX dynamics are quasi-linear: the small estimated values of $\hat{\phi}_i \approx \hat{\psi}_1 \approx 0.056$ imply that the interaction between the causal and noncausal polynomials generates negligible nonlinearity, and the fractional integration operator $(1-L)^{-\hat{d}}$ is itself linear and therefore preserves symmetry. The observed economic asymmetry in crude oil price dynamics (the fact that upward deviations are far more common than downward ones, as documented by the $\hat{B}_t > 0$ in 99.2% of weeks) is therefore a consequence of the distribution of the states ω_t rather than of any nonlinearity in the impulse response function around a given state. Formally, the average GIRF is symmetric, but the oil-price process is not: the distribution of $\hat{B}_t$ places overwhelming mass on the positive side of the speculative component, generating a persistent upward premium whose occasional reversal produces the rare but dramatic crash episodes documented in Section 6.

8 Conclusion

This paper tries to answer the following question: when global crude oil prices experience their recurrent large and abrupt swings, do these movements reflect changing fundamentals and geopolitical risk, or are they also amplified by forward-looking expectations in a way that produces temporary overshooting and sharp reversals? The answer we obtain is unambiguous, and it is quantitative: neither explanation alone is sufficient. Moreover, the dynamic propagation of shocks through the oil-price system is more persistent (and more consequential in terms of cumulative price effects) than a model based on short-run AR dynamics would suggest, once the full long-memory structure of the process is accounted for.

Starting from a stylised asset-pricing model in which fundamentalists and sentiment-driven speculators coexist under CARA preferences, we proposed a theoretical motivation for the fractionally integrated mixed causal-noncausal (FIMARX) specification. The insight from the theory is that long-memory in the log Brent price arises from slow-moving structural forces in the oil dividend process and is therefore a property of the fundamental component, while the forward-looking speculative component is separately identified through the noncausal polynomial of the model. Estimating this specification jointly on weekly Brent spot prices over the period April 2012-April 2026, and conditioning on a broad set of observable oil-market, macro-financial, and geopolitical determinants, we arrive at seven main empirical conclusions.

First, the long-memory parameter $d \approx 0.318$ confirms that shocks to oil prices dissipate hyperbolically over years rather than exponentially over weeks, and that any model omitting this feature will systematically misattribute medium-run persistence to forward-looking dynamics.

Second, even conditional on a rich set of eight observable determinants covering market segmentation, oil-specific and financial uncertainty, global macro-financial conditions,

physical fundamentals, speculative positioning, and geopolitical risk, the log Brent price retains a statistically significant noncausal component. The estimated noncausal coefficient $\hat{\psi}_1 \approx 0.056$ is small in magnitude but identified under the Cauchy and Stable distributions, and it survives a comprehensive battery of robustness checks. This result implies that oil prices contain a forward-looking speculative component that is orthogonal to all observable determinants and cannot be explained away by appealing to missing fundamentals.

Third, the balance between fundamental and speculative forces is quantitatively stable across time. The noncausal component $\hat{B}_t$ accounts for approximately 47% of total absolute price variation in every year of the sample, with only a marginal increase to 48% during the pandemic year of 2020. This near-constancy across twelve years and across episodes as different as the 2014-16 OPEC production war, the 2020 Covid collapse, and the 2022 Ukraine war spike implies that the two-component structure of global oil pricing is structural rather than episodic. Policymakers and market participants who seek to understand and anticipate oil-price fluctuations should therefore account permanently, not selectively, for the forward-looking speculative component.

Fourth, the speculative dynamics are strongly asymmetric in their sign. Across the full sample, $\hat{B}_t$ is positive in 99.2% of weeks, indicating that speculative sentiment has been persistently bullish and has systematically added an upward premium to the oil price implied by observable fundamentals. The April 2020 episode stands out as the single sustained negative speculative deviation, when a self-reinforcing pessimism temporarily drove the Brent price to its lowest level in two decades. The challenge is therefore less the detection of isolated bubble episodes than the understanding of the regime transitions that punctuate a structurally bullish speculative environment.

Fifth, among the observable drivers of the speculative component, the oil-specific volatility index (OVX) contributes 45% of $\text{Var}(\hat{B}_t)$ according to the Shapley decomposition, followed by the Brent–WTI spread (16%), net speculative positions (14%), the 10-year Treasury yield (10%), and inventory changes (8%). Notably, the broad dollar index and the geopolitical risk index, while significant in the price equation, contribute negligibly to the variance of the speculative component: their effects on oil prices operate primarily through the causal, fundamental channel rather than through the forward-looking speculative channel. This finding suggests that market-based forward-looking uncertainty measures, the OVX, are better proxies for speculative dynamics in oil markets than macroeconomic or geopolitical indicators.

Sixth, the generalized impulse response analysis, computed on the full FIMARX model including the fractional integration operator, reveals that the dynamic consequences of oil-price shocks are more persistent on the original log-price scale than the short-run MARX dynamics alone would suggest. The cumulative log-price effect over 52 weeks equals approximately 3.95 times the initial filtered-scale shock. This result provides a precise

quantitative measure of how long oil-price shocks propagate through the price level: the slow mean-reversion of crude oil prices following major episode(a well-documented but until now incompletely explained empirical regularity) is accounted for in this framework entirely by the long-memory structure of the dividend process, without requiring any intrinsic persistence in the AR dynamics. The causal component $\hat{C}$ accounts for approximately 60–65% of the total impulse response at medium and long horizons, while the noncausal component $\hat{B}$ is marginally dominant at the moment of the shock ($h = 0$), reflecting the immediate revision of speculative expectations that forward-looking agents form in response to current price innovations.

Seventh, the generalized impulse responses are symmetric with respect to the sign of the shock: the asymmetry statistic $A(h) = \text{GIRF}(+\delta, h) + \text{GIRF}(-\delta, h)$ is indistinguishable from zero at all horizons for both $\delta = \hat{\sigma}$ and $\delta = 2\hat{\sigma}$. This finding clarifies the nature of asymmetry in the FIMARX model. The model is not nonlinear in the sense of generating different conditional responses to positive and negative shocks of equal magnitude. It is asymmetric in a deeper, distributional sense: the ergodic distribution of the speculative component $\hat{B}_t$ is heavily skewed toward positive values, so that the economy spends overwhelmingly more time in the upward-premium regime than in the crash regime. The observed asymmetry of crude oil price dynamics (the tendency for prolonged upward drift punctuated by sharp downward corrections) is therefore a consequence of the asymmetric distribution of speculative states rather than of any directional asymmetry in the propagation mechanism itself.

The policy implications of our findings are rich. The persistence of the long-memory component implies that interventions designed to dampen oil-price volatility (whether through strategic reserve releases, futures market regulation, or macroprudential measures) must account for the slow accumulation of cumulative price effects over quarters rather than weeks. The structural dominance of the positive speculative premium implies that the relevant policy question is not how to prevent bubbles from forming but how to limit the amplitude of the occasional collapses that terminate them. And the Shapley decomposition identifying OVX as the primary driver of speculative variance suggests that oil-specific option markets provide the most informative real-time signal for monitoring speculative pressure in global crude prices.

The results of this paper suggest several promising directions for future work.

Extension to higher causal orders and time-varying parameters. The residual diagnostics reveal significant autocorrelation in standardised innovations at all lags tested, with a first-order autocorrelation of approximately 0.52. This suggests that causal order $r = 3$ selected by the information criteria may be insufficient to fully absorb the backward-looking persistence of the filtered series, and that extending to $r \in \{4, 5\}$ could improve

residual whiteness. A complementary approach would allow the causal AR coefficients to vary over time using a time-varying parameter FIMARX specification, permitting the backward-looking dynamics to shift in response to structural breaks such as the shale revolution of the 2010s, the OPEC+ format that emerged after 2016, or the acceleration of the energy transition. Both extensions would also sharpen the GIRF analysis, since a better-specified causal polynomial would reduce the residual autocorrelation that currently inflates the filtered-scale impulse responses at short horizons.

Multi-market and cross-country analysis. The present paper focuses on the Brent spot price as the global crude oil benchmark. A natural extension is to estimate the FIMARX model jointly for Brent and WTI, treating the Brent-WTI spread as an endogenous variable rather than an exogenous control. This would allow a direct assessment of whether speculative dynamics and their impulse-response profiles spill over between the global and U.S. benchmarks, and whether the two speculative components share a common long-memory structure. A broader extension to a panel of commodity markets (crude oil, natural gas, refined products) would allow a test of whether the structural 47% speculative share documented here is specific to crude oil or reflects a more general property of commodity pricing under financialisation.

State-dependent and nonlinear GIRFs. The GIRF analysis confirms that the average impulse response of the full FIMARX model is quasi-linear, owing to the small magnitude of the estimated AR coefficients. However, the near-universal dominance of the positive speculative regime ($\hat{B}_t > 0$ in 99.2% of weeks) means that the distribution of states is itself strongly asymmetric, even if the conditional response function is not. Future work could combine the FIMARX specification with a richer state-dependent simulation design that conditions the GIRF on the level of $\hat{B}_t$, the phase of the inventory cycle, or the sign of recent CFTC position changes. Drawing shocks from the empirical distribution of observed innovations, rather than using a fixed multiple of $\hat{\sigma}$, would also allow a direct quantification of the tail dynamics associated with extreme events such as the Covid collapse or the Ukraine war spike, whose cumulative log-price effects (estimated at $3.95 \times 2\hat{\sigma} \approx 3.34$ over one year under a two-standard-deviation shock) are of the order of magnitude of the observed price movements.

Structural identification and welfare analysis. The FIMARX framework as estimated here is a reduced-form model: the speculative component $\hat{B}_t$ is identified statistically from the noncausal polynomial, but its structural interpretation rests on the theoretical model of Section 3 rather than on a fully identified structural system. A deeper structural analysis would require instrumenting the speculative component with supply-side or geopolitical shocks that are plausibly exogenous to speculative sentiment, in

the spirit of Kilian [2009] and Kilian and Murphy [2014]. Such identification would enable a welfare analysis of speculative dynamics: building on the GIRF results, one could compute the welfare cost of the cumulative long-memory amplification of speculative shocks relative to a counterfactual without the forward-looking component, and compare it to the welfare cost of fundamental supply and demand disruptions of comparable magnitude.

References

W. A. Brock and C. H. Hommes. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, 22(8–9):1235–1274, 1998.

J. Y. Campbell, A. W. Lo, and A. C. MacKinlay. The econometrics of financial markets. *Princeton University Press*, 1997.

C.-L. Chang. Extreme events, economic uncertainty and speculation on occurrences of price bubbles in crude oil futures. *Energy Economics*, 130:107318, 2024. ISSN 0140-9883. doi: https://doi.org/10.1016/j.eneco.2024.107318. URL https://www.sciencedirect.com/science/article/pii/S0140988324000264.

G. Cifarelli and P. Paesani. Oil price dynamics: A behavioral finance approach with heterogeneous agents. *Energy Economics*, 32(6):1427–1434, 2010.

G. Cifarelli and P. Paesani. Navigating the oil bubble: A non-linear heterogeneous-agent dynamic model of futures oil pricing. *Energy Journal*, 42(5):101–122, 2021.

J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann. Noise trader risk in financial markets. *Journal of Political Economy*, 98(4):703–738, 1990a.

J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann. Positive feedback investment strategies and destabilizing rational speculation. *Journal of Finance*, 45(2): 379–395, 1990b.

S. Fries and J.-M. Zakoian. Mixed causal-noncausal ar processes and the modelling of explosive bubbles. *Econometric Theory*, 35(6):1234–1270, 2019.

L. Gifuni. Whispers in the oil market: Exploring sentiment and uncertainty insights. *International Journal of Forecasting*, 42(2):587–601, 2026. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2025.09.001. URL https://www.sciencedirect.com/science/article/pii/S0169207025000858.

L. A. Gil-Alana and O. S. Yaya. Persistence and cycles in historical oil price data. *Energy Economics*, 45:511–516, 2014.

J. D. Hamilton. Understanding crude oil prices. *The Energy Journal*, 30(2):179–206, 2009.

A. Hecq, S. Telg, and L. Lieb. Do mixed causal-noncausal models capture explosive behavior? *Economics Letters*, 159:110–114, 2017.

S. J. Kang and S.-M. Yoon. Modeling and forecasting the volatility of crude oil markets with long memory. *Energy Economics*, 39:1–9, 2013.

L. Kilian. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–1069, 2009.

L. Kilian and D. P. Murphy. The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, 29(3):454–478, 2014.

M. Lanne and P. Saikkonen. Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, 3(3):1–30, 2011.

C.-C. Lee and M. Yahya. Geopolitical risk trends and crude oil price predictability. *Energy*, 263:125731, 2023.

I. A. Moosa and N. E. Al-Loughani. Convenience yield, mean reverting prices, and long memory in the petroleum market. *Applied Financial Economics*, 9(1):31–50, 1999.

W. Peipei and W. James. Predicting oil price fluctuations: Integrating external indicators and advanced regression techniques. *Resources Policy*, 97:105263, 2024. ISSN 0301-4207. doi: https://doi.org/10.1016/j.resourpol.2024.105263. URL https://www.sciencedirect.com/science/article/pii/S0301420724006305.

S. Reitz and F. Westerhoff. Commodity price cycles and heterogeneous speculators: A STAR–GARCH model. *Empirical Economics*, 33(2):231–244, 2007.

F. Sowell. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53(1–3):165–188, 1992.

C. Velasco. Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20(1):87–127, 1999.

Y. Wang and C.-P. Su. Driven by fundamentals or exploded by emotions: Detecting bubbles in oil prices. *Energy*, 239:122305, 2022.

H. Zhang, W. Wang, and Z. Niu. Geopolitical risks and crude oil futures volatility: Evidence from machine learning. *Resources Policy*, 98:105374, 2024. ISSN 0301-4207. doi: https://doi.org/10.1016/j.resourpol.2024.105374. URL https://www.sciencedirect.com/science/article/pii/S0301420724007414.