ENSAE PARIS – INSTITUT POLYTECHNIQUE DE PARIS

Research Report — Extension

Stochastic and Machine-Learning Temperature Models for Weather-Derivative Pricing

Application to natural-gas coupling

July 2026

AUTHOR:
Nayel BENABDESADOK

Extension of the report *Stochastic and Machine-Learning Temperature Models for Weather-Derivative Pricing*

SELF-DIRECTED QUANTITATIVE RESEARCH PROJECT

Contents

Abstract

This second part examines whether the temperature models developed in the main report can improve the modeling of Henry Hub natural gas prices. Two angles are explored. The first, structural, casts the Schwartz–Smith two-factor model in state-space form and estimates it by Kalman filter on the futures curve. It does not succeed, for lack of sufficiently long maturities, an identifiability problem that we document precisely. The second targets volatility: a GARCH(1,1) augmented with degree-day anomalies reveals a real but modest weather coupling, concentrated in the tails of the distribution and in late winter.

1 Introduction

This document extends the main report, *Stochastic and Machine-Learning Temperature Models for Weather-Derivative Pricing*, which built temperature models (Ornstein–Uhlenbeck and machine learning) and compared them on the pricing of HDD weather derivatives. The present part explores another application, with commodities. It seems fairly natural that a commodity would be influenced by weather conditions. Consider a winter that is particularly cold and unexpected for a gas company: the quantity of gas planned for household heating is consumed more quickly, and the company must obtain additional gas on the spot market, which drives the price up. Thus, weather conditions would be a factor in the supply and demand of certain commodities, and therefore in their price. The aim of this document is to examine whether the temperature models built in the main report can be used to improve the modeling of natural gas prices and to try to propose a coupling between these two markets.

Here we focus on U.S. Henry Hub natural gas: does this weather information, injected no longer into the payoff of an HDD contract but into the dynamics of the gas underlying, add anything?

We approach the problem from two successive angles. The first targets the *level* of the price: a structural two-factor model, that of Schwartz and Smith, estimated by Kalman filter on the futures curve, would have provided a complete risk-neutral framework in which to then plug in temperature. However, in this document we will detail why this path did not succeed.

The second angle targets *volatility*: daily spot returns then suffice, and a GARCH augmented with weather variables (degree-day anomalies) makes it possible to test directly whether cold moves the variance of gas.

2 Henry Hub Data and Stylised Facts

The underlying for this entire part is natural gas at the *Henry Hub*, the Louisiana delivery point that serves as the benchmark for US contracts. The daily spot price comes from the *Energy Information Administration*[1], which makes the entire historical series available. It spans January 1997 to April 2026, that is 7350 daily values.

The series is quoted in $/MMBtu, the *million British thermal units*, a unit of energy rather than volume. A BTU represents the energy required to raise the temperature of one pound of water by one degree Fahrenheit, and the MMBtu is the standard sales unit for natural gas in the United States: the price reflects the energy content delivered rather than the raw volume.

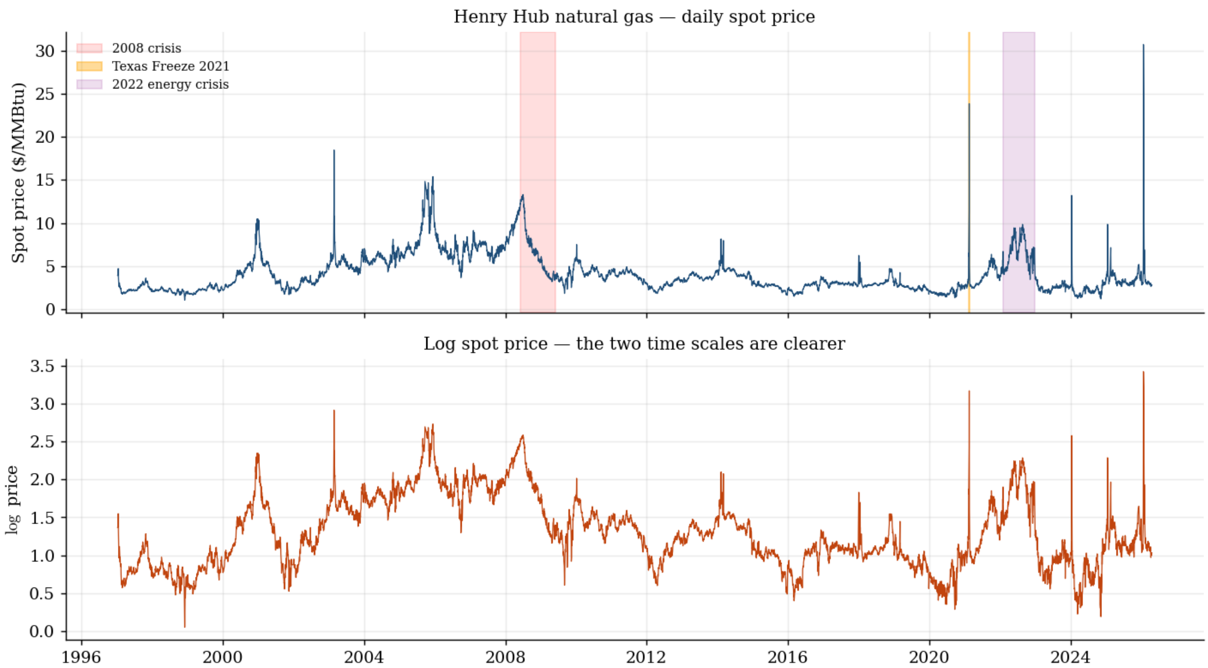

Figure 1: Daily natural gas spot price at the Henry Hub ($/MMBtu) and its logarithm, 1997–2026.

From this series, and in light of the literature on commodities, two superimposed dynamics emerge. On one side, the price oscillates around a long-term equilibrium level, between 3 and 5 $/MMBtu over most of the period (figure 1), a level that drifts slowly with the structural evolution of the market. On the other, this level is punctuated by violent and asymmetric spikes, systematically upward, and systematically traceable to a weather or geopolitical event: a cold winter on low storage in 2003, hurricanes Katrina and Rita in 2005, the Texas freeze of February 2021, the LNG rush to Europe in 2022. These shocks are short and intense, with enormous price swings that subside within a few weeks, or even a few months. Taking the logarithm (lower panel of figure 1) makes these two time scales even more perceptible, by reducing multiplicative shocks to additive deviations.

A Hodrick–Prescott decomposition of the log price, purely illustrative, visually separates these two components (figure 2): the smooth trend plays the role of the long-term factor, the cycle that of the transitory deviations. This is exactly the decomposition that the Schwartz–Smith model formalizes, and with which we begin.

[1] https://www.eia.gov/dnav/ng/hist/rngwhhdD.htm

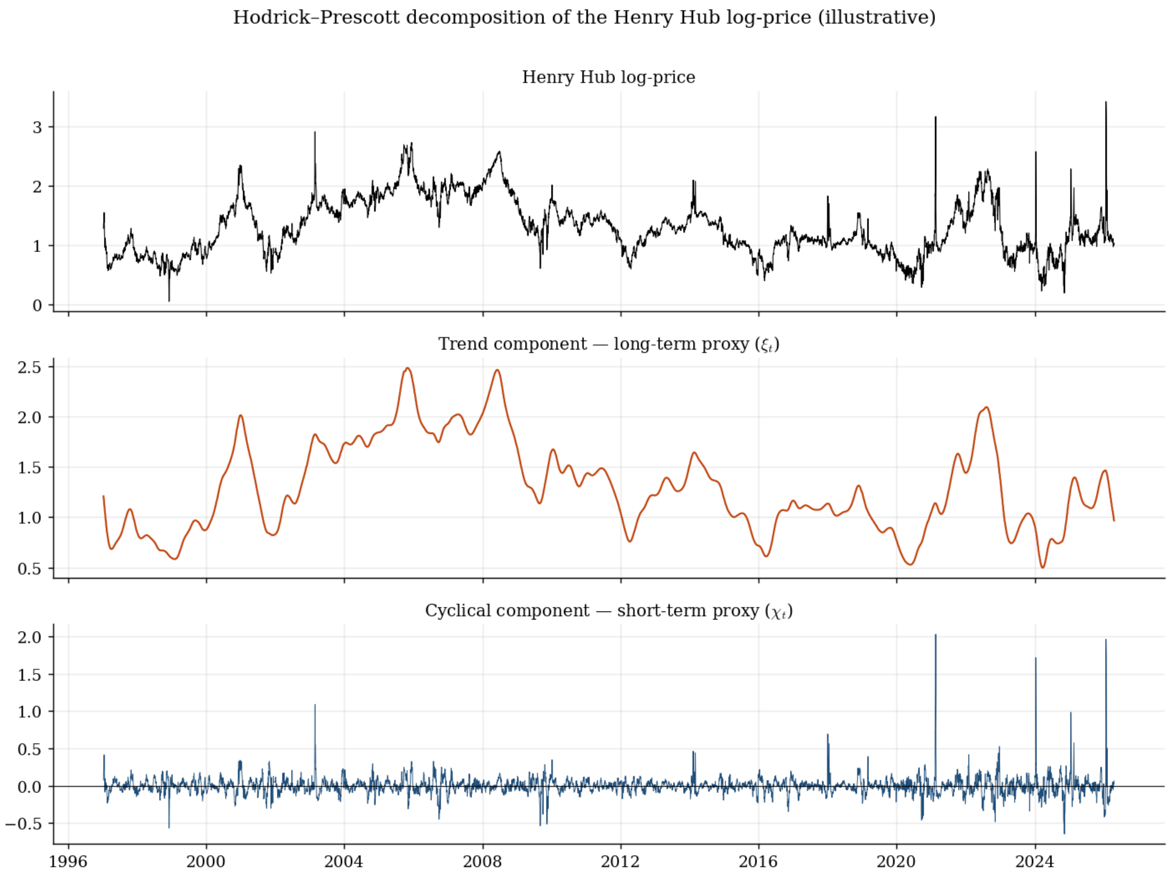

Figure 2: Hodrick–Prescott decomposition of the Henry Hub log price (illustrative): slow trend and transitory cycle.

3 Modelling Natural Gas: the Schwartz–Smith Model

Thanks to the extensive literature on commodities, we can implement the model of Schwartz and Smith (2000), which decomposes the log-price into two latent factors, one short-term and the other long-term. The idea of this detour is to extract these two dynamics from the gas price, in particular the short-term factor: if one could isolate the transitory shocks from the equilibrium level, one could directly test their correlation with large temperature deviations, that is, the winters with high degree-days. This approach did not, however, succeed, owing to a data problem that we make explicit throughout the section.

3.1 Two Factors for One Price: Specification

As observed earlier, the gas spot price blends two dynamics of a different nature. In the short term, transitory shocks (a cold spell, a supply tension, a geopolitical event) push the price away from its equilibrium level, then dissipate within a few days or weeks. In the long term, the equilibrium level itself drifts slowly, following the structural evolutions of the market (production capacity, long-term demand). Schwartz and Smith (2000) formalise this intuition by decomposing the logarithm of the spot price S_t into two latent factors:

$$\log S_t = \chi_t + \xi_t, \tag{1}$$

where χ_t carries the short-term dynamics and ξ_t the long-term one. Each follows a mean-reverting Ornstein–Uhlenbeck process, the two being correlated:

$$\begin{cases} \mathrm{d}\chi_t = -\zeta\,\chi_t\,\mathrm{d}t + \sigma_\chi\,\mathrm{d}Z_t^\chi, \\ \mathrm{d}\xi_t = \left(\mu_\xi - \gamma\,\xi_t\right)\mathrm{d}t + \sigma_\xi\,\mathrm{d}Z_t^\xi, \end{cases} \qquad \mathrm{d}Z_t^\chi\,\mathrm{d}Z_t^\xi = \rho\,\mathrm{d}t. \tag{2}$$

The factor χ_t has zero mean and ξ_t reverts toward its equilibrium level μ_ξ/γ. The two reversion speeds satisfy $\zeta \gg \gamma$: the short-term factor dissipates within a few weeks (short half-life $\ln 2/\zeta$), whereas the long-term factor takes several years to return to equilibrium (γ close to zero). The canonical model of Schwartz and Smith (2000), in which ξ_t is an arithmetic Brownian motion with drift μ_ξ, is obtained as the limiting case $\gamma \to 0$.

Since the purpose of the model is to price futures contracts, we place ourselves directly under the risk-neutral measure $\mathbb{Q}$, without first developing the physical dynamics. By Girsanov's theorem, the change to $\mathbb{Q}$ corrects the drifts of the two factors through *market prices of risk* λ_χ and λ_ξ:

$$\begin{cases} \mathrm{d}\chi_t = \left(-\lambda_\chi - \zeta\,\chi_t\right)\mathrm{d}t + \sigma_\chi\,\mathrm{d}Z_t^{\chi,\mathbb{Q}}, \\ \mathrm{d}\xi_t = \left(\mu_\xi - \lambda_\xi - \gamma\,\xi_t\right)\mathrm{d}t + \sigma_\xi\,\mathrm{d}Z_t^{\xi,\mathbb{Q}}. \end{cases} \tag{3}$$

In everything that follows, all expectations and variances are taken under $\mathbb{Q}$.

Unlike the OU process of the temperature pillar, the volatilities σ_χ and σ_ξ are here assumed constant over time. Solving the two equations by variation of the constant (detailed computation in appendix A), we obtain, for all $s < t$:

$$\begin{cases} \chi_t = \chi_s\,e^{-\zeta(t-s)} - \dfrac{\lambda_\chi}{\zeta}\left(1 - e^{-\zeta(t-s)}\right) + \sigma_\chi \displaystyle\int_s^t e^{-\zeta(t-u)}\,\mathrm{d}Z_u^\chi, \\ \xi_t = \xi_s\,e^{-\gamma(t-s)} + \dfrac{\mu_\xi - \lambda_\xi}{\gamma}\left(1 - e^{-\gamma(t-s)}\right) + \sigma_\xi \displaystyle\int_s^t e^{-\gamma(t-u)}\,\mathrm{d}Z_u^\xi. \end{cases} \tag{4}$$

The state vector $X_t = (\chi_t, \xi_t)^\top$ is Gaussian. Since the stochastic integrals of equation (4) are centred, its mean retains only the deterministic parts, and its variance is computed by Itô isometry. The full derivation, including the cross-covariance term, is deferred to appendix A. With $s = 0$, we obtain

$$\mathbb{E}[X_t] = \begin{pmatrix} e^{-\zeta t}\chi_0 - \dfrac{\lambda_\chi}{\zeta}\left(1-e^{-\zeta t}\right) \\ e^{-\gamma t}\xi_0 + \dfrac{\mu_\xi - \lambda_\xi}{\gamma}\left(1-e^{-\gamma t}\right) \end{pmatrix}, \quad \mathrm{Var}(X_t) = \begin{pmatrix} \dfrac{\sigma_\chi^2}{2\zeta}\left(1-e^{-2\zeta t}\right) & \dfrac{\rho\,\sigma_\chi\sigma_\xi}{\zeta+\gamma}\left(1-e^{-(\zeta+\gamma)t}\right) \\ \cdot & \dfrac{\sigma_\xi^2}{2\gamma}\left(1-e^{-2\gamma t}\right) \end{pmatrix}. \tag{5}$$

The only delicate point is the covariance $\mathrm{Cov}(\chi_t, \xi_t)$, which comes from the *generalised* Itô isometry applied to the two correlated stochastic integrals (appendix A).

The logarithm of the spot, the sum of the two factors (equation 1), is itself Gaussian, with expectation $\mathbb{E}[\log S_t] = \mathbb{E}[\chi_t] + \mathbb{E}[\xi_t]$ and variance

$$\mathrm{Var}(\log S_t) = \mathrm{Var}(\chi_t) + \mathrm{Var}(\xi_t) + 2\,\mathrm{Cov}(\chi_t, \xi_t) = \frac{1-e^{-2\zeta t}}{2\zeta}\,\sigma_\chi^2 + \frac{1-e^{-2\gamma t}}{2\gamma}\,\sigma_\xi^2 + 2\,\frac{1-e^{-(\zeta+\gamma)t}}{\zeta+\gamma}\,\sigma_\chi\sigma_\xi\,\rho, \tag{6}$$

so that S_t is log-normal, with $\mathbb{E}[S_t] = \exp\!\left(\mathbb{E}[\log S_t] + \frac{1}{2}\mathrm{Var}(\log S_t)\right)$.

3.2 From Spot to Futures

As announced, we seek to separate the two factors χ_t and ξ_t from the observed prices. The spot alone does not suffice: at each date, it provides only their sum $\log S_t = \chi_t + \xi_t$, and a scalar decomposes into two terms in infinitely many ways. We therefore need a source of information that loads the two factors *differently*: this is the role of the futures curve, each maturity of which mixes χ_t and ξ_t in distinct proportions.

Since the price of a future with maturity T is, by definition, $F(t,T) = \mathbb{E}_t^{\mathbb{Q}}[S_T]$, conditional log-normality gives:

$$\boxed{\log F(t,T) = e^{-\zeta(T-t)}\chi_t + e^{-\gamma(T-t)}\xi_t + A(T-t)} \tag{7}$$

where the deterministic function $A(\tau)$, with $\tau = T - t$, groups together risk premia and variance corrections:

$$A(\tau) = -\frac{\lambda_\chi}{\zeta}\left(1-e^{-\zeta\tau}\right) + \frac{\mu_\xi - \lambda_\xi}{\gamma}\left(1-e^{-\gamma\tau}\right) + \frac{1}{2}\left(\frac{1-e^{-2\zeta\tau}}{2\zeta}\,\sigma_\chi^2 + \frac{1-e^{-2\gamma\tau}}{2\gamma}\,\sigma_\xi^2 + 2\,\frac{1-e^{-(\zeta+\gamma)\tau}}{\zeta+\gamma}\,\sigma_\chi\sigma_\xi\,\rho\right). \tag{8}$$

It is therefore necessary to obtain futures contract data to carry out our research.

3.3 Analysis of Henry Hub Futures Contracts

To carry out this separation, we obtain as many Henry Hub gas futures contracts as possible. For this we use the Databento API[2], a historical financial data platform offering free access upon registration. We thus have twelve daily contracts from June 2010 to April 2024, denoted $C_1, \ldots, C_{12}$ (table 1): the contract number indicates its relative maturity, C_1 being the contract closest to expiration (the *front-month*) and C_{12} the one delivering in twelve months.

[2] https://databento.com

Table 1: Extract of the prices of the twelve Henry Hub futures contracts (Databento, \$/MMBtu).

Date	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}
2010-06-07	4.980	5.031	5.069	5.140	5.365	5.650	5.822	5.769	5.644	5.392	5.382	5.422
2010-06-08	4.799	4.856	4.894	4.994	5.246	5.521	5.699	5.646	5.522	5.289	5.304	5.359
2010-06-09	4.672	4.728	4.787	4.870	5.171	5.480	5.657	5.612	5.494	5.263	5.281	5.333
2010-06-10	4.686	4.751	4.803	4.883	5.172	5.477	5.674	5.628	5.500	5.270	5.287	5.329
2010-06-11	4.793	4.846	4.899	4.969	5.227	5.515	5.686	5.626	5.510	5.265	5.278	5.325

These contracts are *rolling*: at each expiration, C_1 is replaced by the former C_2, C_2 by C_3, and so on, with a new C_{12} being issued for the corresponding month. We therefore observe at each date exactly twelve active contracts, whose residual maturities $\tau_i(t)$ vary continuously and jump at each roll. The prices are given to us, but not their expiries: the expiration dates $T_i(t)$ must be reconstructed, which is done in three steps via the NYMEX convention. First, the expiration rule: a natural gas contract expires on the third business day before the first of the delivery month. Next, the delivery month of the front-month C_1 at date t is the current month as long as its contract has not yet expired, the following month otherwise, this test advancing C_1 by one step at each roll. Finally, since C_i delivers $(i-1)$ months after C_1, we apply the expiration rule to its delivery month to obtain $T_i(t)$, hence the residual maturity in years

$$\tau_i(t) = \max\left(\frac{T_i(t)-t}{365.25}, \frac{1}{365.25}\right), \tag{9}$$

the one-day floor avoiding a zero or negative maturity right at expiration. table 2 gives the order of magnitude: the maturity progresses steadily from the front-month to one year, and *caps at* $\tau_{\max} \approx 1$ *year.*

Table 2: Residual maturity τ_i (in years) of three representative contracts.

Contract	**mean** τ_i	**max** τ_i
C_1 (front-month)	0.033	0.25
C_6	0.49	0.74
C_{12}	0.95	0.99

3.3.1 Contango, Backwardation and the Samuelson Effect

The shape of the futures curve reflects the balance between the cost of carrying a commodity and the benefit of holding it right away. We speak of *contango* when the forward curve is upward-sloping and lies above the spot, $F_i(t) > S_t$: the normal regime of abundant supply, where storage and carry dominate. We speak of *backwardation* in the opposite case, a downward-sloping curve below the spot, $F_i(t) < S_t$: the market pays a premium, the *convenience yield*, to have the good immediately available, a sign of supply tension. For natural gas, a harsh winter thus pushes the front of the curve into backwardation, and the seasonal bumps of the curve reflect heating demand. figure 3 shows two examples anchored to their spot price. Empirically, the front-month C_1 frequently switches between the two regimes (54 % versus 46 %) whereas C_8 remains in contango 77 % of the time, the transitory shocks carried by χ_t stirring above all the short end of the curve.

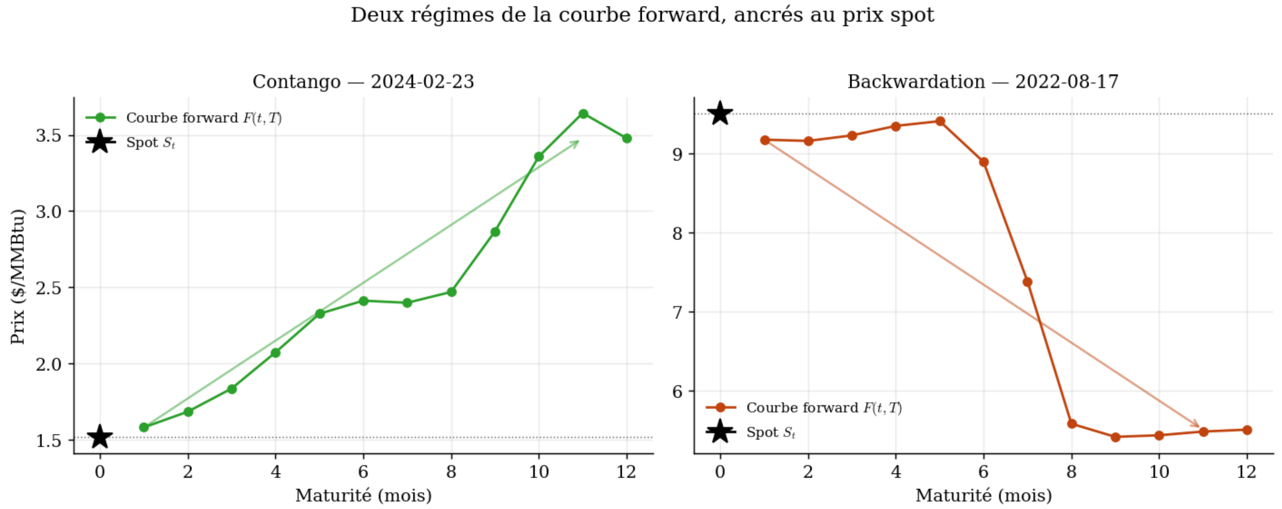

Figure 3: Two regimes of the Henry Hub forward curve, anchored to the spot price (black star at zero maturity): contango (forwards above the spot, upward-sloping curve) and backwardation (forwards below the spot, downward-sloping curve).

This reading is confirmed on volatility. We compute the annualised volatility of each contract as the standard deviation of its daily log-returns, $\sigma_i^{\text{ann}} = \sqrt{252}\,\widehat{\text{std}}(\Delta \log C_{i,t})$. figure 4 shows that it decreases markedly from the front-month (53.9 % for C_1) toward the far maturities (26.7 % for C_{12}). This decrease of volatility with maturity, the *Samuelson effect*, is the empirical signature of a fast-reverting short-term factor: the shocks of χ_t strike the front-month head-on, but are damped, through the weight $e^{-\zeta\tau}$, on the far maturities, which then barely reflect anything but the long-term factor ξ_t. The two-factor structure is thus validated empirically even before estimation. Complementary views of the futures chain, the term structure of the twelve contracts and the contango/backwardation basis over time, appear in appendix D.

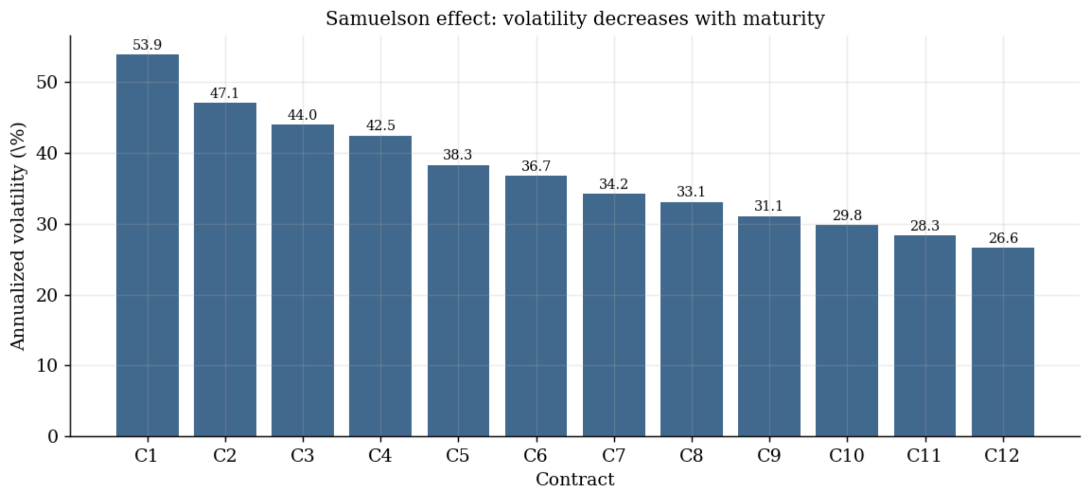

Figure 4: Samuelson effect: annualised volatility of the daily returns of each Henry Hub futures contract, from the front-month C_1 to the one-year maturity C_{12}.

4 State-Space Estimation via the Kalman Filter

Since the two factors χ_t and ξ_t are not observed, their estimation from the futures curve is a matter of filtering a hidden state.

4.1 State-Space Form

We formalise the two equations governing the dynamic system. It is composed of:

- a *state equation*, which describes the dynamics of the latent variables $X_t = (\chi_t, \xi_t)^\top$.
- an *observation equation*, which relates these latent variables to the prices actually quoted.

Since the contracts are rolling (section 3.3), we denote by $F_i(t)$ the price of contract C_i at date t and by $\tau_i(t) = T_i(t) - t$ its residual maturity, which varies over time. We adopt the calibration convention by fixing, for simplicity, the risk premia to zero, $\lambda_\chi = \lambda_\xi = 0$.

The system then writes in linear Gaussian state-space form, as the pair of equations

$$\begin{cases} X_t = c + G X_{t-1} + \omega_t, & \omega_t \sim \mathcal{N}(0, W), \\ Y_t = d_t + F_t^\top X_t + v_t, & v_t \sim \mathcal{N}(0, V), \end{cases} \tag{10}$$

where:

- $X_t = (\chi_t, \xi_t)^\top \in \mathbb{R}^2$ is the latent (unobserved) state vector.
- $Y_t = \big(\log F_1(t), \dots, \log F_M(t)\big)^\top \in \mathbb{R}^M$ stacks the log-prices of the $M = 12$ contracts quoted at t.
- $c \in \mathbb{R}^2$ and $G \in \mathbb{R}^{2\times 2}$ carry the drift and the transition of the state, and ω_t is the state noise with covariance W.
- $d_t \in \mathbb{R}^M$ and $F_t \in \mathbb{R}^{2\times M}$ carry the drift and the observation matrix, and v_t is the measurement noise with covariance $V = s^2 I_M$ (quotation errors, independent across maturities).

The *state equation* comes from the exact discretisation of equation (3) over a step $\Delta t = 1$, which gives

$$c = \begin{pmatrix} 0 \\ \dfrac{\mu_\xi}{\gamma}\big(1 - e^{-\gamma}\big) \end{pmatrix} \quad G = \begin{pmatrix} e^{-\zeta} & 0 \\ 0 & e^{-\gamma} \end{pmatrix} \quad W = \begin{pmatrix} \dfrac{\sigma_\chi^2}{2\zeta}\big(1 - e^{-2\zeta}\big) & \dfrac{\rho\,\sigma_\chi\sigma_\xi}{\zeta+\gamma}\big(1 - e^{-(\zeta+\gamma)}\big) \\ \cdot & \dfrac{\sigma_\xi^2}{2\gamma}\big(1 - e^{-2\gamma}\big) \end{pmatrix}$$

the state noise $\omega_t = (\omega_t^\chi, \omega_t^\xi)^\top$ grouping the stochastic integrals $\omega_t^\chi = \sigma_\chi \int_{t-1}^t e^{-\zeta(t-u)}\,\mathrm{d}Z_u^\chi$ and $\omega_t^\xi = \sigma_\xi \int_{t-1}^t e^{-\gamma(t-u)}\,\mathrm{d}Z_u^\xi$ (appendix A)

The *observation equation* reads directly from the futures formula equation (7), applied to each quoted contract:

$$d_t = \big(A(\tau_1(t)), \dots, A(\tau_M(t))\big)^\top, \qquad F_t = \begin{pmatrix} e^{-\zeta\tau_1(t)} & \dots & e^{-\zeta\tau_M(t)} \\ e^{-\gamma\tau_1(t)} & \dots & e^{-\gamma\tau_M(t)} \end{pmatrix}$$

The vector of parameters to estimate is $\Theta = (\zeta, \gamma, \mu_\xi, \sigma_\chi, \sigma_\xi, \rho, s)$.

4.2 Estimation via the Kalman Filter

Since the state X_t is hidden, we estimate it with the *Kalman filter*, the optimal recursive algorithm for a linear Gaussian system with latent state. We denote by $\mathscr{F}_t$ the information of the observations up to date t. The filter rests on five assumptions:

(H1) *linearity*: the two equations of equation (10) are affine in X_t.

(H2) *Gaussianity* of the noises, $\omega_t \sim \mathcal{N}(0, W)$ and $v_t \sim \mathcal{N}(0, V)$.

(H3) *independence* of the noises, from one another and over time ($\omega_t \perp v_s$ for all t, s, and $\omega_t \perp \omega_s$, $v_t \perp v_s$ for $t \neq s$).

(H4) covariances W and V *known*: here parameterised by Θ, the filter being applied conditionally on Θ at each iteration of the optimiser.

(H5) *initialisation* a_0, P_0 specified.

Without (H2), the filter remains the best *linear* estimator (generalised Gauss–Markov). Gaussianity guarantees its global optimality. It recursively propagates the conditional moments of the state, separating the instant *before* and *after* the arrival of the observation Y_t:

$$\underbrace{a_{t|t-1} = \mathbb{E}[X_t \mid \mathscr{F}_{t-1}],\;\; P_{t|t-1} = \mathrm{Cov}(X_t \mid \mathscr{F}_{t-1})}_{\text{before } Y_t \text{ : prediction}} \qquad \underbrace{a_t = \mathbb{E}[X_t \mid \mathscr{F}_t],\;\; P_t = \mathrm{Cov}(X_t \mid \mathscr{F}_t)}_{\text{after } Y_t \text{ : update}}$$

to which is added the covariance of the prediction error on the observation, $L_{t|t-1} = \mathrm{Cov}(e_t \mid \mathscr{F}_{t-1})$ with $e_t = Y_t - \mathbb{E}[Y_t \mid \mathscr{F}_{t-1}]$. figure 5 summarises the recursive cycle chained at each date.

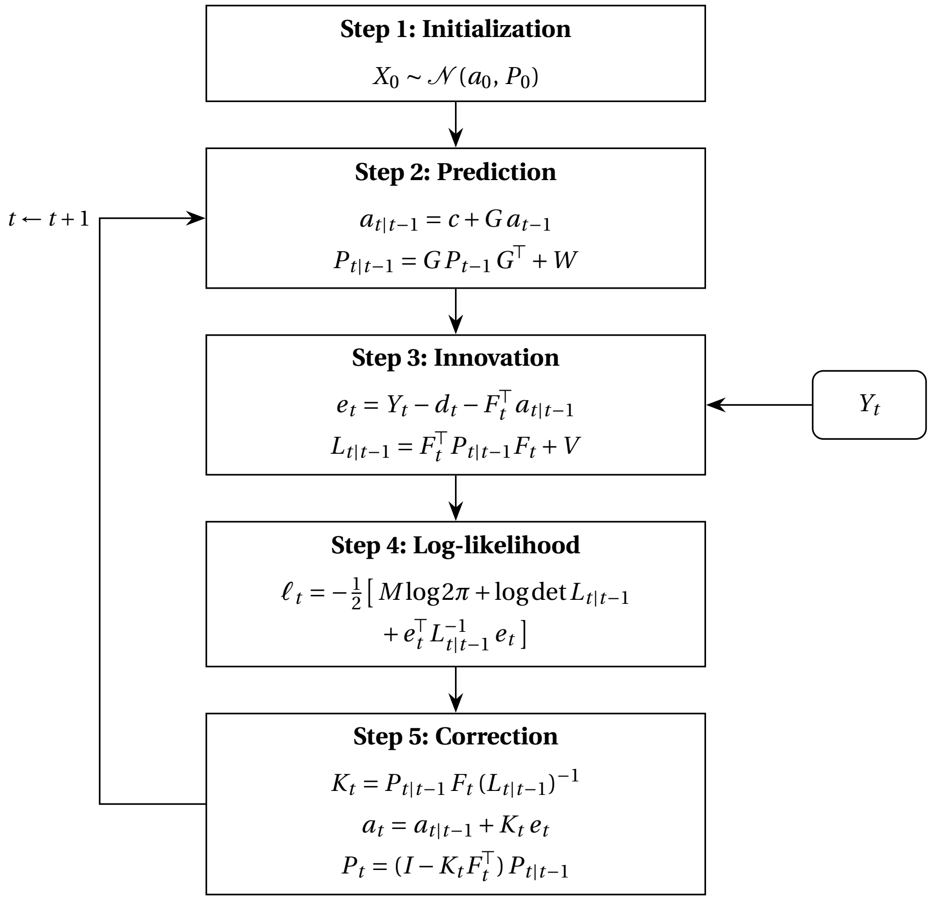

Figure 5: Flowchart of the Kalman filter

For fixed Θ and initialised (a_0, P_0), the filter chains four steps at each date t, whose results we give here, the derivations appearing in appendix B.

Prediction. We propagate the state estimated at the previous step through the dynamics of equation (10), to obtain the best prediction of the current state before observing the new data:

$$a_{t|t-1} = c + G\,a_{t-1}, \qquad P_{t|t-1} = G\,P_{t-1}\,G^\top + W. \tag{11}$$

Innovation. We confront the prediction with the new observation Y_t: the prediction error e_t and its covariance $L_{t|t-1}$ measure the gap between what the model anticipated and what the market quotes,

$$e_t = Y_t - d_t - F_t^\top a_{t|t-1}, \qquad L_{t|t-1} = F_t^\top P_{t|t-1} F_t + V, \tag{12}$$

with $e_t \mid \mathscr{F}_{t-1} \sim \mathcal{N}(0, L_{t|t-1})$.

Computation of the log-likelihood. Since $e_t \mid \mathscr{F}_{t-1}$ is Gaussian in high dimension, the innovation immediately yields the contribution of date t to the log-likelihood:

$$\ell(\Theta, Y) = -\frac{1}{M_T} \sum_{t=1}^{M_T} \log(p(Y_t \mid \mathscr{F}_{t-1})) \tag{13}$$

Moreover, $p(Y_t \mid \mathscr{F}_{t-1}) = \frac{1}{2\pi^{\frac{M_T}{2}}} \det(L_{t|t-1})^{\frac{1}{2}} \exp(-\frac{1}{2} e_t^T L_{t|t-1}^{-1} e_t)$. Passing to the log and summing with the filter, the log-likelihood to maximise in Θ is:

$$\ell(\Theta; Y) = -\frac{1}{2} \sum_{t=1}^{M_T} \Big[M \log(2\pi) + \log\det\big(L_{t|t-1}\big) + e_t^\top L_{t|t-1}^{-1} e_t \Big]. \tag{14}$$

Correction. We finally correct the prediction proportionally to the error, through the *Kalman gain* K_t, which minimises the mean squared error between X_t and a_t and thus balances the confidence between prediction and observation. The updated state (a_t, P_t) then moves on to the next step:

$$\boxed{K_t = P_{t|t-1} F_t L_{t|t-1}^{-1}, \qquad a_t = a_{t|t-1} + K_t e_t, \qquad P_t = \big(I - K_t F_t^\top\big) P_{t|t-1}} \tag{15}$$

The innovation covariance $L_{t|t-1}$, the minimisation giving the gain K_t and the update of P_t are derived in appendix B.

4.3 Why Estimation Fails with Our Data

The model is estimated by maximum likelihood (L-BFGS-B) on the weekly training data 2010–2018 (447 observations of the twelve-maturity curve), the search bounds being economically motivated by the plausible half-lives of the two factors, from two weeks to six months for χ_t and from two to twenty years for ξ_t. The symptom is immediate and reproducible: four different initialisations all converge to the same log-likelihood (NLL = −30.739), but with parameters crushed onto the bounds. The short speed ζ lands on its low economic bound ($\hat{\zeta}$ = 1.403, half-life of six months) and, above all, the long speed $\hat{\gamma}$ = 0.0017 sticks to the lower numerical bound, that is, an implicit half-life of the long-term factor of about 400 years, economically absurd. The optimiser finds no interior minimum: the likelihood is flat in the direction of γ. How can this phenomenon be explained?

4.4 An Identifiability Problem

Our Databento chain stops at a maximum maturity $\tau_{\max} = 1$ year, whereas an economically credible long-term factor has a half-life of several years, hence a low reversion speed γ. Over such a short range of maturities, the weights $e^{-\gamma\tau_i}$ are all close to one, and the second row of the observation matrix F_t flattens onto a constant vector:

$$F_t = \begin{pmatrix} e^{-\zeta\tau_1} & \cdots & e^{-\zeta\tau_M} \\ e^{-\gamma\tau_1} & \cdots & e^{-\gamma\tau_M} \end{pmatrix} \xrightarrow[\gamma\to 0]{} \begin{pmatrix} e^{-\zeta\tau_1} & \cdots & e^{-\zeta\tau_M} \\ 1 & \cdots & 1 \end{pmatrix}.$$

The short-term factor χ_t keeps a row that decreases markedly along the curve, but ξ_t now loads all maturities identically: the observation matrix becomes ill-conditioned and the filter can no longer separate ξ_t from a constant level. figure 6 confirms this at the estimated parameters, where $e^{-\gamma\tau_i} \in [0.998,\ 1]$, a variation of 0.0016 over the twelve contracts, to be compared with the clear decrease of $e^{-\zeta\tau_i}$.

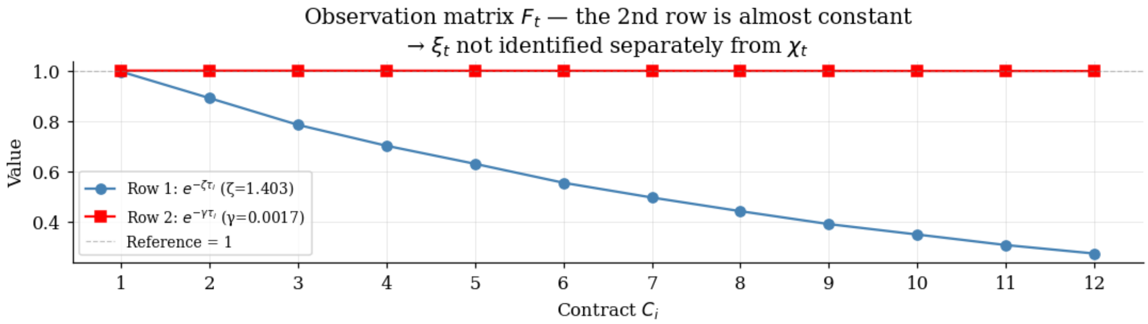

Figure 6: The two rows of the observation matrix F_t at the estimated parameters: $e^{-\zeta\tau_i}$ decreases markedly along the twelve contracts, $e^{-\gamma\tau_i}$ stays flat at 1.

This degeneracy is an *identifiability* defect. A model is identifiable when distinct parameters produce distinct observation laws, that is, when the map $\Theta \mapsto p(y;\Theta)$ is injective. This is no longer the case here: when the second row of F_t is constant, the drift $A(\tau)$ becomes linear in maturity and a shift $(\mu_\xi + \delta,\ \xi_t - \delta\tau)$ leaves each observation unchanged, the additional slope $\delta\tau$ being exactly absorbed by the drop in level.

Yet the injectivity of $\Theta \mapsto p(y;\Theta)$ is precisely the assumption that grounds the consistency and uniqueness of the maximum likelihood estimator. The log-likelihood exhibits a flat direction, along which (μ_ξ, ξ_t, γ) drift without the criterion moving, and the optimiser, for lack of an interior minimum, pushes $\hat{\gamma}$ to its bound. The formal proof, $p(y;\Theta_1) = p(y;\Theta_2)$ with $\Theta_1 \neq \Theta_2$, as well as the quantitative condition on the maturities ($\tau_{\max} \gg 1/\gamma$, that is, about ten years for natural gas), are detailed in appendix C.

Figure 7: Filtered states of the Schwartz–Smith model on the Henry Hub: short-term factor $\hat{\chi}_t$, long-term factor $\hat{\xi}_t$ and reconstruction of the log-price (red zone: out-of-sample).

In figure 7, the reconstruction $\hat{\chi}_t + \hat{\xi}_t$ visually matches the log-price, including out-of-sample, but this quality validates nothing, since any sufficiently flexible additive model reconstructs the series it is given. Out-of-sample (2019–2024), the frozen parameters moreover degrade markedly, the log-price reconstruction RMSE doubling from 0.071 to 0.156. The objective was the *economic* identification of the two factors.

The conclusion of this detour is twofold. On the one hand, it is a *data* problem and not a method problem: with a long curve, beyond five to ten years of maturity, the same code would identify the model, as Schwartz and Smith (2000) do on futures going up to nine years.

5 Targeting Volatility: a Weather-Coupled GARCH

The structural detour showed that the price *level* resists our data. Its *volatility*, by contrast, requires only the daily spot returns, which we have over nearly thirty years. We therefore work on the log returns $r_t = \log(S_t/S_{t-1})$. We observe volatility regimes (*clustering*), a characteristic of financial series, markedly more pronounced in winter. To model this effect, a GARCH model retains a memory of the past. Since the mean of the returns is statistically zero, we take a GARCH(1,1) as our baseline:

$$r_t = \varepsilon_t, \qquad \sigma_t^2 = \omega + \alpha\,\varepsilon_{t-1}^2 + \beta\,\sigma_{t-1}^2, \tag{16}$$

We seek to enrich it with exogenous weather variables, each addition being motivated by an empirical diagnostic. All models are estimated by maximum likelihood on the daily Henry Hub returns (training sample, 2785 observations).

5.1 Building a winter temperature anomaly

A day with high HDD does not mechanically imply a rise in volatility: the market has already priced in the normal seasonal drop in temperatures. What carries information is the *surprise*, the days markedly colder than the seasonal norm. We define the positive HDD anomaly on day t as

$$\widetilde{\mathrm{HDD}}_t^+ = \max\big(\mathrm{HDD}_t - \overline{\mathrm{HDD}}(m_t, d_t),\ 0\big), \tag{17}$$

where $\overline{\mathrm{HDD}}(m_t, d_t)$ is the historical mean HDD for the same calendar day (m_t the month, d_t the day) over the 2000–2019 reference period. We keep only the positive part: it is the days colder than the norm that create unanticipated pressure on gas demand.

To represent heating demand over the study area, we aggregate four cities into a weighted composite:

$$\widetilde{\mathrm{HDD}}_t^{+,\mathrm{agg}} = \sum_{c\in\mathscr{C}} w_c\,\widetilde{\mathrm{HDD}}_{c,t}^+, \qquad w_c = \frac{\bar{Q}_c}{\sum_{c'\in\mathscr{C}} \bar{Q}_{c'}}, \tag{18}$$

with $\mathscr{C}$ = {Chicago, Boston, New York, Minneapolis}. Rather than population, we use each state's *residential gas consumption* (EIA data), which directly captures the intensity of temperature-sensitive demand: a city of comparable size in Florida structurally consumes less heating gas than a Midwestern city. The weights $\bar{Q}_c$, each state's share in the total consumption of the four, averaged over 2000–2024 to stabilize them, give Chicago 38.8 %, New York 38.1 %, Minneapolis 12.0 % and Boston 11.1 %.

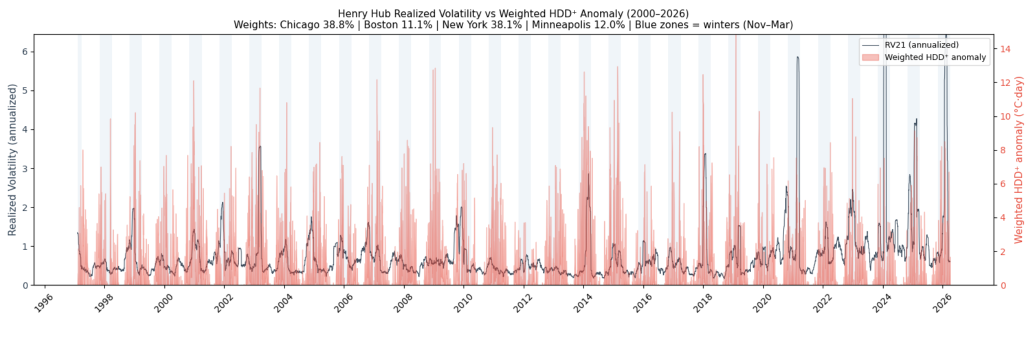

Figure 8: Henry Hub realized volatility (annualized RV21) and weighted composite HDD$^+$ anomaly, 2000–2026. Blue zones over the winters (November–March).

figure 8 overlays the 21-day realized volatility on this composite anomaly. First, the volatility spikes are almost exclusively wintertime, which suggests that cold-related demand shocks are a determinant of gas volatility. Second, the relationship is not mechanical: some winters with large anomalies produce no spike (2011, 2013), while others explode (2014, 2021). The effect of cold is therefore *conditional* on other factors. As exogenous factors, one might think of the moment when the surprise hits as well as the natural gas storage level.

5.2 Distribution of anomalies by city

Before coupling this variable to volatility, we examine its distribution city by city (figure 9). On the days colder than the norm alone, the four distributions share the same shape: a mode near zero and a *fat tail* toward the large anomalies. Positive-anomaly days represent about a third of the observations, with a mean on the order of 3 to 4 °C·day, but extremes exceeding 10 to 20 °C·day and a markedly positive excess kurtosis. It is precisely these extreme values, rare but intense, that are liable to trigger the high-volatility episodes, and that motivate the tail analysis that follows.

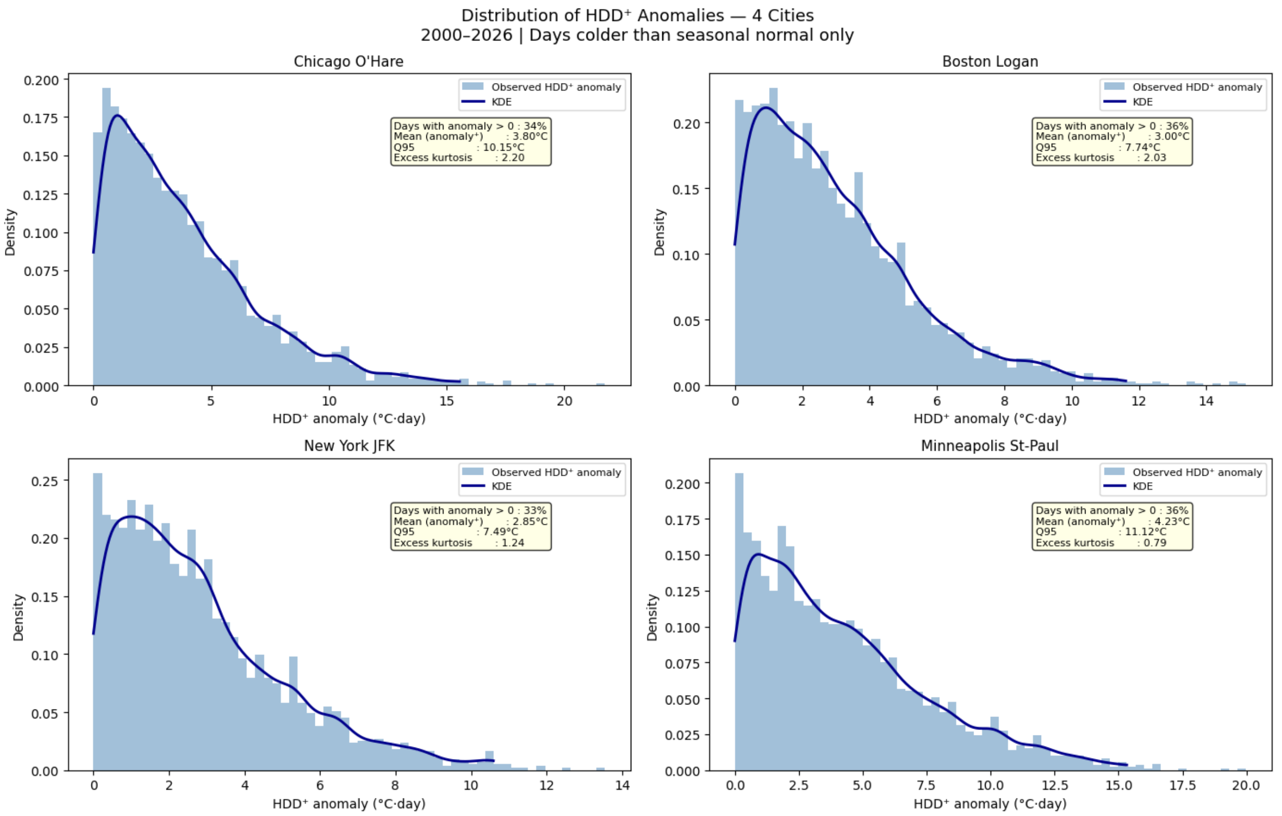

Figure 9: Distribution of the HDD$^+$ anomaly by city (days colder than the seasonal norm, 2000–2026)

5.3 Fat tail investigation

To study the behavior of the extreme values, we resort to a *quantile regression*, which generalizes least-squares regression by estimating, instead of the conditional mean of Y given X, its *conditional quantile* $Q_\tau[Y \mid X]$ for $\tau \in (0,1)$. On weekly winter data (579 observations), we regress realized volatility on the composite anomaly,

$$Q_\tau\big[RV_t \mid \widetilde{\mathrm{HDD}}_t^{+,\mathrm{agg}}\big] = \alpha_\tau + \gamma_\tau\, \widetilde{\mathrm{HDD}}_t^{+,\mathrm{agg}}, \tag{19}$$

the coefficient γ_τ being estimated by minimizing the asymmetric (pinball) loss

$$\hat{\gamma}_\tau = \arg\min_{\gamma} \sum_t \rho_\tau\big(RV_t - \alpha - \gamma\, \widetilde{\mathrm{HDD}}_t^{+,\mathrm{agg}}\big), \qquad \rho_\tau(u) = u\big(\tau - \mathbf{1}_{u<0}\big), \tag{20}$$

which penalizes positive and negative residuals differently depending on τ. We seek to see whether the effect is *concentrated in the tails*: what is hard to anticipate is not a moderate anomaly, but an intense winter.

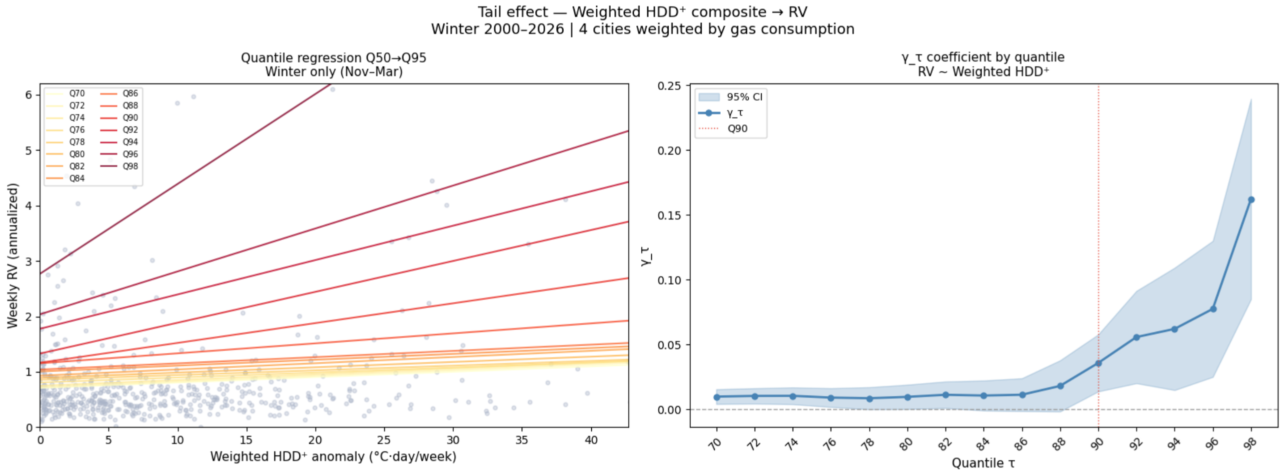

Figure 10: Tail effect of the weighted HDD$^+$ composite on the weekly realized volatility (winters 2000–2026, 579 observations): quantile regression lines Q50–Q95 (left) and coefficient γ_τ with 95 % CI (right).

figure 10 plots on the left, as a scatter plot, the HDD anomaly on the x-axis and the realized volatility on the y-axis, with the quantile regression lines overlaid. The coefficient γ_τ (the slope of the regression) is indistinguishable from zero between Q50 and Q80, then grows sharply beyond, as the right-hand figure shows. The weather effect does not shift the center of the volatility distribution, it inflates its *tails*.

The HDD–volatility relationship is therefore **nonlinear** and amplifies in the high-volatility regimes. One caveat, however: the small number of intense winters in the sample greatly widens the confidence intervals at the highest quantiles, the ones we care about most.

5.4 Timing of the anomaly within the winter

To be a little more precise, the γ_τ coefficients above are global: they average the effect over the whole season. But does a November anomaly have the same impact as a February anomaly? To test this, we re-estimate γ_τ on *rolling six-week windows* along the winter, aggregating all the available years:

$$\hat{\gamma}_\tau(s) = \arg\min_{\gamma} \sum_{t \in \mathscr{W}(s)} \rho_\tau\big(RV_t - \alpha - \gamma\, \widetilde{\mathrm{HDD}}_t^{+,\mathrm{agg}}\big), \tag{21}$$

where $\mathscr{W}(s)$ denotes the set of observations from the six weeks centered on s.

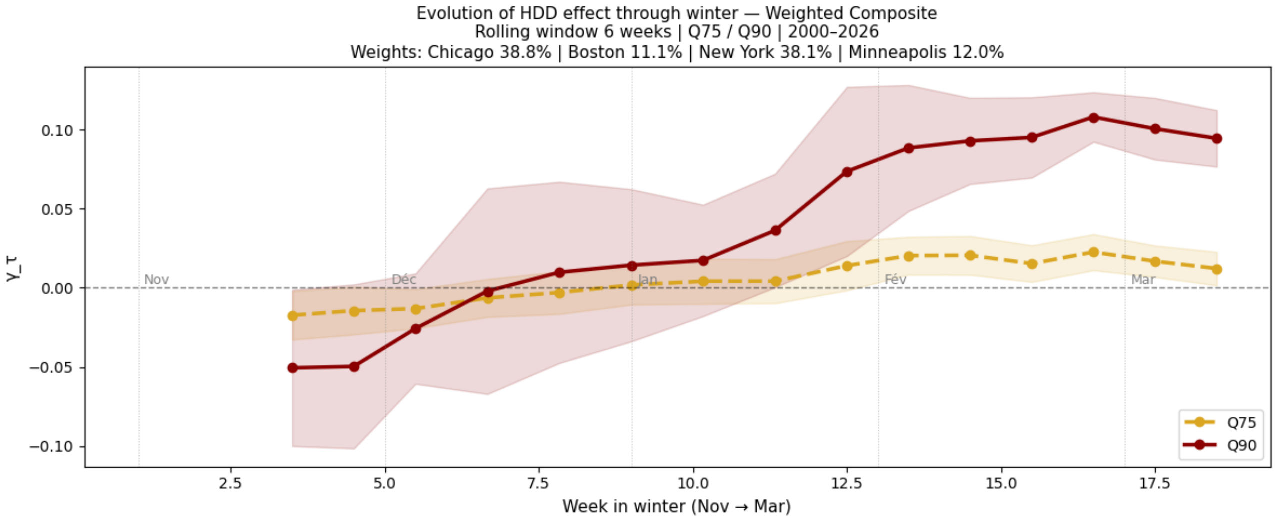

Figure 11: Evolution of the coefficient γ_τ (quantiles 75 and 90) along the winter, quantile regression on rolling six-week windows, weighted composite, 2000–2026.

figure 11 confirms a *late-winter* effect. The coefficient γ_{90} is null, or even negative, in November–December, becomes positive as early as January and peaks in February around 0.10, well above the 75th quantile, which stays flat. The same degree-day of anomaly is therefore not worth the same volatility depending on the moment of the winter, and only the upper tail (Q90) carries this effect.

The explanation is economic: in early winter, gas storage is full and cushions the demand shock. As the winter progresses it empties, and a late cold episode can then only be absorbed by the price. This mechanism calls for a storage variable in the GARCH.

5.5 Relative storage and models A, B, C

We measure storage by the weekly level I_t (EIA data, in Bcf, billion cubic feet). What governs the amplitude of the spikes is not the absolute level but its deviation from the seasonal norm, captured by the *relative storage anomaly*

$$\tilde{I}_t = \frac{I_t - \bar{I}_w}{\bar{I}_w}, \tag{22}$$

where $\bar{I}_w$ is the historical mean storage for calendar week w.

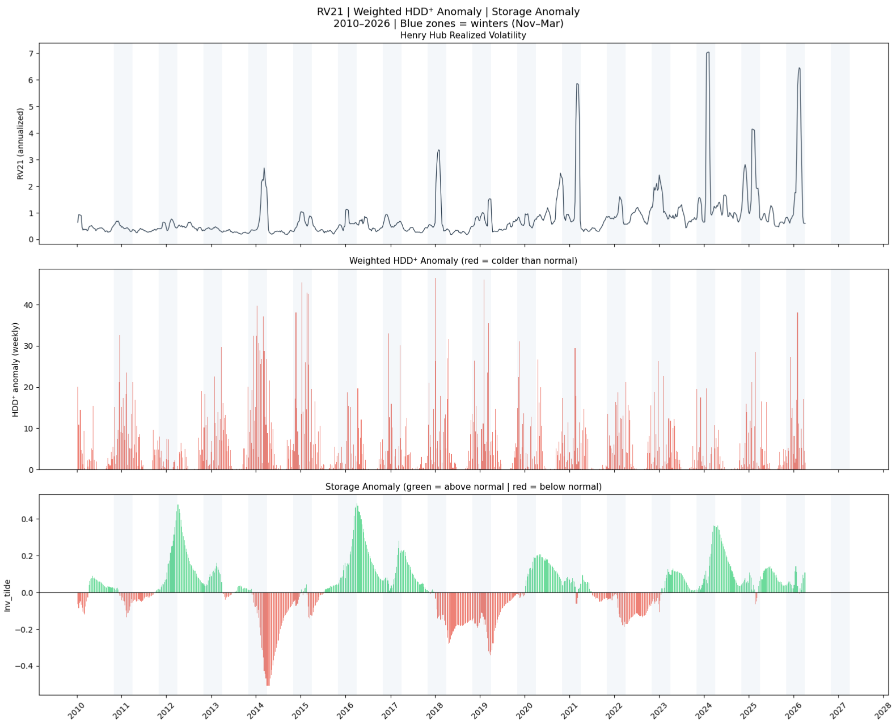

Figure 12: 2010–2026: realized volatility RV21, weighted weekly HDD^+ anomaly, and relative storage anomaly $\tilde{I}_t$ (green: above the norm. Red: below).

figure 12 reveals the asymmetry of transmission. In late winter, storage is structurally low after several weeks of withdrawal, but not all winters are alike. When $\tilde{I}_t > 0$ (2012, 2016), the market keeps a safety cushion and a late cold shock remains absorbable. When $\tilde{I}_t < 0$ (2014, 2018, 2022), storage is below the norm, the market would anticipate a risk of physical shortage, and even a moderate anomaly is enough to make volatility explode. The largest RV21 spikes in the sample correspond most of the time to these configurations.

We then test three couplings, all estimated by maximum likelihood, whose mean equation reduces to $r_t = \varepsilon_t$. *Model A* is the GARCH(1,1) baseline of equation (16). *Model B* adds to it the day's composite anomaly, the term $\gamma\,\widehat{\mathrm{HDD}}_t^+$. *Model C* further adds a storage × late-winter interaction, $\delta\,\widehat{\mathrm{HDD}}_t^+ \cdot \max(-\tilde{I}_t, 0) \cdot \mathbf{1}_{\text{jan–mar}}$, which captures a precise nonlinearity: the same cold shock amplifies the variance only if two conditions are met, storage below the norm ($\tilde{I}_t < 0$) and an occurrence in late winter.

Table 3: Models A, B and C estimated by maximum likelihood

Model	ω	α	β	γ (p)	δ (p)	**AIC**
A – GARCH(1,1)	0.267	0.116	0.866	—	—	14250.2
B – $+\gamma\,\widehat{\mathrm{HDD}}^+$	0.235	0.113	0.862	0.198 (0.005)	—	**14241.3**
C – $+\delta$ storage×winter	0.235	0.113	0.862	0.198 (0.007)	0.000 (1.000)	14243.3

The coefficient $\hat{\gamma} = 0.198$ of model B is significant ($p = 0.005$) and its AIC gains nine points over the baseline (14241 versus 14250), a result confirmed by the BIC: there is indeed a coupling between temperature anomalies and conditional variance. The persistence $\hat{\alpha} + \hat{\beta}$ drops from 0.982 to 0.976. Part of the “persistent” variance was in reality seasonal weather volatility, absorbed by the GARCH term for lack of an exogenous variable. Model B is retained as the reference specification of the coupling. Model C, on the other hand, fails: $\hat{\delta} = 0$ is not significant ($p = 1.000$), the log-likelihood is identical to that of B and the AIC deteriorates.

This does not mean, however, that the storage effect is economically absent: two structural factors explain the non-identification of δ. On one hand a data problem: the episodes meeting the three conditions (cold anomaly, $\tilde{I}_t < 0$, January–March) amount to only 276 observations out of 2785, less than 10 % of the sample. On the other hand, absorption by persistence: with $\hat{\beta} = 0.862$, the late-winter variance is already high via $\beta\sigma^2_{t-1}$, inherited from the early-season shocks, and the optimizer no longer finds any residual signal to attribute to δ.

5.6 Splitting the effect within the winter: model D

The failure of model C invites us to reformulate the timing effect not through an interaction, but through γ coefficients specific to each half of the winter. *Model D* distinguishes the effect of an early-winter anomaly (November–December) from that of a late-winter anomaly (January–March):

$$\sigma^2_t = \omega + \alpha\,\varepsilon^2_{t-1} + \beta\,\sigma^2_{t-1} + \gamma_{\text{early}}\,\widetilde{\text{HDD}}^+_t\,\mathbf{1}_{\{\text{nov, dec}\}} + \gamma_{\text{late}}\,\widetilde{\text{HDD}}^+_t\,\mathbf{1}_{\{\text{jan, feb, mar}\}}, \tag{23}$$

and allows us to test the hypothesis $H_0 : \gamma_{\text{early}} = \gamma_{\text{late}}$.

Table 4: Model D (early/late-winter split) against the baseline A. Wald test $H_0 : \gamma_{\text{early}} = \gamma_{\text{late}}$: $p = 0.061$. For reference, model B AIC = 14241.3.

Model	ω	α	β	γ_{early} (p)	γ_{late} (p)	**AIC**
A – GARCH(1,1)	0.267	0.116	0.866	—	—	14250.2
D – early/late	0.283	0.110	0.861	**0.618 (0.001)**	0.190 (0.128)	**14231.2**

The result (table 4) is counterintuitive in light of the timing analysis: it is the *early*-winter coefficient that carries the information. $\hat{\gamma}_{\text{early}} = 0.618$ is significant at 1 % ($p = 0.001$), while $\hat{\gamma}_{\text{late}} = 0.190$ is not ($p = 0.128$). Model D dominates the baseline A as well as model B in AIC (14231, versus 14250 and 14241).

This apparent paradox resolves once the model's memory is removed. Re-estimated as ARCH-X (model D′, without the $\beta\sigma^2_{t-1}$ term), the ranking reverses: both coefficients become highly significant again and it is γ_{late} that dominates ($\hat{\gamma}_{\text{late}} = 6.15$ versus $\hat{\gamma}_{\text{early}} = 3.17$). Without $\beta\sigma^2_{t-1}$, the model has no way to condition the late-winter variance on the recent past. γ_{late} then absorbs both the marginal effect of late cold and the variance inherited from previous shocks, and is mechanically overestimated. The AIC nonetheless comes down unambiguously in favor of the GARCH-X (14231 versus 14661, a gap of 430 points): the GARCH memory, via $\hat{\beta} = 0.861$, is indispensable to the dynamics of gas volatility. The late-winter effect is indeed real, but it is already **encoded in this memory**: the variance from January to March inherits the cold accumulated since November, so that the late-winter anomaly no longer adds anything at the margin.

Model D, with its early-winter effect γ_{early}, is retained as the best specification of the weather–volatility coupling. The analysis of models A through D sketches a common logic: to bring new information to the conditional variance, a variable must contain what neither the memory $\beta\sigma^2_{t-1}$ nor the realized degree-days already capture.

6 Conclusion

This component followed the trace of temperature in the price of natural gas, in two stages. The structural route first: the two-factor Schwartz–Smith model, cast in state-space form and estimated by Kalman filter on the futures curve, did not succeed. With maturities limited to one year, the long-term factor is non-identifiable and the optimizer collapses its parameters onto their bounds. The failure remains instructive, because it is quantified and reproducible: a data limitation, not a methodological one, which a long curve would lift.

The volatility route then, which succeeds. A GARCH(1,1) already captures on its own most of the spot dynamics, and while our analyses of degree-day anomalies, storage, and winter timing made it possible to build a weather-coupled model that beats it, model D with its significant early-winter effect, the coupling remains modest, largely absorbed by the GARCH memory. The research results are nonetheless satisfactory given the data available.

Two natural extensions emerge: estimating the Schwartz–Smith model on a long futures curve to lift its identification, and replacing the r_t^2 proxy with an intraday realized volatility to refine the GARCH component.

References

Binkowski, Karol, Peilun He, Nino Kordzakhia, and Pavel Shevchenko (2021). “On the Parameter Estimation in the Schwartz–Smith’s Two-Factor Model”. In: *arXiv preprint arXiv:2108.01881*. DOI: 10.48550/arXiv.2108.01881.

Chen, Yanting, Peter R. Hartley, and Yihui Lan (2023). “Temperature, storage, and natural gas futures prices”. In: *Journal of Futures Markets* 43.7, pp. 1006–1031. DOI: 10.1002/fut.22402.

Cheng, Tuoyuan, Saikiran Reddy Poreddy, and Kan Chen (2025). “Tail risk in weather derivatives”. In: *Commodities* 4.2, p. 11. DOI: 10.3390/commodities4020011.

Müller, Jan, Guido Hirsch, and Alfred Müller (2015). “Modeling the Price of Natural Gas with Temperature and Oil Price as Exogenous Factors”. In: *Innovations in Quantitative Risk Management*. Ed. by Kathrin Glau, Matthias Scherer, and Rudi Zagst. Vol. 99. Springer Proceedings in Mathematics & Statistics. Springer, pp. 109–128. DOI: 10.1007/978-3-319-09114-3_7.

Schwartz, Eduardo and James E. Smith (2000). “Short-term variations and long-term dynamics in commodity prices”. In: *Management Science* 46.7, pp. 893–911. DOI: 10.1287/mnsc.46.7.893.12034.

Appendix A Moments of the Schwartz–Smith Model

This appendix establishes the conditional moments of the state vector $X_t = (\chi_t, \xi_t)^\top$ of section 3.1, directly from the closed-form solution of the two stochastic differential equations under $\mathbb{Q}$, without resorting to a discretisation. From these it derives the distribution of the log-price and the futures formula, then gives the exact discretisation used by the filter.

A.1 Solving the two processes

The two factors are Ornstein–Uhlenbeck processes. We solve by variation of constants, applying Itô's formula to $e^{\zeta t}\chi_t$ for the short-term factor under $\mathbb{Q}$ (dynamics equation (3)):

$$\mathrm{d}\big(e^{\zeta t}\chi_t\big) = e^{\zeta t}\big(\mathrm{d}\chi_t + \zeta\,\chi_t\,\mathrm{d}t\big) = e^{\zeta t}\big(-\lambda_\chi\,\mathrm{d}t + \sigma_\chi\,\mathrm{d}Z_t^\chi\big).$$

Integrating from s to t and then multiplying by $e^{-\zeta t}$, and likewise for ξ_t, we obtain the solved form equation (4):

$$\begin{cases} \chi_t = \chi_s\,e^{-\zeta(t-s)} - \dfrac{\lambda_\chi}{\zeta}\big(1-e^{-\zeta(t-s)}\big) + \sigma_\chi \displaystyle\int_s^t e^{-\zeta(t-u)}\,\mathrm{d}Z_u^\chi, \\ \xi_t = \xi_s\,e^{-\gamma(t-s)} + \dfrac{\mu_\xi-\lambda_\xi}{\gamma}\big(1-e^{-\gamma(t-s)}\big) + \sigma_\xi \displaystyle\int_s^t e^{-\gamma(t-u)}\,\mathrm{d}Z_u^\xi. \end{cases} \tag{A.1}$$

Each factor is the sum of a deterministic term and a Wiener integral, hence Gaussian, and the pair (χ_t, ξ_t) is a Gaussian vector.

A.2 Expectation

The stochastic integrals in equation (A.1) have zero expectation. Setting $s = 0$, only the deterministic parts remain:

$$\mathbb{E}[\chi_t] = \chi_0\,e^{-\zeta t} - \frac{\lambda_\chi}{\zeta}\big(1-e^{-\zeta t}\big), \qquad \mathbb{E}[\xi_t] = \xi_0\,e^{-\gamma t} + \frac{\mu_\xi-\lambda_\xi}{\gamma}\big(1-e^{-\gamma t}\big).$$

A.3 Variances

The variance of each factor depends only on its stochastic part. By the Itô isometry, $\mathbb{E}\big[\big(\int_0^t f(u)\,\mathrm{d}Z_u\big)^2\big] = \int_0^t f(u)^2\,\mathrm{d}u$, hence

$$\mathrm{Var}(\chi_t) = \sigma_\chi^2 \int_0^t e^{-2\zeta(t-u)}\,\mathrm{d}u = \frac{\sigma_\chi^2}{2\zeta}\big(1-e^{-2\zeta t}\big), \qquad \mathrm{Var}(\xi_t) = \sigma_\xi^2 \int_0^t e^{-2\gamma(t-u)}\,\mathrm{d}u = \frac{\sigma_\xi^2}{2\gamma}\big(1-e^{-2\gamma t}\big).$$

A.4 Covariance between the two factors

In this case, only the stochastic parts contribute:

$$\mathrm{Cov}(\chi_t, \xi_t) = \mathrm{Cov}\left(\sigma_\chi \int_0^t e^{-\zeta(t-u)}\,\mathrm{d}Z_u^\chi,\ \sigma_\xi \int_0^t e^{-\gamma(t-u)}\,\mathrm{d}Z_u^\xi\right).$$

Since the two Brownian motions are correlated, $\mathrm{d}Z_t^\chi\,\mathrm{d}Z_t^\xi = \rho\,\mathrm{d}t$, we can write $Z_t^\xi = \rho\,Z_t^\chi + \sqrt{1-\rho^2}\,B_t$, where B is a Brownian motion independent of Z^χ. As the integral against B contributes no covariance, it follows that

$$\mathrm{Cov}(\chi_t, \xi_t) = \rho\,\sigma_\chi\sigma_\xi\,\mathrm{Cov}\left(\int_0^t e^{-\zeta(t-u)}\,\mathrm{d}Z_u^\chi,\ \int_0^t e^{-\gamma(t-u)}\,\mathrm{d}Z_u^\chi\right).$$

By the generalised Itô isometry, $\mathbb{E}\big[\big(\int_0^t f\,\mathrm{d}Z\big)\big(\int_0^t g\,\mathrm{d}Z\big)\big] = \int_0^t f(u)\,g(u)\,\mathrm{d}u$, we finally obtain

$$\mathrm{Cov}(\chi_t, \xi_t) = \rho\,\sigma_\chi\sigma_\xi \int_0^t e^{-(\zeta+\gamma)(t-u)}\,\mathrm{d}u = \boxed{\frac{\rho\,\sigma_\chi\sigma_\xi}{\zeta+\gamma}\big(1-e^{-(\zeta+\gamma)t}\big)}.$$

These three quantities make up the matrix $\mathrm{Var}(X_t)$ of equation (5).

A.5 Log-normality and futures prices

The log-spot $\log S_t = \chi_t + \xi_t$ is Gaussian, with mean $\mathbb{E}[\chi_t] + \mathbb{E}[\xi_t]$ and variance $\mathrm{Var}(\chi_t) + \mathrm{Var}(\xi_t) + 2\,\mathrm{Cov}(\chi_t, \xi_t)$, which gives equation (6). The spot $S_t = e^{\log S_t}$ is therefore log-normal. For $X \sim \mathcal{N}(m, v^2)$, completing the square in the numerator and then setting $u = \big(x - (m + v^2)\big)/v$,

$$\mathbb{E}\big[e^X\big] = \int_{\mathbb{R}} e^x \frac{1}{\sqrt{2\pi v^2}}\, e^{-\frac{(x-m)^2}{2v^2}}\,\mathrm{d}x = e^{m+\frac{v^2}{2}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u^2}{2}}\,\mathrm{d}u = e^{m+\frac{v^2}{2}},$$

the remaining Gaussian integral being equal to one. Since the price of a futures contract with maturity T is $F(t,T) = \mathbb{E}_t^{\mathbb{Q}}[S_T]$, conditional log-normality gives

$$\log F(t,T) = \mathbb{E}_t^{\mathbb{Q}}\big[\log S_T\big] + \tfrac{1}{2}\mathrm{Var}_t^{\mathbb{Q}}\big(\log S_T\big) = e^{-\zeta(T-t)}\chi_t + e^{-\gamma(T-t)}\xi_t + A(T-t),$$

where $A(\tau)$ collects the deterministic and variance terms of equation (8), which establishes equation (7).

A.6 Exact discretisation for the filter

For estimation, we apply the solved form equation (A.1) over a step Δt (from $t-1$ to t), with zero risk premia ($\lambda_\chi = \lambda_\xi = 0$). We recover exactly the state equation equation (10), $X_t = c + G\,X_{t-1} + \omega_t$, with $G = \mathrm{diag}\big(e^{-\zeta\Delta t}, e^{-\gamma\Delta t}\big)$ and $c = \big(0, \frac{\mu_\xi}{\gamma}(1-e^{-\gamma\Delta t})\big)^\top$. The state noise ω_t is the vector of stochastic integrals over $[t-1, t]$. Its exact covariance is obtained by the same computations as above, over an interval of length Δt:

Appendix B Derivation of the Kalman Filter

This appendix derives the Kalman filter results stated in section 4.2, on the state-space system equation (10) and under assumptions (H1)–(H5). We write $\mathscr{F}_t$ for the information generated by the observations up to date t, and we recall the filter quantities: $a_{t|t-1} = \mathbb{E}[X_t \mid \mathscr{F}_{t-1}]$ and $a_t = \mathbb{E}[X_t \mid \mathscr{F}_t]$, the predicted and then updated state. $P_{t|t-1} = \mathrm{Cov}(X_t \mid \mathscr{F}_{t-1})$ and $P_t = \mathrm{Cov}(X_t \mid \mathscr{F}_t)$, their uncertainties, and $L_{t|t-1} = \mathrm{Cov}(e_t \mid \mathscr{F}_{t-1})$, the covariance of the prediction error on the observation $e_t = Y_t - \mathbb{E}[Y_t \mid \mathscr{F}_{t-1}]$. By induction, X_t is a linear combination of independent Gaussians, so Y_t and $Y_t \mid \mathscr{F}_{t-1}$ are Gaussian: all the conditional distributions of the filter are Gaussian and fully described by these moments.

B.1 Prediction

Applying the conditional expectation and covariance to the state equation, with $\omega_t \perp \mathscr{F}_{t-1}$:

$$a_{t|t-1} = c + G\,a_{t-1}, \qquad P_{t|t-1} = \mathrm{Cov}\big(GX_{t-1} + \omega_t \mid \mathscr{F}_{t-1}\big) = G\,P_{t-1}\,G^\top + W. \tag{B.1}$$

B.2 Innovation

Since the observation $Y_t = d_t + F_t^\top X_t + v_t$ is Gaussian, its prediction is $\mathbb{E}[Y_t \mid \mathscr{F}_{t-1}] = d_t + F_t^\top a_{t|t-1}$, the measurement noise v_t being centered and independent of $\mathscr{F}_{t-1}$. The prediction error is therefore written

$$e_t = Y_t - d_t - F_t^\top a_{t|t-1} = F_t^\top \left(X_t - a_{t|t-1}\right) + v_t. \tag{B.2}$$

Since $\mathbb{E}[e_t \mid \mathscr{F}_{t-1}] = 0$, its conditional covariance is $L_{t|t-1} = \mathbb{E}\left[e_t e_t^\top \mid \mathscr{F}_{t-1}\right]$. Substituting equation (B.2) and expanding, the two cross terms in $(X_t - a_{t|t-1})\, v_t^\top$ vanish, because v_t is centered and independent of X_t and of $\mathscr{F}_{t-1}$:

$$\mathbb{E}\left[(X_t - a_{t|t-1})\, v_t^\top \mid \mathscr{F}_{t-1}\right] = \mathbb{E}\left[X_t - a_{t|t-1} \mid \mathscr{F}_{t-1}\right] \mathbb{E}[v_t]^\top = 0.$$

There remain the state term, where $\mathbb{E}\left[(X_t - a_{t|t-1})(X_t - a_{t|t-1})^\top \mid \mathscr{F}_{t-1}\right] = P_{t|t-1}$ by definition, and the measurement term $\mathbb{E}\left[v_t v_t^\top\right] = V$:

$$L_{t|t-1} = F_t^\top P_{t|t-1} F_t + V, \tag{B.3}$$

so that $e_t \mid \mathscr{F}_{t-1} \sim \mathcal{N}\left(0, L_{t|t-1}\right)$.

B.3 Log-likelihood computation

Since each innovation satisfies $e_t \mid \mathscr{F}_{t-1} \sim \mathcal{N}(0, L_{t|t-1})$, the joint distribution of the observations factorizes into a product of conditionals, $\mathscr{L}(\Theta; Y) = \prod_{t=1}^{M_T} p\left(Y_t \mid \mathscr{F}_{t-1}\right)$, and each conditional is Gaussian, $Y_t \mid \mathscr{F}_{t-1} \sim \mathcal{N}\left(d_t + F_t^\top a_{t|t-1},\, L_{t|t-1}\right)$. The log-likelihood is therefore written directly from the filter outputs:

$$\ell(\Theta; Y) = -\frac{1}{2} \sum_{t=1}^{M_T} \left[M \log(2\pi) + \log\det\left(L_{t|t-1}\right) + e_t^\top L_{t|t-1}^{-1} e_t \right], \tag{B.4}$$

each pass of the filter at date t providing its contribution. Estimation consists in numerically maximizing ℓ over Θ, the filter being rerun at each evaluation.

B.4 Correction and Kalman gain

Once Y_t is observed, we correct the prediction linearly in the error, $a_t = a_{t|t-1} + K_t e_t$, and we choose the gain matrix K_t that minimizes the mean squared error $\mathbb{E}\left[\|X_t - a_t\|^2 \mid \mathscr{F}_t\right]$.

The criterion reduces to $\mathrm{tr}(P_t)$. Expanding the norm component by component,

$$\mathbb{E}\left[\|X_t - a_t\|^2 \mid \mathscr{F}_t\right] = \mathbb{E}\left[(X_t - a_t)^\top (X_t - a_t) \mid \mathscr{F}_t\right] = \sum_i \mathbb{E}\left[(X_t - a_t)_i^2 \mid \mathscr{F}_t\right],$$

and, since $a_{t,i} = \mathbb{E}[X_{t,i} \mid \mathscr{F}_t]$, each term of the sum is a conditional variance:

$$\mathbb{E}\left[(X_t - a_t)_i^2 \mid \mathscr{F}_t\right] = \mathbb{E}\left[X_{t,i}^2 \mid \mathscr{F}_t\right] - 2\, a_{t,i} \underbrace{\mathbb{E}\left[X_{t,i} \mid \mathscr{F}_t\right]}_{=\, a_{t,i}} + a_{t,i}^2 = \mathbb{E}\left[X_{t,i}^2 \mid \mathscr{F}_t\right] - a_{t,i}^2 = \mathrm{Var}\left(X_{t,i} \mid \mathscr{F}_t\right) = [P_t]_{ii}.$$

The sum of the conditional variances is the trace, $\mathbb{E}\left[\|X_t - a_t\|^2 \mid \mathscr{F}_t\right] = \sum_i [P_t]_{ii} = \mathrm{tr}(P_t)$, so that K_t is the matrix that minimizes $\mathrm{tr}(P_t)$. We therefore express P_t as a function of K_t.

Error after correction. We set $\varepsilon_t = X_t - a_{t|t-1}$, the prediction error on the state. Substituting $e_t = F_t^\top \varepsilon_t + v_t$ (equation (B.2)) into the correction,

$$X_t - a_t = \varepsilon_t - K_t e_t = \varepsilon_t - K_t\left(F_t^\top \varepsilon_t + v_t\right) = \left(I - K_t F_t^\top\right) \varepsilon_t - K_t\, v_t.$$

Expansion of P_t into four terms. With the rule $(A-B)(A-B)^\top = AA^\top - AB^\top - BA^\top + BB^\top$ applied to $A = (I - K_t F_t^\top)\varepsilon_t$ and $B = K_t v_t$,

$$
\begin{aligned}
(X_t - a_t)(X_t - a_t)^\top = &\left(I - K_t F_t^\top\right)\varepsilon_t \varepsilon_t^\top \left(I - K_t F_t^\top\right)^\top && \text{(i)}\\
&- \left(I - K_t F_t^\top\right)\varepsilon_t v_t^\top K_t^\top && \text{(ii)}\\
&- K_t v_t \varepsilon_t^\top \left(I - K_t F_t^\top\right)^\top && \text{(iii)}\\
&+ K_t v_t v_t^\top K_t^\top. && \text{(iv)}
\end{aligned}
$$

In conditional expectation, the two cross terms (ii) and (iii) vanish: v_t is centered and independent of ε_t (a function of X_t and of $\mathscr{F}_{t-1}$) as of $\mathscr{F}_t$, so

$$\mathbb{E}\left[\text{(ii)} \mid \mathscr{F}_t\right] = \left(I - K_t F_t^\top\right)\mathbb{E}\left[\varepsilon_t v_t^\top \mid \mathscr{F}_t\right] K_t^\top = \left(I - K_t F_t^\top\right)\mathbb{E}\left[\varepsilon_t \mid \mathscr{F}_t\right]\mathbb{E}[v_t]^\top K_t^\top = 0,$$

and likewise for (iii). Only the terms (i) and (iv) remain:

$$P_t = \left(I - K_t F_t^\top\right)\mathbb{E}\left[\varepsilon_t \varepsilon_t^\top \mid \mathscr{F}_t\right]\left(I - K_t F_t^\top\right)^\top + K_t \mathbb{E}\left[v_t v_t^\top\right] K_t^\top.$$

Covariance of the state error. The dynamics give $\varepsilon_t = (c + GX_{t-1} + \omega_t) - (c + Ga_{t-1}) = G\varepsilon_{t-1} + \omega_t$, hence

$$\varepsilon_t \varepsilon_t^\top = G\varepsilon_{t-1}\varepsilon_{t-1}^\top G^\top + G\varepsilon_{t-1}\omega_t^\top + \omega_t \varepsilon_{t-1}^\top G^\top + \omega_t \omega_t^\top.$$

Since ω_t is centered and independent of ε_{t-1} and of $\mathscr{F}_{t-1}$, the two cross terms disappear in expectation, and it follows that

$$\mathbb{E}\left[\varepsilon_t \varepsilon_t^\top \mid \mathscr{F}_{t-1}\right] = G\,\mathbb{E}\left[\varepsilon_{t-1}\varepsilon_{t-1}^\top \mid \mathscr{F}_{t-1}\right] G^\top + W = G P_{t-1} G^\top + W = P_{t|t-1},$$

consistent with the prediction step. With, moreover, $\mathbb{E}[v_t v_t^\top] = V$, we obtain the symmetric form of P_t, which we expand:

$$
\begin{aligned}
P_t &= \left(I - K_t F_t^\top\right) P_{t|t-1} \left(I - K_t F_t^\top\right)^\top + K_t V K_t^\top\\
&= \left(P_{t|t-1} - K_t F_t^\top P_{t|t-1}\right)\left(I - F_t K_t^\top\right) + K_t V K_t^\top\\
&= P_{t|t-1} - P_{t|t-1} F_t K_t^\top - K_t F_t^\top P_{t|t-1}\\
&\quad + K_t F_t^\top P_{t|t-1} F_t K_t^\top + K_t V K_t^\top\\
&= P_{t|t-1} - K_t F_t^\top P_{t|t-1} - P_{t|t-1} F_t K_t^\top + K_t\left(F_t^\top P_{t|t-1} F_t + V\right) K_t^\top.
\end{aligned}
$$

Recognizing $L_{t|t-1} = F_t^\top P_{t|t-1} F_t + V$ (equation (B.3)) in the last term,

$$P_t = P_{t|t-1} - K_t F_t^\top P_{t|t-1} - P_{t|t-1} F_t K_t^\top + K_t L_{t|t-1} K_t^\top. \tag{B.5}$$

Minimization. We differentiate $\operatorname{tr}(P_t)$ with respect to K_t term by term, using the rules $\partial_K \operatorname{tr}(KA) = A^\top$, $\partial_K \operatorname{tr}(AK^\top) = A$ and $\partial_K \operatorname{tr}(KAK^\top) = 2KA$ (A symmetric):

$$
\begin{aligned}
\frac{\partial}{\partial K_t} \operatorname{tr}\left(-K_t F_t^\top P_{t|t-1}\right) &= -P_{t|t-1} F_t,\\
\frac{\partial}{\partial K_t} \operatorname{tr}\left(-P_{t|t-1} F_t K_t^\top\right) &= -P_{t|t-1} F_t,\\
\frac{\partial}{\partial K_t} \operatorname{tr}\left(K_t L_{t|t-1} K_t^\top\right) &= 2 K_t L_{t|t-1},
\end{aligned}
$$

the first using $\partial_K \operatorname{tr}(KA) = A^\top$ with $A = F_t^\top P_{t|t-1}$, whose transpose $P_{t|t-1}F_t$ exploits the symmetry of $P_{t|t-1}$. Summing, the first-order condition is written

$$\frac{\partial \operatorname{tr}(P_t)}{\partial K_t} = -2\,P_{t|t-1}F_t + 2\,K_t\,L_{t|t-1} = 0 \iff K_t\,L_{t|t-1} = P_{t|t-1}\,F_t \iff \boxed{K_t = P_{t|t-1}\,F_t\,L_{t|t-1}^{-1}} \tag{B.6}$$

Reinjection. The relation $K_t\,L_{t|t-1} = P_{t|t-1}\,F_t$ entails $K_t\,L_{t|t-1}\,K_t^\top = P_{t|t-1}\,F_t\,K_t^\top$, which exactly cancels the third term $-P_{t|t-1}F_tK_t^\top$ of equation (B.5):

$$P_t = P_{t|t-1} - K_tF_t^\top P_{t|t-1} - P_{t|t-1}F_tK_t^\top + P_{t|t-1}F_tK_t^\top = P_{t|t-1} - K_tF_t^\top P_{t|t-1},$$

that is, the update in its usual form:

$$\boxed{a_t = a_{t|t-1} + K_t\,e_t, \qquad P_t = \left(I - K_tF_t^\top\right)P_{t|t-1}} \tag{B.7}$$

The gain K_t tunes the correction: if the prediction uncertainty $P_{t|t-1}$ dominates the measurement noise V, the gain is strong and the observation prevails, and conversely.

Appendix C Non-identifiability with a short futures curve

This appendix formalizes the argument of section 4.3. We recall the definition: a model parameterized by Θ is *identifiable* if $p(y;\Theta_1) = p(y;\Theta_2)$ implies $\Theta_1 = \Theta_2$, in other words if the map $\Theta \mapsto p(y;\Theta)$ is injective. If two distinct parameter sets produce exactly the same distribution of the observations, no amount of data can tell them apart and the maximum likelihood estimator is not consistent: it does not converge to any point, the likelihood exhibiting a flat direction.

C.1 Model degeneracy as $\gamma \to 0$

We start from the measurement equation equation (10). The observation matrix stacks, for each maturity τ_i, the weights of the two factors:

$$F_t = \begin{pmatrix} e^{-\zeta\tau_1} & \cdots & e^{-\zeta\tau_M} \\ e^{-\gamma\tau_1} & \cdots & e^{-\gamma\tau_M} \end{pmatrix} \xrightarrow[\gamma\to 0]{} \begin{pmatrix} e^{-\zeta\tau_1} & \cdots & e^{-\zeta\tau_M} \\ 1 & \cdots & 1 \end{pmatrix}:$$

the long-term factor row becomes constant, all maturities loading on it identically. In the same regime, the component of $A(\tau)$ carried by the drift becomes linear in maturity, $\frac{\mu_\xi-\lambda_\xi}{\gamma}\left(1-e^{-\gamma\tau}\right) \to (\mu_\xi - \lambda_\xi)\,\tau$, and each observation reads

$$y_{t,i} = \left(\mu_\xi - \lambda_\xi + \tfrac{1}{2}\sigma_\xi^2\right)\tau_i + g(\tau_i) + e^{-\zeta\tau_i}\,\chi_t + \xi_t + v_{t,i}, \tag{C.1}$$

where $g(\tau_i)$ collects the terms in ζ, σ_χ, σ_ξ, ρ and λ_χ, which depend neither on the drift nor on the level of the long-term factor.

Consider first a curve reduced to a single maturity τ. For any $\delta \neq 0$, the translated parameter set

$$\Theta_2: \quad \mu_\xi \mapsto \mu_\xi + \delta, \qquad \xi_t \mapsto \xi_t - \delta\,\tau,$$

leaves equation (C.1) unchanged, term by term: the extra slope $\delta\tau$ is exactly absorbed by the drop in level. Since the noise $v_{t,i}$ is the same in both formulations, $p(y;\Theta_1) = p(y;\Theta_2)$ for every trajectory: there exists a continuous family of pairs (μ_ξ, ξ_t) producing exactly the same observations. The model is not identifiable and the maximum likelihood does not converge in probability to an optimal parameter.

With several maturities, the translation above depends on τ_i and no longer defines a common state: the exact identity breaks, and it is precisely the spread of maturities that must identify the factor. But if the observed range $[\tau_1, \tau_M]$ is narrow relative to $1/\gamma$, the weights $e^{-\gamma\tau_i}$ are nearly constant from one maturity to the next and the direction $(\mu_\xi + \delta,\ \xi_t - \delta\bar{\tau})$ modifies the likelihood only at second order, by an amount proportional to the dispersion of the τ_i around their mean $\bar{\tau}$. The likelihood then exhibits a nearly flat valley: theoretical identifiability in the strict sense, but inference impossible in practice, the estimates drifting freely along the valley at the mercy of the measurement noise.

C.2 Orders of magnitude: the condition on maturities

Identifying the two factors requires that their weights vary detectably along the observed curve, which imposes two conditions on the maximum maturity $\tau_{\max}$:

$$\text{(i)}\ e^{-\gamma\tau_{\max}} \ll 1 \quad \Longleftrightarrow \quad \tau_{\max} \gg \frac{1}{\gamma}, \qquad\qquad \text{(ii)}\ e^{-\zeta\tau_{\max}} \approx 0 \quad \Longleftrightarrow \quad \tau_{\max} \gg \frac{1}{\zeta}.$$

Condition (ii) guarantees that the long contracts are informative only about ξ_t, condition (i) that the decay of the weight of ξ_t be visible, allowing level and drift to be separated. The short contracts, by difference with the long contracts, are then informative about χ_t.

With the orders of magnitude of natural gas, $\gamma \approx 0.1$ (long-term factor half-life of several years) and $\zeta \approx 1.3$ (shocks absorbed within a few months), and our Databento chain with maximum maturity $\tau_{\max} = 1$ year:

$$e^{-\gamma\tau_{\max}} = e^{-0.1} = 0.905, \qquad e^{-\zeta\tau_{\max}} = e^{-1.3} = 0.272.$$

For all our maturities, $e^{-\gamma\tau_i} \in [0.905,\ 1]$: the total variation of the long-term factor weight along the curve is below ten percent, indistinguishable from the measurement noise. Condition (ii) is approximately satisfied ($1/\zeta \approx 0.8$ year), while condition (i) would require $\tau_{\max} \gg 1/\gamma \approx 10$ years. This is a data problem, not a method problem: Schwartz and Smith (2000) calibrate their model on futures with maturities of up to nine years, whereas our chain stops at one year.

Figure C.1 shows the optimizer trace for the best starting point: the log-likelihood stabilizes while ζ and γ slide toward their respective lower bounds, without ever finding an interior minimum, the numerical manifestation of the flat valley described above.

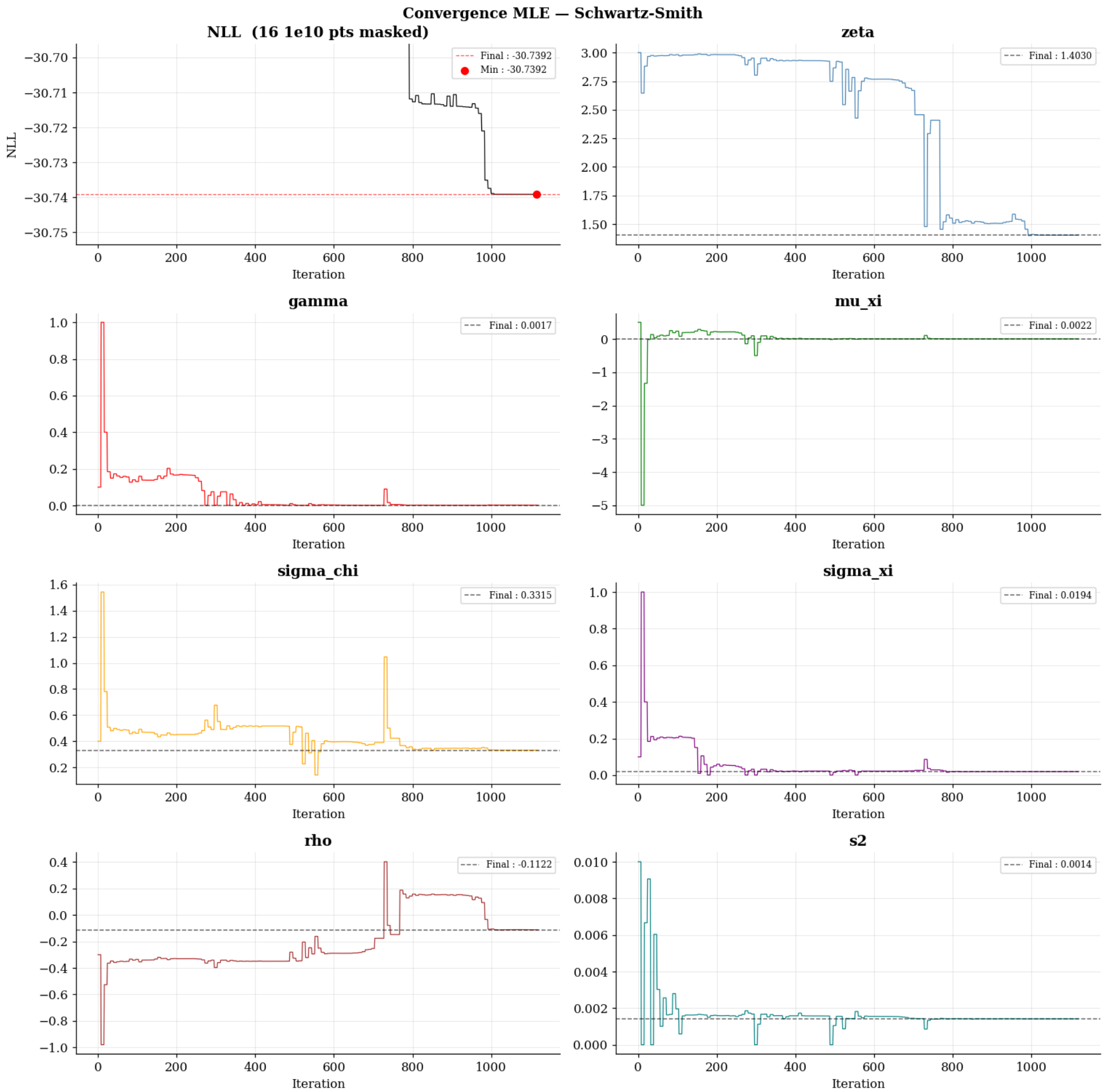

Figure C.1: Trace of the MLE optimization of the Schwartz–Smith model (best initialization): log-likelihood and parameter trajectories over the iterations.

Appendix D Supplementary Graphs on the Futures Contracts

This appendix gathers the notebook views on the Henry Hub futures chain that complement the analysis in section 3.3.

figure D.1 overlays the spot and the twelve contracts over the last thousand days, grouped by maturity bucket. The Samuelson effect is directly visible: the short-term contracts (C_1–C_4) track the spot and follow its violent spikes, whereas the long-term contracts (C_9–C_{12}) are markedly smoother.

Figure D.1: Henry Hub spot and log-prices of the twelve futures contracts $C_1, \ldots, C_{12}$ over the last thousand days, by maturity bucket (short, medium, long term).

Figures D.2 and D.3 show the log-basis $\log C_i - \log S$ over time, which distinguishes the *contango* regime (positive basis, $F_t > S_t$) from the *backwardation* regime (negative basis, $F_t < S_t$). The front-month C_1 switches very frequently between the two (54 % contango, 46 % backwardation), with strong occasional excursions. The intermediate maturity C_8, by contrast, stays in contango most of the time (77 %), the cost of carry dominating at the far maturities.

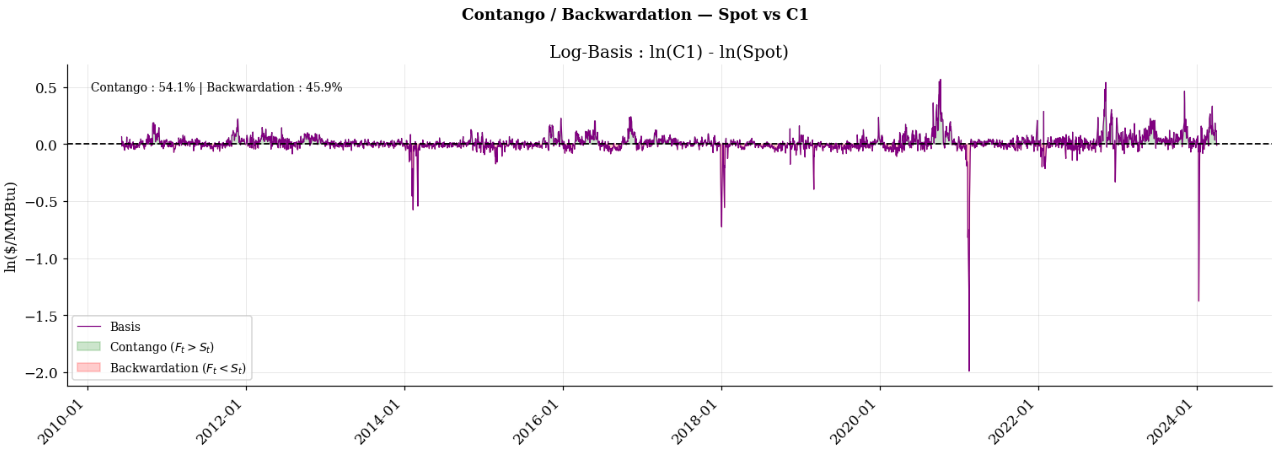

Figure D.2: Log-basis $\log C_1 - \log S$ (front-month): contango and backwardation alternate frequently at the shortest maturity.

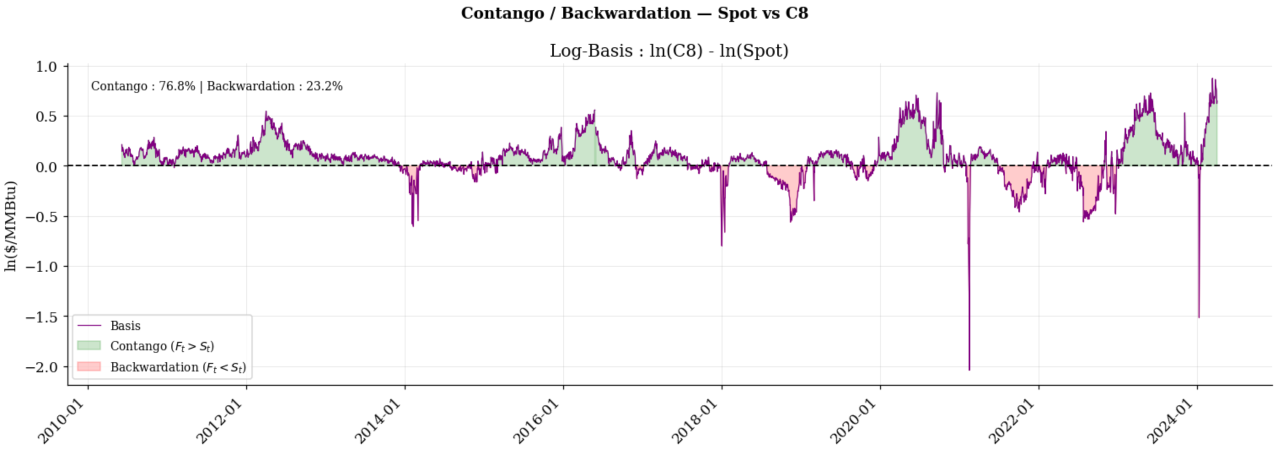

Figure D.3: Log-basis $\log C_8 - \log S$ (intermediate maturity): contango dominates, the carry structure taking precedence over transient shocks.