# A Regime-Aware Equity Timing Strategy Derived from the Marketron Model

Implementing, Calibrating, and Extending Halperin & Itkin (2025) for Model-Based Equity Risk Management

TFA Quant Research Track, EPFL, Spring 2026

Pablo Habib David Attias

April 2026

**Abstract**

We implement the Marketron model of Halperin & Itkin (2025), a physics-inspired framework for stock-price dynamics with multiple stable regimes, and use it to construct a regime-aware timing signal for the S&P 500. The signal is derived from the model-implied instantaneous drift, estimated through a particle filter on monthly market data, and translated into simple long/flat and signal-sized allocation rules.

Using a clean in-sample/out-of-sample protocol, we calibrate the model on data from January 2000 to December 2018 and evaluate the resulting strategies on data from January 2019 to October 2024. The filtered model remains stable out of sample, with nearly identical prediction errors in and out of sample, as reflected in an RMSE ratio of 1.010. The main practical benefit of the drift-based long/flat strategy is improved downside protection: it reduces the out-of-sample maximum drawdown from −24.8% under buy-and-hold to −20.1%, while also achieving a marginally higher out-of-sample Sharpe ratio (0.864 vs. 0.831). However, the in-sample Sharpe ratio of the strategy remains below buy-and-hold (0.084 vs. 0.212), and the out-of-sample improvement rests on only 70 months of data during a strong bull-market period, so it should not be interpreted as statistically conclusive evidence of predictive superiority.

On the modeling side, the calibrated Marketron system reproduces the main in-sample return moments closely across horizons from 2 to 18 years, including horizon-dependent mean returns, negative skewness at medium and long horizons, and non-Gaussian tail behavior. On our in-sample moment-matching criterion, Marketron substantially outperforms a geometric Brownian motion benchmark, with a total moment MSE of 0.032 against 8.644 for GBM. Overall, our results suggest that, in this setting, the main value of the Marketron framework lies in combining nonlinear regime-sensitive interpretation, strong in-sample fit to higher-order return moments, and a potentially useful defensive overlay for equity risk management.

## Contents

# 1 Introduction

## 1.1 Motivation

A standard starting point for equity price modeling is geometric Brownian motion (GBM):

$$dS_t = \mu\, S_t\, dt + \sigma\, S_t\, dW_t, \tag{1}$$

where $S_t$ denotes the stock price, $\mu$ its average growth rate, and $\sigma$ its volatility. The term $dW_t$ represents an increment of Brownian motion, i.e. a source of continuous random fluctuations. GBM underlies the classical Black–Scholes framework and remains a central benchmark in quantitative finance.

However, GBM implies that log-returns are Gaussian, and therefore symmetric and light-tailed. In such a framework, large market moves, and especially large downside moves, should be exceedingly rare. Real equity markets do not behave this way. The S&P 500 displays skewness, excess kurtosis, and recurrent periods of market stress that are difficult to reconcile with a single-regime Gaussian diffusion.

The motivation of the Marketron model goes beyond this classical limitation [Halperin and Itkin, 2025]. Halperin and Itkin build their framework around the interaction between investor flows, price impact, and market inelasticity, with an additional memory variable that stores information about past flows. This feedback mechanism generates a nonlinear stochastic system whose effective potential may exhibit multiple extrema, corresponding to metastable market configurations. In this picture, market dynamics can be interpreted as those of a particle evolving in an energy landscape: local minima represent persistent market regimes, while transitions between them occur through barrier-crossing events induced by stochastic fluctuations. This structure provides a natural mechanism for regime changes, asymmetric return distributions, and heavy tails.

## 1.2 Our Contribution

Our project has two main objectives. First, we implement the Marketron model of Halperin and Itkin [2025] and calibrate it on S&P 500 data under a clean train/test protocol: parameters are estimated on an in-sample period running from January 2000 to December 2018, and the resulting strategy is evaluated on the genuinely unseen period from January 2019 to October 2024 (Sections 2–2.10). Second, we use the filtered hidden state of the model to construct a drift-based timing signal (Section 3) and assess whether this latent-state information can be used as a practical equity-risk-management overlay out of sample (Section 4).

More precisely, at each month the particle filter produces estimates of the model's hidden variables. We then plug these filtered states into the drift term of the log-price equation to obtain a model-implied directional signal, namely the model's current estimate of whether the local drift of the log-price is positive or negative. We translate this signal into simple allocation rules, in particular a long/flat strategy that is invested in the S&P 500 when the filtered drift is positive and moves to cash otherwise.

This trading rule is not an external indicator added independently of the model. Rather, it is a project-specific extension built directly from the model-implied drift, which allows us to test whether the filtered Marketron state contains economically useful timing information. Our results suggest that this information is more convincing as a defensive overlay than as a source of robust predictive superiority: the long/flat strategy reduces drawdowns out of sample and delivers a marginally higher out-of-sample Sharpe ratio than buy-and-hold, but its in-sample Sharpe ratio is lower, and the out-of-sample improvement is based on only 70 months of data. On the statistical side, the calibrated Marketron model reproduces the main in-sample return moments closely across horizons and substantially outperforms a simple GBM benchmark on our moment-matching criterion.

## 2 The Marketron Model

The original paper derives the model progressively, starting from a stock-price equation in $S_t$, together with specific choices for the latent signal, the investor policy, and the price-impact function. In the present project, however, our implementation, filtering, and calibration all operate directly on the final log-price formulation of the model. For this reason, we do not reproduce the full derivation from $S_t$ here, and instead take as our starting point the coupled three-dimensional system in $(x_t, y_t, \theta_t)$ given in Eq. (29) of Halperin and Itkin [2025]. This is the formulation that is directly relevant for the numerical methods and empirical analysis developed in the sequel.

### 2.1 Core Idea: The Market as a Physical System

The Marketron model represents the stock market as a ball rolling on a hilly landscape, subject to random shocks:

- valleys represent stable market regimes in which the price tends to remain;
- hills represent unstable tipping points between these regimes;
- random shocks, modeled by Brownian motion, can push the ball over a hill and into another valley; this is how the model accounts for regime transitions and episodes of market stress.

The shape of this landscape is determined by three interacting state variables that evolve over time, which we describe next.

### 2.2 State Variables

The model tracks three state variables over time:

1. $x_t = \log(S_t/S^*)$: the log-price of the S&P 500 relative to a reference level $S^* = 1000$. This is the only directly observed variable. When $x_t > 0$, the market is above the reference level; when $x_t < 0$, it is below it.
2. $y_t$: the memory variable. It represents the accumulated effect of past investor activity, that is, a form of inertia generated by recent money flows. Higher values of $y_t$ correspond to stronger recent buying pressure, while lower values reflect weaker demand or net selling. This variable is latent and must be inferred from the observed price series.
3. $\theta_t$: the signal variable. This is a latent, slowly varying component that affects the dynamics of the system. It follows an Ornstein–Uhlenbeck process, meaning that it evolves randomly while being continuously pulled back toward a long-run level $\hat{\theta}$.

### 2.3 Equations of Motion

The three state variables evolve according to a system of coupled stochastic differential equations (SDEs). In general, an SDE takes the form

$$dX_t = \underbrace{a(X_t)\,dt}_{\text{predictable drift}} + \underbrace{b(X_t)\,dW_t}_{\text{random shock}},$$

where the drift term $a(X_t)\,dt$ represents the systematic and predictable part of the dynamics, while $b(X_t)\,dW_t$ represents a random shock whose magnitude is controlled by the volatility coefficient $b$.

The full Marketron system (Eq. 29, p. 19 of [Halperin and Itkin, 2025]) is given by

$$dx_t = \Big[\ \underbrace{v\, f(\theta_t, t)}_{\substack{\text{signal influence}\\ \text{on price}}} \ + \ \underbrace{\eta}_{\substack{\text{structural}\\ \text{drift}}} \ - \ \underbrace{c\, y_t\, V_M'(x_t)}_{\substack{\text{nonlinear feedback}\\ \text{from investor flows}}}\ \Big]\, dt + \sigma\, dW_t, \tag{2}$$

$$dy_t = \Big[\ \underbrace{v\, h(\theta_t, t)}_{\substack{\text{signal influence}\\ \text{on memory}}} \ + \ \underbrace{\mu(\bar{y} - y_t)}_{\substack{\text{memory}\\ \text{mean-reversion}}} \ - \ \underbrace{c\, V_M(x_t)}_{\substack{\text{price feedback}\\ \text{into memory}}}\ \Big]\, dt + \sigma_y\, d\tilde{W}_t, \tag{3}$$

$$d\theta_t = \underbrace{k(\hat{\theta} - \theta_t)}_{\substack{\text{signal}\\ \text{mean-reversion}}} \ dt + \sigma_z\, dZ_t. \tag{4}$$

Here $W_t$, $\tilde{W}_t$, and $Z_t$ are three independent Brownian motions, while $\sigma$, $\sigma_y$, and $\sigma_z$ govern the intensity of randomness in each equation. The functions $f(\theta_t, t)$ and $h(\theta_t, t)$ describe how the latent signal affects price and memory dynamics, whereas $V_M(x)$ is the nonlinear potential governing the price-feedback structure of the model. The parameter $k$ controls the speed at which $\theta_t$ reverts to its long-run level $\hat{\theta}$. The explicit forms of these objects are introduced below.

These three equations play distinct but interconnected roles. Equation (2) governs the evolution of the observed log-price $x_t$: it combines a latent signal component, a structural drift, a nonlinear feedback term driven by investor memory, and random shocks. Equation (3) describes the hidden memory variable $y_t$, which is influenced both by the latent signal and by the current price level, while also reverting toward a long-run level $\bar{y}$. Finally, Eq. (4) specifies the latent signal $\theta_t$ as a mean-reverting Ornstein–Uhlenbeck process. Together, these equations couple price, memory, and latent signal into a single nonlinear stochastic system.

We now describe the main building blocks of this system.

**The Morse-like potential $V_M(x)$.** This function, defined in Eq. (16) of Halperin and Itkin [2025], governs the nonlinear price-feedback structure of the model. It is called "Morse-like" because its shape is related to the Morse potential used in physics, although the expression adopted here is adapted to the financial setting:

$$V_M(x) = \frac{1}{\varepsilon}\left[(\varepsilon - 1)(e^{-x} - 1) + \frac{1}{\kappa}\log\frac{1 + \kappa e^{-x}}{1 + \kappa}\right], \qquad \kappa = g\varepsilon. \tag{5}$$

The parameter $g \in [0, 1]$ is a coupling constant controlling how strongly investor behavior reacts to price deviations, while $\varepsilon = 0.02$ is a small regularization parameter. The derivative $V_M'(x)$ enters directly into the price equation through the nonlinear feedback term and therefore helps determine how the current price level interacts with the memory variable $y_t$.

For qualitative analysis, the paper also introduces the inverted-Morse approximation

$$\bar{V}_M(x) = \frac{g}{2} - 1 + e^{-x} - \frac{g}{2}e^{-2x}. \tag{6}$$

This approximation is analytically more tractable and is useful for studying the geometry of the potential and the existence of multiple regimes. By contrast, the exact expression in Eq. (5) is used in the full state dynamics, in the particle filter, and in calibration, where fidelity to the original specification is more important than analytical simplicity.

Figure 1 compares the exact Morse-like potential with its inverted-Morse approximation and also displays the corresponding derivative. It illustrates the region in which the approximation remains close to the exact specification, as well as the range in which the two expressions begin to diverge.

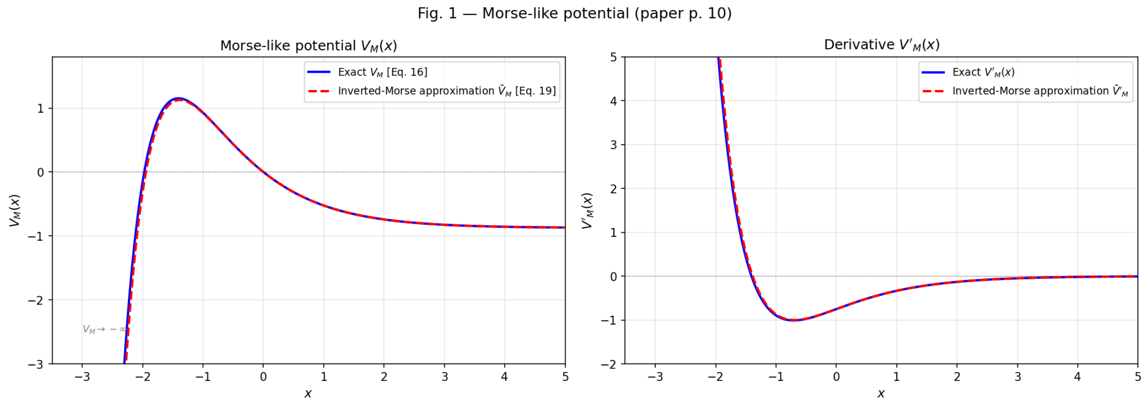



Figure 1: The exact Morse-like potential $V_M(x)$, its derivative $V'_M(x)$, and the corresponding inverted-Morse approximation used for qualitative analysis.

**The signal functions $f(\theta, t)$ and $h(\theta, t)$.** These functions determine how the hidden signal $\theta_t$ enters the price and memory dynamics (Eq. 35 of Halperin and Itkin [2025]):

$$f(\theta, t) = \frac{a_1(t)}{1 + e^{-b_1 \theta}}, \quad a_1(t) = k_{1,x} \cos(k_{2,x} + k_{3,x} t), \tag{7}$$

$$h(\theta, t) = \frac{a_2(t)}{1 + e^{-b_2 \theta}}, \quad a_2(t) = k_{1,y} \sin(k_{2,y} + k_{3,y} t). \tag{8}$$

Each signal function combines a sigmoid component, $\frac{1}{1+e^{-b\theta}}$, with a time-varying amplitude. The sigmoid acts as a smooth switching mechanism, modulating how strongly the latent signal $\theta_t$ affects the dynamics. The periodic amplitudes $a_1(t)$ and $a_2(t)$ introduce explicit time dependence into these effects and form part of the empirical specification adopted in the paper's calibration procedure.

**The mean-reversion terms.** In Eq. (3), the term $\mu(\bar{y} - y_t)$ pulls the memory variable back toward its equilibrium level $\bar{y}$ at speed $\mu$. In Eq. (4), the term $k(\hat{\theta} - \theta_t)$ pulls the latent signal back toward its long-run level $\hat{\theta}$ at speed $k$. In both cases, the restoring force increases with the distance from equilibrium.

## 2.4 The Energy Landscape and Three Market Regimes

The purpose of introducing a potential is not to replace the original equations, but to make their deterministic structure easier to interpret. In the stochastic formulation of the model, the price and memory variables evolve under the combined action of signal terms, random shocks, and a nonlinear feedback mechanism. If one temporarily ignores the signal contributions and the stochastic shocks, the remaining drift terms in the coupled $(x_t, y_t)$ dynamics can be written as minus the partial derivatives of a two-dimensional potential. This provides a geometric representation of the deterministic skeleton of the model and allows the dynamics to be interpreted in terms of valleys, barriers, and metastable states. Specifically, we introduce

$$V(x, y) = -\eta\, x + c\, y\, V_M(x) + \tfrac{1}{2}\mu\, (y - \bar{y})^2, \tag{9}$$

which corresponds to the exact deterministic formulation before the tractable approximation used for qualitative analysis in the paper. Indeed,

$$-\frac{\partial V}{\partial x} = \eta - c\, y\, V'_M(x), \qquad -\frac{\partial V}{\partial y} = \mu(\bar{y} - y) - c\, V_M(x),$$

which are exactly the structural drift terms appearing in Eqs. (2) and (3) once the signal terms and stochastic shocks are removed. The potential therefore does not introduce a new assumption; rather, it reorganizes the deterministic part of the model into a form that is easier to visualize and analyze.

Each term has a natural interpretation. The linear term $-\eta x$ tilts the landscape globally: if $\eta > 0$, the potential is biased downward toward larger values of $x$, whereas if $\eta < 0$, it is biased in the opposite direction. The coupling term $c\, y\, V_M(x)$ links the hidden memory state to the nonlinear price potential; in combination with the other terms, this coupling is what makes a multi-well structure possible. Finally, the quadratic term $\frac{1}{2}\mu(y-\bar{y})^2$ confines the memory variable near its equilibrium level $\bar{y}$, with strength controlled by $\mu$.

Although the two-dimensional potential is useful conceptually, the regime structure becomes even clearer in the fast-memory limit. When $\mu \gg 1$, the memory variable $y_t$ evolves on a much faster timescale than the price variable $x_t$. In that regime, $y_t$ is no longer treated as a fully independent slow state; instead, it is approximated by its local quasi-equilibrium value conditional on $x_t$. Starting from the memory equation (3), and ignoring for this qualitative reduction the signal and noise terms, the deterministic drift of $y_t$ is

$$\mu(\bar{y} - y) - c\, V_M(x).$$

Assuming that the memory variable rapidly relaxes to its local quasi-equilibrium, we set this drift approximately equal to zero, which gives

$$y \approx \bar{y} - \frac{c}{\mu} V_M(x).$$

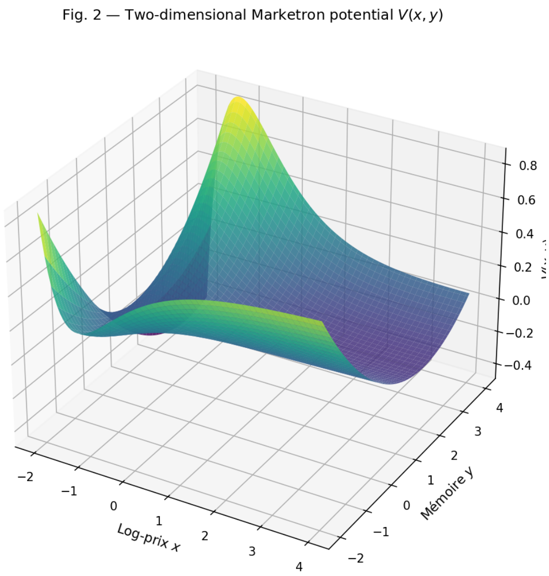
Substituting this expression into the two-dimensional potential $V(x, y)$ yields an effective one-dimensional potential,

$$U_{\text{eff}}(x) = -\eta\, x + c\bar{y}\, V_M(x) - \frac{c^2}{2\mu}\, V_M(x)^2, \tag{10}$$

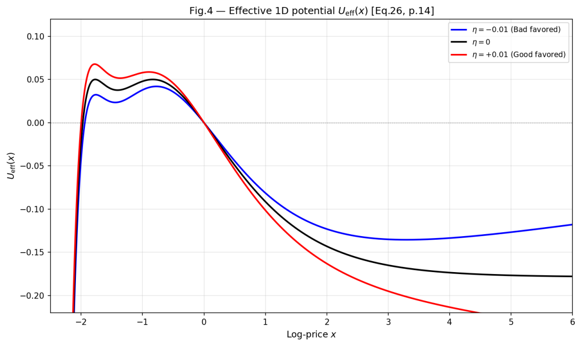
which captures the reduced deterministic dynamics of the model in the fast-memory limit.

This reduction is useful because it makes the regime structure much easier to read. In a potential picture, a local minimum corresponds to a metastable state: if the system is slightly perturbed away from it, the deterministic force tends to pull it back. By contrast, a local maximum is unstable: even a small perturbation pushes the system away from it. For suitable parameter values, the interplay between the global tilt $-\eta x$ and the nonlinear terms involving $V_M(x)$ and $V_M(x)^2$ produces a double-well structure, with two local minima separated by a local maximum. Following the interpretation proposed in the paper, these extrema correspond to the Good and Bad market regimes, separated by an intermediate Ugly tipping point. The key idea is that the model can generate more than one metastable state, together with a barrier separating them.

These geometric features are illustrated in Figure 2. The left panel shows the full two-dimensional landscape in the $(x, y)$ variables, while the right panel shows the reduced one-dimensional effective potential in the fast-memory limit. Both figures are computed using qualitative parameter values taken from the paper, and are intended to illustrate the regime mechanism rather than the empirically calibrated landscape of our final specification.




(a) The full 2D potential $V(x, y)$ viewed as a surface. The valleys and the ridge separating them are clearly visible. The $x$-axis is the log-price, the $y$-axis is the memory variable, and height represents potential energy.




(b) The effective 1D potential $U_{\text{eff}}(x)$ for different values of $\eta$. For suitable values of $\eta$, the landscape displays a double-well structure, with two local minima separated by a local maximum. Changing $\eta$ tilts the landscape and changes the relative depth of the metastable states.

Figure 2: The Marketron energy landscape, computed with qualitative parameter values from the paper ($c = 0.13$, $g = 0.25$, $\mu = 0.1$, $\bar{y} = 1$). A useful intuition is that of a ball sitting in one of the valleys: under small random shocks it remains near a local minimum, but sufficiently large fluctuations may push it across the barrier into another metastable region.

A complementary perspective is obtained by fixing the memory variable and examining how the potential changes along the price direction alone. Figure 3 shows that different values of $y$ can substantially reshape the landscape seen by $x$. This helps explain why the hidden memory state matters for the stability structure of the model, and why estimating $y_t$ is important in the filtered version of the dynamics.

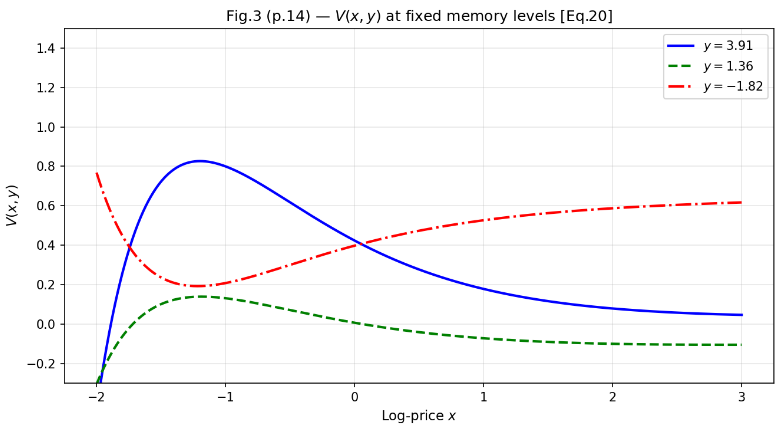



Figure 3: The two-dimensional potential $V(x, y)$ sliced at three fixed values of the memory variable $y$. Each curve shows how the potential landscape in the price direction changes with the hidden memory state of the market. When $y$ varies, the local shape of the landscape in $x$ changes substantially, illustrating how investor memory can alter the effective forces acting on the price.

### 2.5 Model Parameters

Although the model contains 18 parameters, they play distinct conceptual roles and can be grouped into three broad families:

- **Volatility parameters:** $\sigma$ (price noise), $\sigma_y$ (memory noise), and $\sigma_z$ (signal noise).
- **Structural parameters:** $\eta$ (global drift or tilt), $g$ (investor coupling in the Morse-like potential), $c$ (strength of the price–memory coupling), $\mu$ (memory mean-reversion speed), and $\bar{y}$ (equilibrium memory level).
- **Signal parameters:** $k$ (signal mean-reversion speed), $\hat{\theta}$ (long-run signal level), $b_1, b_2$ (sigmoid steepness), and $k_{1,x}, k_{2,x}, k_{3,x}, k_{1,y}, k_{2,y}, k_{3,y}$, which determine the amplitude, phase, and frequency of the time-varying signal components in the price and memory equations.

The regularization parameter $\varepsilon = 0.02$ is fixed throughout. In our implementation, the parameter bounds are taken from the empirical specification reported in Table 1 of Halperin and Itkin [2025], while the final calibrated values are obtained from our own in-sample calibration procedure described below.

### 2.6 Data

We use monthly closing levels of the S&P 500 index from January 2000 to October 2024, downloaded from Yahoo Finance through the `yfinance` Python library. The dataset contains 298 monthly observations. As in Halperin and Itkin [2025], we work with price levels only, excluding reinvested dividends. We transform the series into log-prices according to

$$x_t = \log(S_t/S^*),$$

where the reference level is fixed at $S^* = 1000$. The resulting series is treated as a monthly time series with time step $\Delta t = 1/12$ year.

For empirical evaluation, we divide the sample into an in-sample period from January 2000 to December 2018 (228 monthly observations) and an out-of-sample period from January 2019 to October 2024 (70 monthly observations). In our main specification, the in-sample subsample is used for calibration, while the out-of-sample subsample is reserved for the evaluation of the trading strategy.

We stop the sample in October 2024 in order to remain aligned with the empirical window considered in Halperin and Itkin [2025]. Extending the dataset to more recent observations would be a natural continuation of the project, but it would also move the analysis away from the paper's original empirical setting and make direct comparisons less transparent.

### 2.7 Numerical Discretization: The Seidel Scheme

To simulate the continuous-time SDE system in discrete time, we need a numerical approximation scheme. Following Halperin and Itkin [2025], we use the Seidel scheme (Eqs. 31–33), which updates the coupled variables sequentially rather than simultaneously. This choice is natural here because the dynamics of $x_t$, $y_t$, and $\theta_t$ are strongly intertwined: later updates can therefore use the most recent values computed earlier within the same time step.

More precisely, at each time step:

1. we first compute $\Delta\theta_n$ from the current state;
2. we then use this updated signal increment to compute the new memory value;
3. finally, we update $x$ using the freshly updated memory variable.

In this way, the discretization respects the directional coupling structure of the model more faithfully than a naive simultaneous Euler update. Figure 4 shows a qualitative simulation of the full system under representative parameter values, before calibration to data.

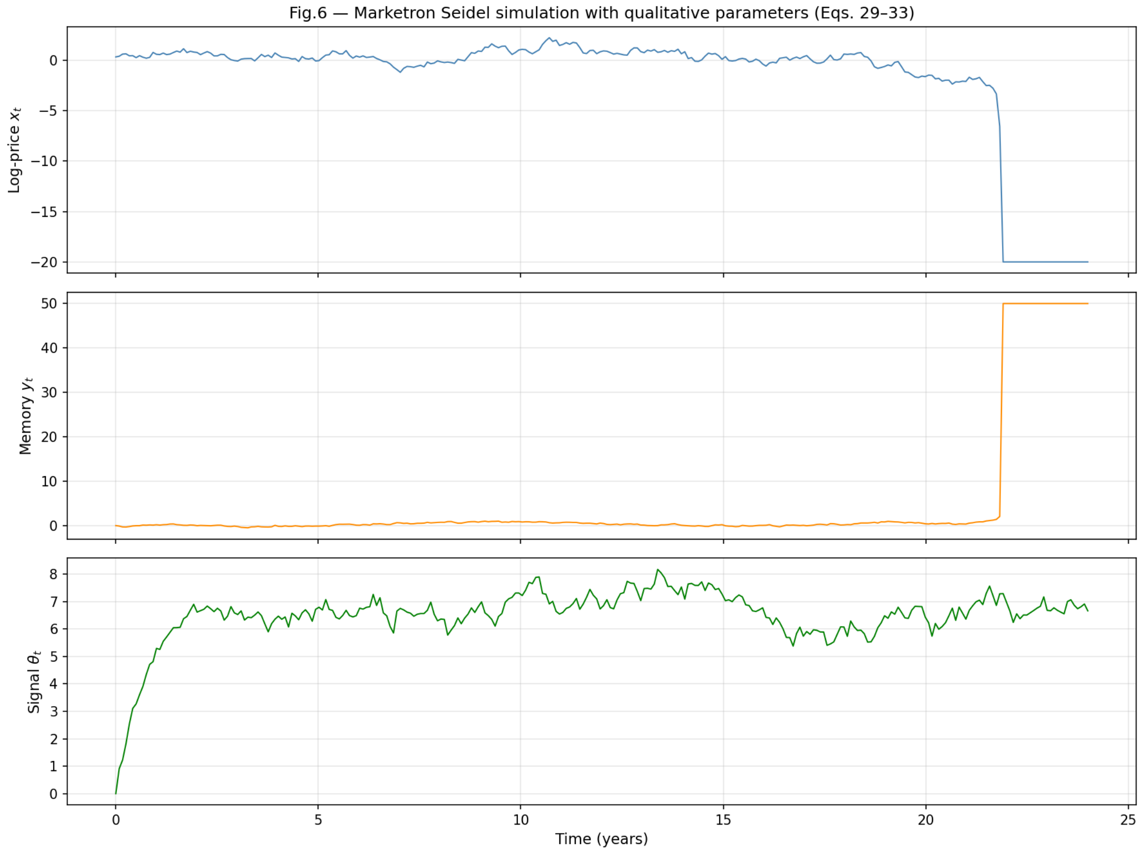



Figure 4: Qualitative simulation of the full Marketron system over 24 years under representative parameter values. The three panels show the joint evolution of the log-price $x_t$ (top), the memory variable $y_t$ (middle), and the latent signal $\theta_t$ (bottom). This simulation is not meant as a calibrated fit to the S&P 500, but as a qualitative check that the Seidel discretization produces coherent coupled trajectories before running the full filtering and calibration procedures. In particular, it illustrates the interaction between price, memory, and latent signal that distinguishes the Marketron dynamics from a single-variable benchmark such as geometric Brownian motion.

## 2.8 The Particle Filter: Estimating Hidden Variables

Only $x_t$ is directly observed, while $y_t$ and $\theta_t$ are latent. Since the model's drift, and therefore our timing signal, depends on these hidden variables, they must be estimated at each date from the observed price series. To do so, we use a bootstrap particle filter [Gordon et al., 1993], following the empirical setup of Section 6.1 in Halperin and Itkin [2025]. The filter maintains a cloud of 1500 particles, each representing one possible latent market state $(x, y, \theta)$. At each monthly time step, it proceeds as follows:

1. **Propagate:** all 1500 particles are advanced by one month using the Seidel discretization, each with its own random shocks. This generates a cloud of possible next-period latent states.

2. **Weight:** each particle is compared with the observed S&P 500 log-price. Particles whose predicted price is closer to the observed value receive higher weights, according to the

Gaussian likelihood

$$w_i \propto \exp\left(-\frac{(x_{\text{pred}}^{(i)} - x_{\text{obs}})^2}{2 \times 0.05}\right). \tag{11}$$

The quantity 0.05 is the observation-noise variance used in the paper. It determines how selective the filter is: the smaller this variance, the more strongly particles are penalized for deviating from the observed price.

3. **Resample:** a new cloud of 1500 particles is drawn from the current one with probability proportional to the particle weights. Particles that better match the data are replicated, while poorly fitting particles are discarded. This step keeps the particle cloud concentrated on latent states that remain consistent with the observations.

4. **Estimate:** the filtered hidden state is estimated from the resampled particle cloud. In particular, the filtered memory variable is approximated by the empirical average

$$\hat{y}_t = \frac{1}{1500} \sum_{i=1}^{1500} y_t^{(i)},$$

and similarly for $\hat{\theta}_t$.

At the initial date, the particle cloud is generated by sampling

$$x_0^{(i)} = x_{\text{obs},0} + \varepsilon_{x,i}, \qquad y_0^{(i)} = \varepsilon_{y,i}, \qquad \theta_0^{(i)} = \varepsilon_{\theta,i}, \qquad i = 1, \ldots, 1500,$$

where $\varepsilon_{x,i}, \varepsilon_{y,i}, \varepsilon_{\theta,i}$ are independent Gaussian perturbations distributed as $\mathcal{N}(0, 0.01^2)$. Thus, the initial particles are centered at the first observed log-price for $x$, and at 0 for the latent variables $y$ and $\theta$, with a small random dispersion.

The full filter is vectorized in NumPy, so that all 1500 particles are processed simultaneously through array operations. In our implementation, this makes repeated filtering computationally inexpensive, which is important because calibration requires many objective evaluations.

The use of 1500 particles, monthly time steps ($\Delta t = 1/12$), and observation variance 0.05 follows the empirical specification adopted in Halperin and Itkin [2025].

### 2.9 Calibration: Finding the Best Parameters

Calibration consists in choosing the 18 model parameters so that the model reproduces the main empirical properties of the S&P 500 as closely as possible, while preserving the qualitative regime structure of the Marketron framework. In our main specification, calibration is performed only on the in-sample period from January 2000 to December 2018. To do so, we minimize the objective function

$$\mathcal{L}(\boldsymbol{\theta}) = \underbrace{\sum_{i=1}^{5} \sum_{j=1}^{4} \left(m_{ij}^{\text{model}} - m_{ij}^{\text{market}}\right)^2}_{\text{moment-matching term}} + \underbrace{\lambda_1 P_{\text{shape}}(\boldsymbol{\theta})}_{\text{shape penalty}} + \underbrace{\lambda_2 P_{\text{bias}}(\boldsymbol{\theta})}_{\text{prediction-bias penalty}}. \tag{12}$$

The first term matches four return moments; mean, volatility, skewness, and excess kurtosis at the five in-sample horizons

$$H = 2, 5, 10, 15, 18 \text{ years}.$$

These targets are computed exclusively from the January 2000–December 2018 subsample, which removes look-ahead bias from the main specification.

The shape penalty $P_{\text{shape}}$ discourages parameter values for which the effective potential $U_{\text{eff}}(x)$ no longer exhibits the multi-extrema structure associated with the Good, Bad, and Ugly regimes.

As in the paper, this penalty is motivated by analytical conditions on the potential shape discussed in Section 6 and Appendix C of Halperin and Itkin [2025].

The prediction-bias penalty $P_{\text{bias}}$ discourages parameter configurations for which the filtered model systematically over- or under-predicts observed prices. In our implementation, this penalty is based on the mean prediction error across particles, so that the calibrated model does not display a persistent directional bias.

In our implementation, both penalty terms are weighted equally, with

$$\lambda_1 = \lambda_2 = 10.$$

The resulting optimization problem is high-dimensional, non-convex, noisy, and computationally expensive, since each objective evaluation requires running the particle filter. We therefore benchmarked several calibration pipelines rather than fixing the optimizer a priori. In particular, we compared DE + L-BFGS-B, CEopt + DE, and CMA-ES + L-BFGS-B under the same in-sample objective and particle-filter setup.

The final benchmark strongly favored **CMA-ES + L-BFGS-B**. It delivered the best overall compromise between objective quality and runtime among the tested methods. For DE + L-BFGS-B and CEopt + DE, the average objective values over three independent seeds were approximately 7.336 and 2.878, respectively; by comparison, the retained CMA-ES + L-BFGS-B run achieved 2.035. In terms of runtime, CMA-ES + L-BFGS-B required only about two minutes per run, compared with roughly 55 minutes for DE + L-BFGS-B and 109 minutes for CEopt + DE.

For these reasons, we retain **CMA-ES + L-BFGS-B** as the final calibration pipeline throughout the remainder of the paper.

### 2.10 Calibration Results

Using the retained CMA-ES + L-BFGS-B calibration pipeline, Table 1 compares the model-implied return moments with their empirical counterparts on the in-sample period only, that is, January 2000–December 2018.

| $H$ | **Mean** | | **Volatility** | | **Skewness** | | **Kurtosis** | |
|---|---|---|---|---|---|---|---|---|
| (yrs) | Mkt | Model | Mkt | Model | Mkt | Model | Mkt | Model |
| 2 | −0.1051 | −0.0962 | 0.1757 | 0.1715 | 0.1720 | 0.1962 | −0.6414 | −0.6264 |
| 5 | −0.0327 | −0.0292 | 0.1623 | 0.1643 | −0.2295 | −0.2694 | −0.2688 | −0.1586 |
| 10 | −0.0259 | −0.0279 | 0.1631 | 0.1617 | −0.7649 | −0.7024 | 1.2930 | 1.1990 |
| 15 | 0.0237 | 0.0244 | 0.1540 | 0.1502 | −0.7629 | −0.7438 | 1.4253 | 1.4331 |
| 18 | 0.0424 | 0.0380 | 0.1457 | 0.1434 | −0.7935 | −0.7681 | 1.6057 | 1.6657 |

Table 1: Annualized return moments on the in-sample period (January 2000–December 2018): S&P 500 data versus calibrated Marketron under the retained CMA-ES + L-BFGS-B pipeline.

Each row corresponds to one return horizon computed within the in-sample window. Mean denotes the average annualized return, volatility the annualized standard deviation, skewness the degree of asymmetry, and excess kurtosis the heaviness of the tails relative to a Gaussian benchmark.

Overall, the retained calibration reproduces the in-sample moments remarkably well. The fit is close not only for the mean and volatility, but also for skewness and excess kurtosis across all horizons considered. In particular, the model captures the sign change in mean returns across horizons, the strongly negative skewness observed from medium horizons onward, and the pronounced non-Gaussian tail behavior reflected in positive excess kurtosis at long horizons.

The main remaining discrepancy appears at the 2-year horizon, where the empirical skewness is slightly positive while the model also produces a small positive skewness, though with somewhat larger magnitude. Beyond that short horizon, however, the fit is consistently strong. At 5, 10, 15, and 18 years, the model-implied moments lie very close to their empirical targets, including the higher-order moments that a Gaussian benchmark cannot reproduce by construction.

This strong fit is also reflected in the global in-sample moment-matching criterion. The total in-sample moment MSE is 0.032 for Marketron, compared with 8.644 for a GBM benchmark. The large gap comes primarily from skewness and kurtosis, where GBM is structurally constrained to zero and therefore cannot match the observed non-Gaussian features of the data. In that sense, the main statistical advantage of the Marketron framework in our setting lies in its ability to reproduce higher-order distributional properties within a state-space model that remains economically interpretable.

Table 2 reports the calibrated parameter values produced by the retained CMA-ES + L-BFGS-B pipeline, together with the benchmark values reported in Halperin and Itkin [2025].

| **Param** | **Paper** | **Ours** | **Param** | **Paper** | **Ours** |
|---|---|---|---|---|---|
| $\sigma$ | 0.791 | 2.9781 | $c$ | 3.931 | 0.1440 |
| $\sigma_y$ | 0.380 | 1.1541 | $b_1$ | 1.682 | 4.4232 |
| $\sigma_z$ | 0.833 | 2.9972 | $b_2$ | $-1.210$ | 6.3317 |
| $\eta$ | $-1.569$ | $-0.2601$ | $k_{1,x}$ | $-3.200$ | 4.9866 |
| $k$ | 1.287 | 1.3953 | $k_{2,x}$ | 2.742 | 2.6569 |
| $\mu$ | 1.667 | 2.4511 | $k_{3,x}$ | $-1.883$ | 4.5372 |
| $g$ | 0.683 | 0.2721 | $k_{1,y}$ | $-0.786$ | 0.2032 |
| $\hat{\theta}$ | 6.787 | 7.2757 | $k_{2,y}$ | 3.890 | $-2.5348$ |
| $\bar{y}$ | 0.473 | 0.0503 | $k_{3,y}$ | 1.559 | $-1.7432$ |

Table 2: Calibrated parameters: benchmark values from Halperin and Itkin [2025] versus our retained CMA-ES + L-BFGS-B calibration pipeline.

Several parameters differ substantially from the benchmark values reported in the paper. This is not unexpected in a high-dimensional latent-variable model with strong nonlinear couplings, partial observability, and a noisy calibration objective. In such settings, different parameter combinations can generate similar qualitative behavior while differing materially in their decomposition of noise, latent-state persistence, and signal structure.

A few differences are especially notable. First, the three volatility parameters $\sigma$, $\sigma_y$, and $\sigma_z$ are all larger than in the paper calibration, especially for the latent coordinates $y$ and $\theta$. Second, the price–memory coupling parameter $c$ is much smaller than in the benchmark calibration, while the memory mean-reversion speed $\mu$ is larger. Third, several signal-related coefficients differ not only in magnitude but also in sign, which suggests that the precise empirical form of the latent signal component is not tightly identified in a unique structural way by the calibration target alone.

For this reason, the economic interpretation of individual parameter values should be treated with caution. In the context of this project, the more relevant benchmark is not the standalone meaning of each estimated coefficient, but rather the joint quality of the in-sample statistical fit, the stability of the filtered latent states, and the usefulness of the resulting drift signal for downstream trading applications.

# 3 Trading Strategy

Up to this point, our project has focused on implementing, understanding, and calibrating the Marketron framework on real S&P 500 data. The next step is the main empirical extension of the project: rather than using the model only as a descriptive or calibration device, we turn its filtered latent state into a concrete trading signal. In this sense, our contribution is not to modify the structure of the Marketron model itself, but to show how its hidden-state dynamics can be translated into a regime-aware allocation rule for equity exposure.

While Halperin and Itkin [2025] develop and calibrate the model, they do not explicitly construct a trading strategy from the filtered state variables. Our approach therefore goes one step further by using the model-implied drift of the log-price equation as a directional signal for the S&P 500.

### 3.1 The Key Idea: Using the Model's Drift as a Trading Signal

The starting point is the log-price dynamics of the model. Recall that the $x_t$ equation is

$$dx_t = \Big[v f(\theta_t, t) + \eta - c\, y_t\, V_M'(x_t)\Big] dt + \sigma\, dW_t. \tag{13}$$

This equation contains two conceptually distinct components. The first is the deterministic term,

$$v f(\theta_t, t) + \eta - c\, y_t\, V_M'(x_t),$$

which captures the local average tendency of the log-price. The second is the stochastic term,

$$\sigma\, dW_t,$$

which represents unpredictable market noise. Since a trading rule should be based on the model's systematic directional information rather than on the noise component, we focus on the deterministic part only.

This deterministic component naturally defines the model-implied local drift. Because the hidden variables $y_t$ and $\theta_t$ are not directly observed, we replace them by their filtered estimates $\hat{y}_t$ and $\hat{\theta}_t$, and similarly use the filtered log-price $\hat{x}_t$. This yields the signal

$$\mu_x(t) = v\, f(\hat{\theta}_t, t) + \eta - c\, \hat{y}_t\, V_M'(\hat{x}_t), \tag{14}$$

which we interpret as the model's current estimate of the instantaneous drift of the log-price.

This signal combines three sources of information:

1. $v\, f(\hat{\theta}_t, t)$, the latent-signal contribution, which captures how the filtered hidden signal affects the current drift through the function $f$;
2. $\eta$, the structural drift component, which represents the baseline deterministic tendency in the log-price dynamics;
3. $-c\, \hat{y}_t\, V_M'(\hat{x}_t)$, the nonlinear feedback component, which depends both on the filtered memory state and on the local slope of the Morse-like potential.

The interpretation is straightforward. When $\mu_x(t) > 0$, the model-implied drift is positive and suggests upward local pressure on the market. When $\mu_x(t) < 0$, it is negative and suggests downward local pressure. The filtered drift therefore provides a simple and model-consistent way to transform the latent Marketron state into a tradable directional signal.

### 3.2 Strategy Rules

We consider two implementations of the drift signal. In both cases, $\text{Position}_t$ denotes the portfolio exposure to the S&P 500 at time $t$: $\text{Position}_t = 0$ corresponds to a fully cash position, $\text{Position}_t = 1$ to full investment in the index, and intermediate values to partial exposure. Values above 1 correspond to leveraged exposure. In particular, the sized strategy allows exposure up to a maximum of 1.5, that is, 150% of the reference notional.

**Long/flat strategy.** The first implementation is a binary allocation rule:

$$\text{Position}_t = \begin{cases} 1 & \text{if } \mu_x(t) > 0, \\ 0 & \text{if } \mu_x(t) \leq 0. \end{cases} \tag{15}$$

When the model-implied drift is positive, the portfolio is fully invested in the S&P 500; otherwise, it moves entirely to cash. This is the most direct way to convert the filtered drift into a tradable allocation rule, and it will also serve as our main benchmark among the Marketron-based strategies.

**Sized strategy.** The second implementation allows the portfolio weight to vary continuously with the strength of the signal:

$$\text{Position}_t = \text{clip}\left(\frac{\mu_x(t)}{\hat{\sigma}_\mu(t)}, 0, 1.5\right), \tag{16}$$

where $\hat{\sigma}_\mu(t)$ denotes a trailing empirical standard deviation of the drift signal, computed causally using only information available up to time $t$. In our implementation, it is estimated over a rolling window of the most recent $L$ observations, together with a small positive floor to prevent excessively large exposures when recent signal variability becomes too small:

$$\hat{\sigma}_\mu(t) = \max\bigl(\text{Std}\bigl(\mu_x(s) : s = \max(0, t - L + 1), \ldots, t\bigr), \varepsilon\bigr). \tag{17}$$

Here $L$ denotes the window length and $\varepsilon > 0$ a fixed lower bound. In the empirical implementation, we set $L = 12$, corresponding to one year of monthly signal history, and $\varepsilon = 0.01$.

This normalization makes exposure depend on the strength of the current drift signal relative to its recent variability rather than on its absolute level alone. A weakly positive signal therefore generates a small position, whereas a strongly positive signal leads to a larger one. The clipping operator restricts the final exposure to the interval $[0, 1.5]$: negative signals imply zero exposure, moderately positive signals imply partial exposure, and sufficiently strong positive signals are capped at the maximum leverage level.

Both strategies incorporate transaction costs of 5 basis points per trade, applied to the absolute change in position size. This provides a simple and reasonable approximation for monthly rebalancing of a liquid index exposure.

### 3.3 Performance Metrics

We evaluate the strategies using standard portfolio-performance measures computed on monthly strategy log-returns, net of transaction costs. Annualized return measures the average yearly growth rate of the strategy and is computed as 12 times the mean monthly log-return. Annualized volatility measures the dispersion of monthly strategy log-returns scaled to a yearly basis. The Sharpe ratio summarizes risk-adjusted performance by dividing annualized return by annualized volatility.

Maximum drawdown is computed from the corresponding portfolio wealth process, obtained by exponentiating cumulative log-returns, and measures the worst peak-to-trough percentage decline in portfolio value over the period considered. Finally, the Calmar ratio relates annualized return to the magnitude of the maximum drawdown, and is therefore especially informative for strategies whose main objective is downside protection.

### 3.4 Backtest Design

Our backtest is based on 298 monthly S&P 500 log-price observations spanning January 2000 to October 2024. We divide the sample into an in-sample period (January 2000–December 2018, 228 months) and an out-of-sample period (January 2019–October 2024, 70 months).

In the main specification, model parameters are calibrated using only the in-sample subsample, i.e. January 2000 to December 2018. The calibration targets, including the return moments entering the objective function, are therefore computed exclusively from the training period. Once this parameter vector has been estimated, it is held fixed throughout the evaluation stage.

Conditional on these fixed parameters, the particle filter is then run forward through time in order to estimate the latent states and generate the corresponding trading signal. In that sense, the signal itself is computed causally: at each date, it only uses information available up to that point through the forward filtering pass. Portfolio positions over the 2019–2024 period are therefore generated without re-estimating parameters on the test sample, so this period remains genuinely out of sample from the perspective of model calibration.

This train/test design is important for methodological reasons. By calibrating the model only on 2000–2018 and evaluating the resulting strategy on 2019–2024, we avoid look-ahead bias at the parameter-estimation stage and obtain a clean out-of-sample assessment of the Marketron-based timing rule.

For visualization or descriptive analysis, the particle filter may still be run over the full 2000–2024 sample using the parameter vector estimated on the in-sample period. This does not alter the calibration protocol, since the parameters remain fixed and the filter is applied sequentially through time. However, all main trading results reported in the paper are based on the train/test logic described above.

# 4 Results

## 4.1 Strategy Performance

| | **In-Sample (2000–2018)** | | | **Out-of-Sample (2019–2024)** | | |
|---|---|---|---|---|---|---|
| **Metric** | B&H | Long/Flat | Sized | B&H | Long/Flat | Sized |
| Ann. return | +3.1% | +0.9% | +0.1% | +14.4% | +10.5% | +7.2% |
| Ann. volatility | 14.6% | 10.3% | 10.8% | 17.4% | 12.2% | 14.3% |
| Sharpe ratio | 0.212 | 0.084 | 0.013 | 0.831 | **0.864** | 0.503 |
| Max drawdown | −52.6% | −39.4% | −42.0% | −24.8% | **−20.1%** | −24.8% |
| Calmar ratio | 0.059 | 0.022 | 0.003 | 0.582 | 0.525 | 0.290 |
| # Trades | 0 | 27 | 122 | 0 | 8 | 35 |

Table 3: Strategy performance metrics, net of 5 bps transaction costs per trade, under the final IS-only calibration.

Table 3 compares the Marketron-based strategies with the buy-and-hold benchmark. The in-sample period contains the dot-com bust and the 2008 financial crisis, while the out-of-sample period contains the COVID crash and the 2022 rate-hiking selloff.

Several conclusions emerge.

1. **The long/flat rule is better interpreted as a defensive overlay than as a source of robust alpha.** In-sample, it underperforms buy-and-hold in both annualized return and Sharpe ratio: return falls from +3.1% to +0.9%, and the Sharpe ratio from 0.212 to 0.084. However, this weaker performance is accompanied by a clear reduction in risk, with annualized volatility declining from 14.6% to 10.3% and maximum drawdown improving from −52.6% to −39.4%.

2. **Out of sample, the long/flat rule looks more attractive.** Over 2019–2024, it delivers an annualized return of +10.5%, compared with +14.4% for buy-and-hold, but with lower volatility (12.2% versus 17.4%), a slightly higher Sharpe ratio (0.864 versus 0.831), and a

smaller maximum drawdown ($-20.1\%$ versus $-24.8\%$). In this sense, the strategy appears to exchange part of the upside for a somewhat smoother path of returns.

3. **These out-of-sample gains should nevertheless be interpreted cautiously.** The Sharpe improvement is economically modest, the test window contains only 70 monthly observations, and the strategy does not display superior in-sample Sharpe performance. The most reasonable interpretation is therefore not that the model has uncovered a robust predictive edge, but rather that the filtered drift may contain limited but potentially useful defensive timing information.

4. **The sized strategy is less convincing than the binary long/flat rule.** In-sample, its Sharpe ratio is close to zero (0.013) and its drawdown remains large ($-42.0\%$). Out of sample, it improves relative to its in-sample behavior, but still remains weaker than the long/flat rule on both Sharpe ratio and drawdown. In our setting, continuous scaling of exposure does not improve the economic usefulness of the signal.

5. **Turnover remains moderate for the long/flat strategy.** The rule makes 27 trades in-sample and 8 trades out-of-sample, for a total of 35 trades over the full sample. At monthly frequency, this remains operationally manageable and keeps transaction-cost drag limited.

Figure 5 makes these trade-offs visually clear. The top panel shows that the model-implied drift alternates regularly between positive and negative phases, which mechanically generates a sequence of long and cash periods in the second panel. The third panel shows that the long/flat wealth path is smoother than buy-and-hold, but also lags behind it during the strongest sustained bull-market phases, especially after 2020. The bottom panel confirms the main economic benefit of the strategy: drawdowns are visibly shallower than under buy-and-hold, both around the 2008 crisis and in the later out-of-sample period.

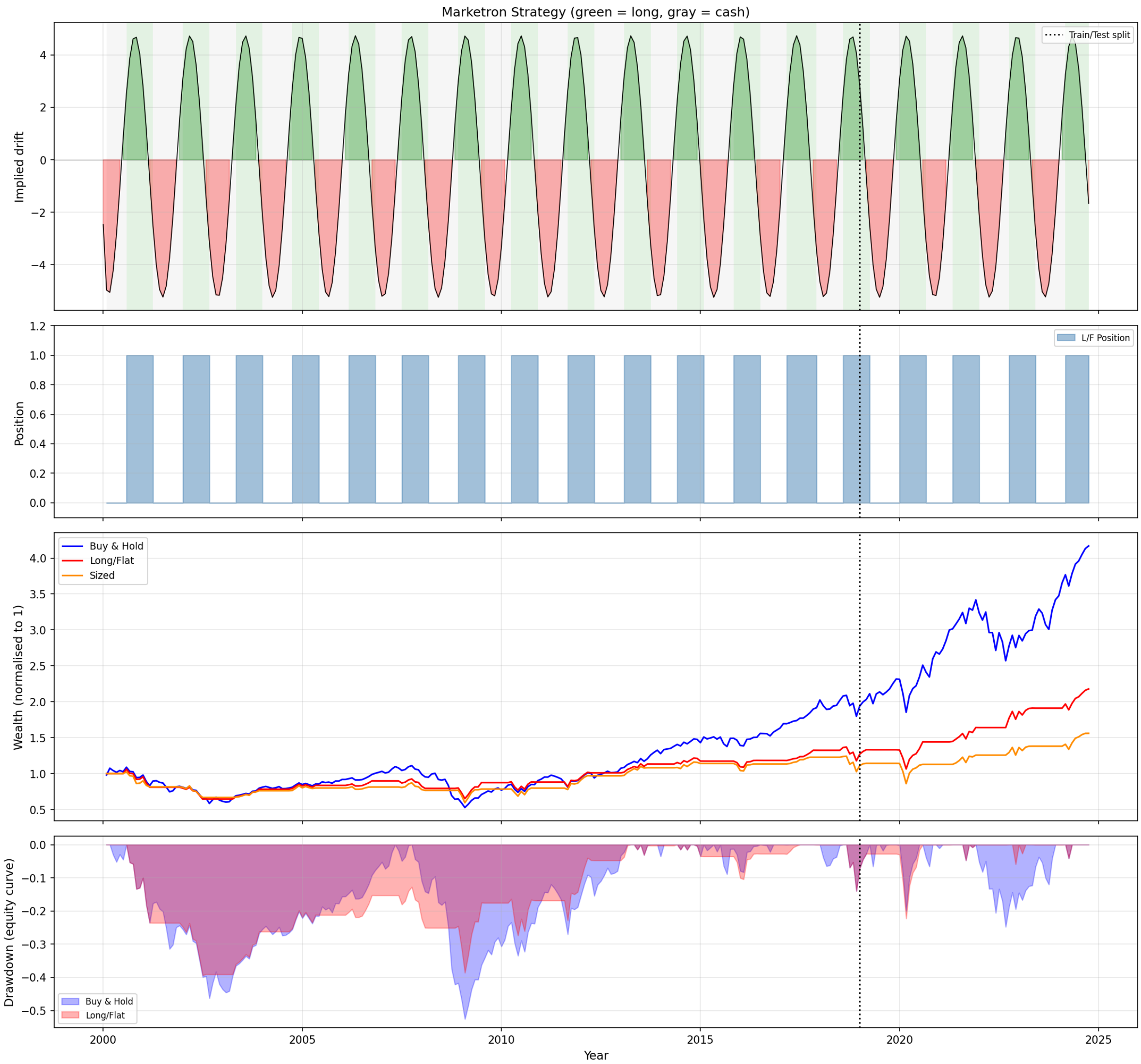



Figure 5: **Strategy backtest (full period, 2000–2024).** Panel (a): the model-implied drift signal $\mu_x(t)$ from Eq. (14), with green and red shading indicating positive and negative values. Panel (b): the corresponding long/flat position rule. Panel (c): normalized wealth paths for buy-and-hold, long/flat, and sized strategies. Panel (d): drawdown computed from the corresponding wealth processes. Under the final IS-only calibration, the long/flat strategy reduces downside risk relative to buy-and-hold, but does not dominate it uniformly in cumulative performance. The dashed vertical line marks the in-sample/out-of-sample split at January 2019.

Figure 6 provides a compact visual summary of the same comparison. It is especially useful for highlighting the asymmetry between the two subsamples: the long/flat rule is weaker than buy-and-hold in-sample on Sharpe ratio, but looks somewhat more favorable out-of-sample, especially from a downside-risk perspective.


Figure 6: **In-sample versus out-of-sample strategy comparison.** Each bar group reports annualized return, annualized volatility, Sharpe ratio, and maximum drawdown for the three strategies. The figure highlights the main asymmetry of the results: the long/flat rule is not stronger than buy-and-hold in-sample on Sharpe ratio, but appears somewhat more attractive out-of-sample, especially from a downside-risk perspective.

## 4.2 Why Does the Strategy Work? Interpretation via the Model

Figure 7 shows the filtered hidden states of the model over the full sample. The top panel compares observed and filtered log-prices and shows that the particle filter tracks the market trajectory very closely throughout the period. The second panel displays the filtered memory variable $y_t$, which fluctuates around its long-run level $\bar{y}$ while evolving more smoothly than the observed price. The third panel reports the filtered signal variable $\theta_t$, which rapidly converges toward its long-run level $\hat{\theta}$ and then remains within a relatively narrow band around it.

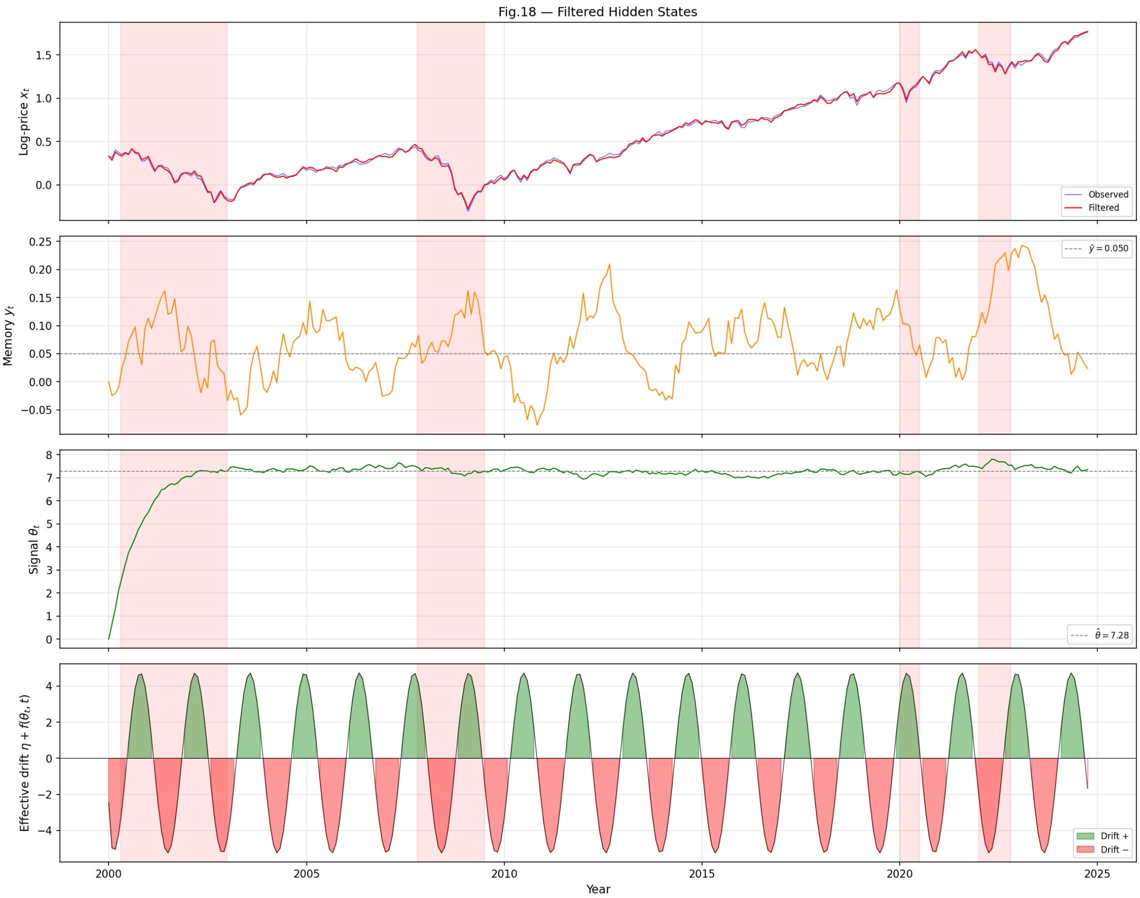



Figure 7: **Filtered hidden states.** Panel (a): observed and filtered log-price. Panel (b): filtered memory variable $y_t$, shown together with its long-run level $\bar{y}$. Panel (c): filtered signal $\theta_t$, shown together with its long-run level $\hat{\theta}$. Panel (d): the signal-driven component $\eta + f(\theta_t, t)$ of the model drift, with green and red areas indicating positive and negative values. Pink shaded regions mark major market stress episodes for visual reference.

The main message of Figure 7 comes from the bottom panel. The quantity $\eta + f(\theta_t, t)$, which captures the structural drift together with the direct contribution of the latent signal, exhibits a strongly periodic pattern, with regular alternations between positive and negative phases. This suggests that an important part of the timing behavior generated by the model comes from a latent cyclical component embedded in the calibrated signal specification.

This point is important for interpretation. The model should not be viewed as a real-time crash detector in which the drift simply turns negative in response to contemporaneous market stress. Rather, the filtered Marketron state combines a smooth latent signal, a memory component, and nonlinear price feedback, and the resulting timing rule behaves more like a regime-aware cyclical overlay. When negative-drift phases happen to overlap with periods of market weakness, the strategy can reduce downside exposure; when they do not, the strategy may instead miss part of the upside.

At the same time, Figure 7 also shows why the model remains economically interpretable despite its mixed performance. The particle filter does not merely produce a black-box allocation rule: it decomposes market dynamics into an observed price trajectory, a persistent memory state, and a slowly varying latent signal. Even if the resulting strategy does not provide robust statistical dominance, the model still offers a coherent state-space interpretation of changing market conditions and a disciplined way to translate that interpretation into portfolio exposure.

### 4.3 Regime Detection on Real Data

One of our original contributions is to build a simple empirical regime-classification scheme inspired by the Good/Bad/Ugly interpretation of the Marketron model and apply it to filtered S&P 500 data. The classification is constructed from rolling 12-month returns and rolling 12-month volatility computed from the filtered log-price. Months with negative 12-month return are classified as *Bad*; among months with positive 12-month return, those with below-median volatility are classified as *Good*, while the remaining positive-return months are classified as *Ugly*, interpreted as more fragile or transitional states.

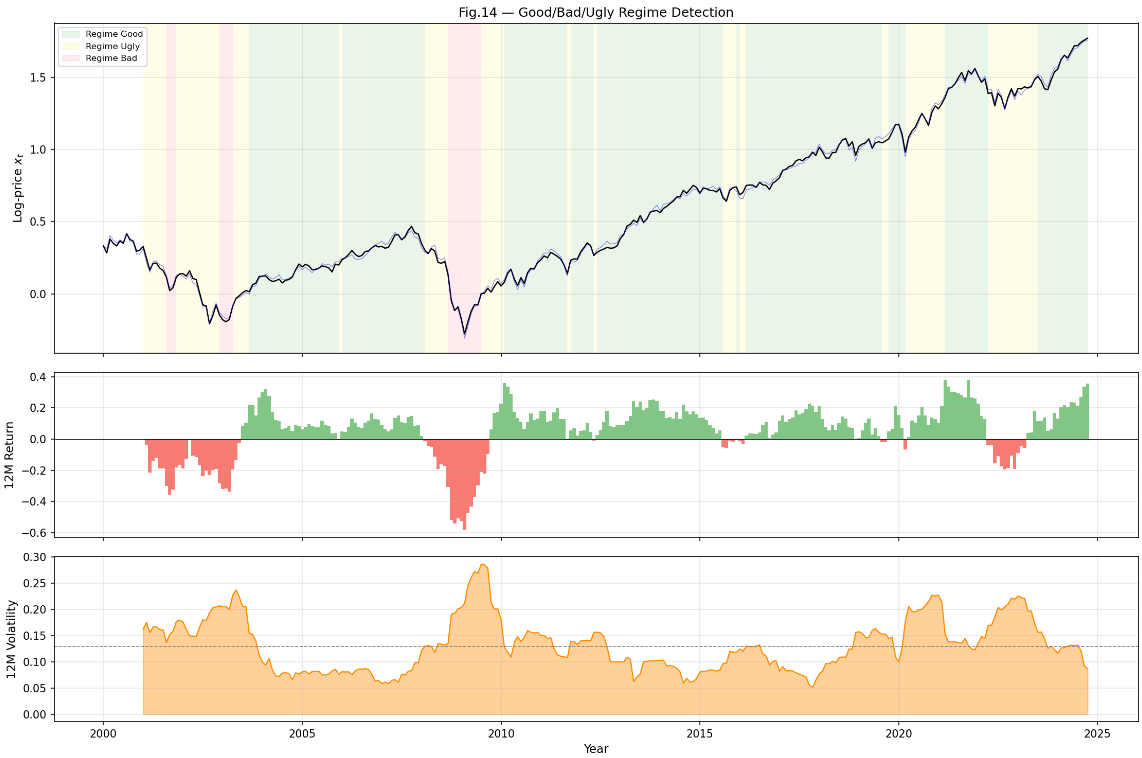



Figure 8: **Market regime classification (2000–2024). Top panel:** observed and filtered log-price with colored background indicating the detected regime at each date. Green = Good, red = Bad, yellow = Ugly. **Middle panel:** rolling 12-month return computed from the filtered log-price; negative values correspond to the Bad classification. **Bottom panel:** rolling 12-month volatility. The dashed line marks the median volatility threshold used to separate Good from Ugly among positive-return months. The figure shows that large market downturns are typically classified as Bad, while long rising periods with more moderate volatility tend to be classified as Good. The Ugly regime captures intermediate phases with positive but more fragile market conditions.

Figure 8 provides an intuitive descriptive link between the filtered Marketron dynamics and the three-regime narrative of the model. In particular, the classification identifies the major bearish episodes of the sample as Bad regimes, while extended upward phases are mostly labeled Good. The Ugly regime appears in intermediate periods where returns remain positive but volatility is elevated, which is consistent with the idea of a transition or unstable plateau between more clearly favorable and unfavorable conditions.

This classification should nevertheless be interpreted with caution. It is not obtained by directly mapping the filtered latent state into the minima and saddle structure of the theoretical potential. Rather, it is a heuristic empirical overlay built from filtered prices, designed to give an economically interpretable real-data counterpart to the Good/Bad/Ugly language of the model.

For that reason, the figure should be read as an illustration that the regime narrative is broadly consistent with the filtered data, rather than as a literal structural identification of potential wells and saddle points in observed market history.

### 4.4 Comparison with Geometric Brownian Motion

To assess whether the added complexity of the Marketron framework is justified, we compare its moment predictions with those of the simplest benchmark, geometric Brownian motion (GBM). Under GBM, returns are Gaussian, so the model can in principle match the mean and volatility through its two parameters $\mu$ and $\sigma$, but it cannot generate nonzero skewness or excess kurtosis. These higher-order moments are identically zero by construction.

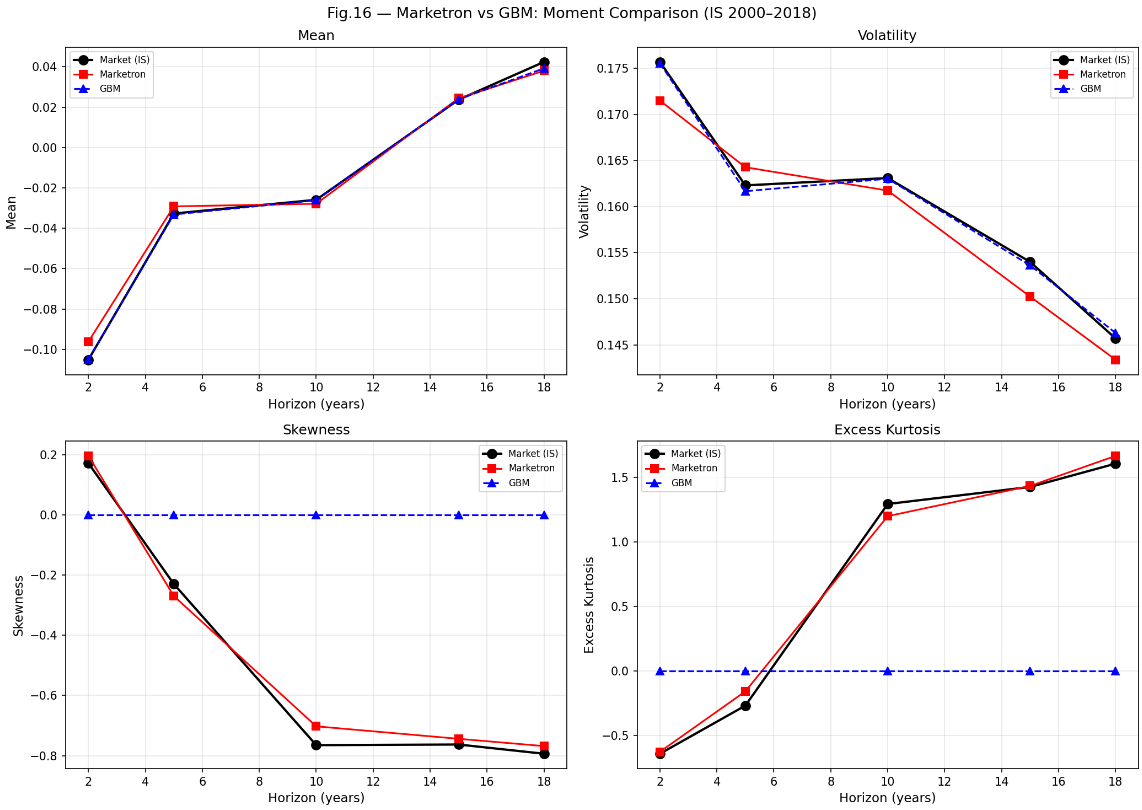



Figure 9: **Moment comparison across in-sample horizons.** Market data (black circles), Marketron (red squares), GBM (blue triangles). **Top-left:** mean return. **Top-right:** volatility. **Bottom-left:** skewness. **Bottom-right:** excess kurtosis. GBM tracks the first two moments reasonably well but is structurally unable to reproduce the nonzero skewness and kurtosis observed in the data, whereas Marketron matches all four moments closely across horizons.

Figure 9 makes the comparison transparent. For mean return and volatility, both models are reasonably close to the empirical targets. The main difference appears in the lower panels. Because GBM imposes Gaussian returns, it remains fixed at zero skewness and zero excess kurtosis for every horizon. By contrast, Marketron reproduces the positive short-horizon skewness, the strongly negative skewness observed from medium horizons onward, and the rise in excess kurtosis at longer horizons.

This difference is also visible in the aggregate error metric. On the in-sample moment targets used in calibration, the total moment MSE is 0.032 for Marketron, compared with 8.644 for GBM. Thus, on our in-sample criterion, Marketron outperforms GBM by a very large margin. The reason is not that it improves dramatically on the first two moments, where GBM already performs adequately, but that it captures the higher-order non-Gaussian structure of the data that GBM cannot match at all.

The advantage of Marketron over GBM is therefore mainly structural. GBM is too rigid to reproduce horizon-dependent asymmetry and tail behavior, whereas the Marketron framework generates these features naturally through its nonlinear state dynamics. In this sense, Figure 9 provides one of the clearest justifications for the additional modeling complexity: the gain lies primarily in the ability to fit skewness and kurtosis while remaining within an interpretable latent-state framework.

At the same time, GBM remains only a baseline benchmark. It is deliberately simple, and stronger alternatives such as stochastic-volatility or regime-switching models could provide a more demanding comparison. We use GBM here because it is the canonical benchmark and therefore the most natural first point of reference.

### 4.5 Model vs. Market: Log-Returns and Higher-Order Moments

Figures 9 and 10 provide a direct visual comparison between the filtered Marketron output and the observed S&P 500 series. They serve two distinct purposes. Figure 9 assesses the time-series fit of the filtered model at the monthly frequency, while Figure 10 focuses on how the distribution of returns changes with the horizon, especially through skewness and excess kurtosis.

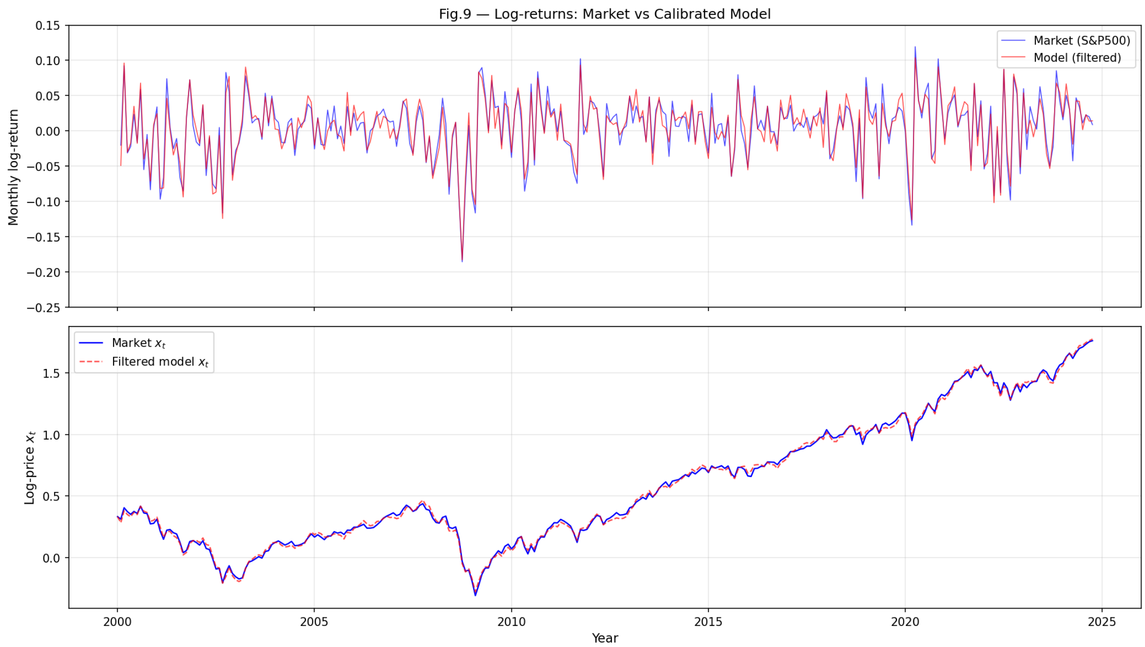



Figure 10: **Model versus market at the monthly frequency. Top panel:** monthly log-returns from the filtered model (red) versus the observed S&P 500 series (blue). **Bottom panel:** observed and filtered log-price levels. The filtered path tracks the market closely over the full sample, while the return series reproduces most of the major monthly fluctuations, including the large negative moves around the 2008 crisis and the COVID shock.

Figure 10 shows that the particle-filtered model remains tightly anchored to the observed market trajectory. This is especially clear in the bottom panel, where the filtered log-price almost overlaps the observed series over the full sample. The top panel shows that the model also reproduces much of the month-to-month variation in returns, including the main spikes and reversals during major stress episodes. This figure is therefore useful primarily as a diagnostic of filtering quality: it confirms that the latent-state model can track the observed price path closely enough to support the downstream strategy and regime analysis.

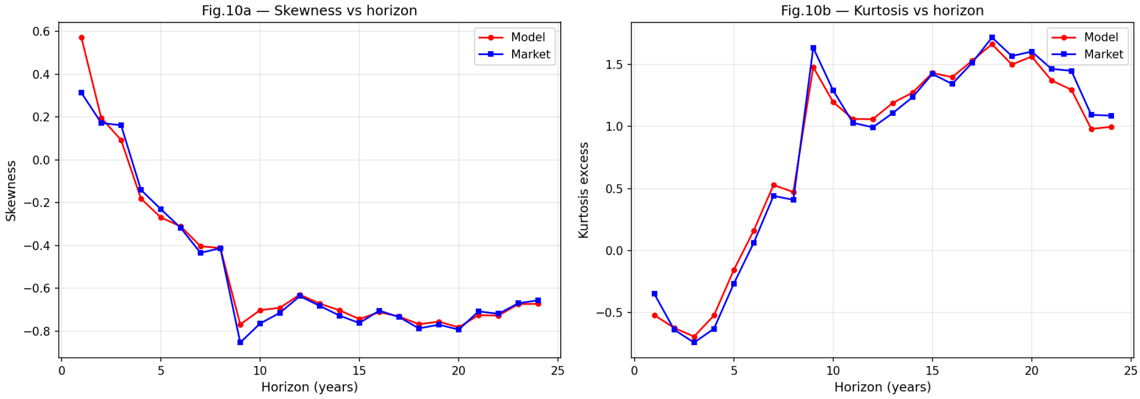



Figure 11: **Skewness and excess kurtosis across horizons. Left:** skewness of cumulative returns as a function of horizon for the market and the calibrated model. **Right:** excess kurtosis as a function of horizon. The model captures the main horizon-dependent non-Gaussian patterns observed in the data, including increasingly negative skewness and elevated excess kurtosis at medium and long horizons.

Figure 11 complements the time-series comparison by focusing on higher-order moments across horizons. The left panel shows that skewness becomes increasingly negative as the horizon grows, a key asymmetry that a Gaussian benchmark cannot reproduce. The right panel shows that excess kurtosis rises markedly at medium and long horizons, indicating heavier tails than under a normal distribution. The calibrated Marketron model follows both patterns closely. This figure is useful because it isolates the two stylized facts that most directly justify moving beyond GBM: horizon-dependent asymmetry and non-Gaussian tail thickness.

## 4.6 Further Results: Instantons, Default Rates, and Phase Diagrams

### 4.6.1 Instanton Trajectories

To complement the empirical results, we also illustrate the transition geometry of the qualitative two-dimensional potential through minimum-energy paths (MEPs). The goal of this figure is not empirical fitting, but rather to visualize how the Good/Bad/Ugly landscape can support regime-to-regime transitions in the deterministic skeleton of the model.

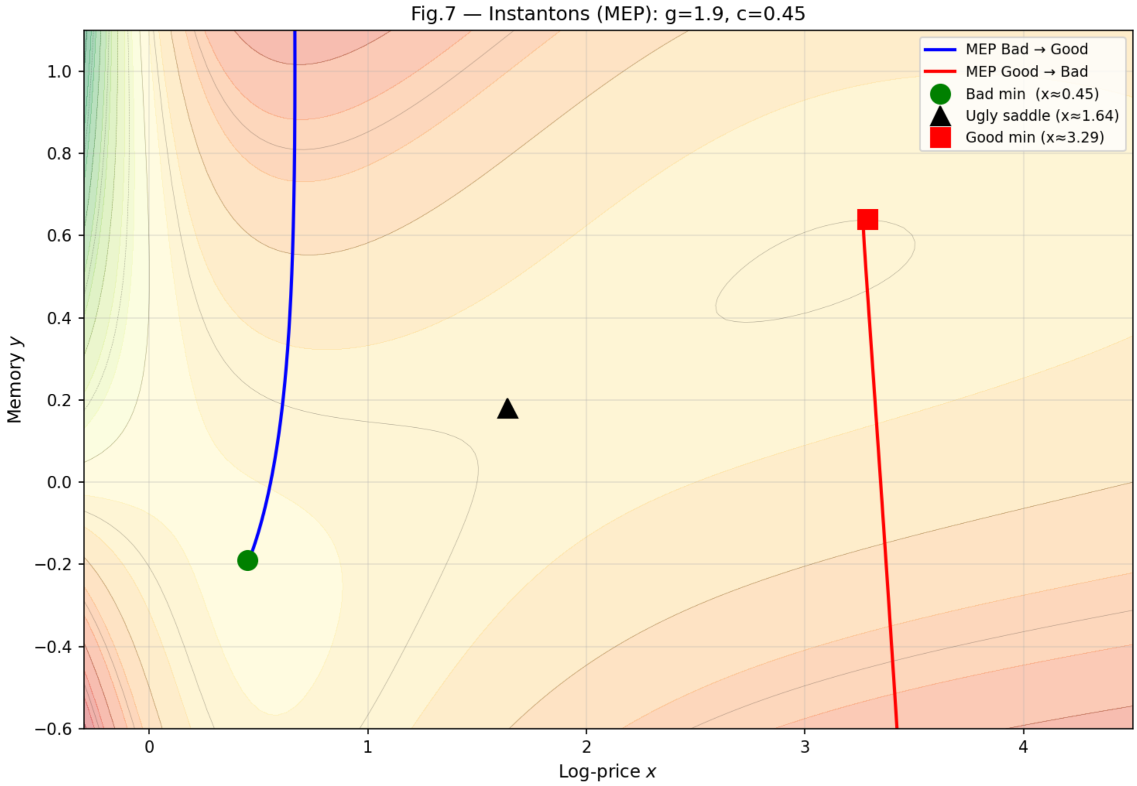



Figure 12: **Minimum-energy-path illustration in the qualitative 2D potential.** The black triangle marks the intermediate Ugly saddle, while the green circle and red square mark the Bad and Good minima, respectively. The blue and red curves show the two local descent branches associated with the saddle structure, one toward the Bad basin and one toward the Good basin. The contour background displays the two-dimensional potential $V(x, y)$ under a qualitative parameter choice chosen to produce three distinct equilibria.

Figure 12 provides a geometric illustration of the multi-regime structure discussed earlier. The potential exhibits two local minima, interpreted as Bad and Good regimes, separated by an intermediate saddle interpreted as Ugly. In that setting, the MEP construction gives a natural qualitative picture of how transitions between basins can occur: the saddle acts as the unstable gateway between the two metastable regions.

The figure is useful mainly as a conceptual complement to the Good/Bad/Ugly narrative of the model. It should not be over-interpreted quantitatively. In particular, the displayed paths are obtained under qualitative parameter values chosen to make the three-equilibria structure clearly visible, rather than under the final empirical calibration used for filtering and trading. The role of the figure is therefore to clarify the geometry of the landscape, not to provide a data-driven estimate of actual market-transition paths.

#### 4.6.2 Default Rate

The Marketron framework also allows a qualitative discussion of rare extreme downside events. In our simulations, we define a default-like event as the crossing of the threshold $x_t < -5$, which corresponds to a loss of more than 99% in price relative to the reference level $S^*$. Using qualitative parameter values and a 24-year Monte Carlo horizon, we simulate 10,000 independent paths and record how often such an extreme barrier is hit.

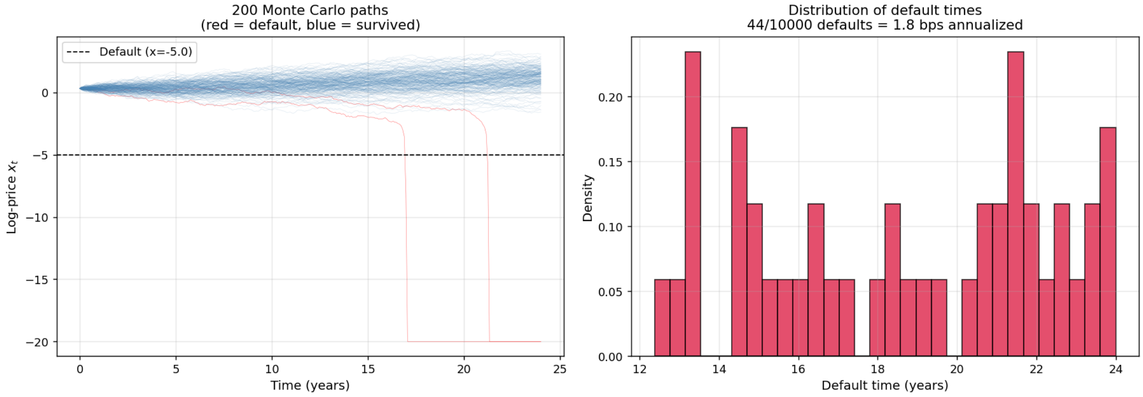


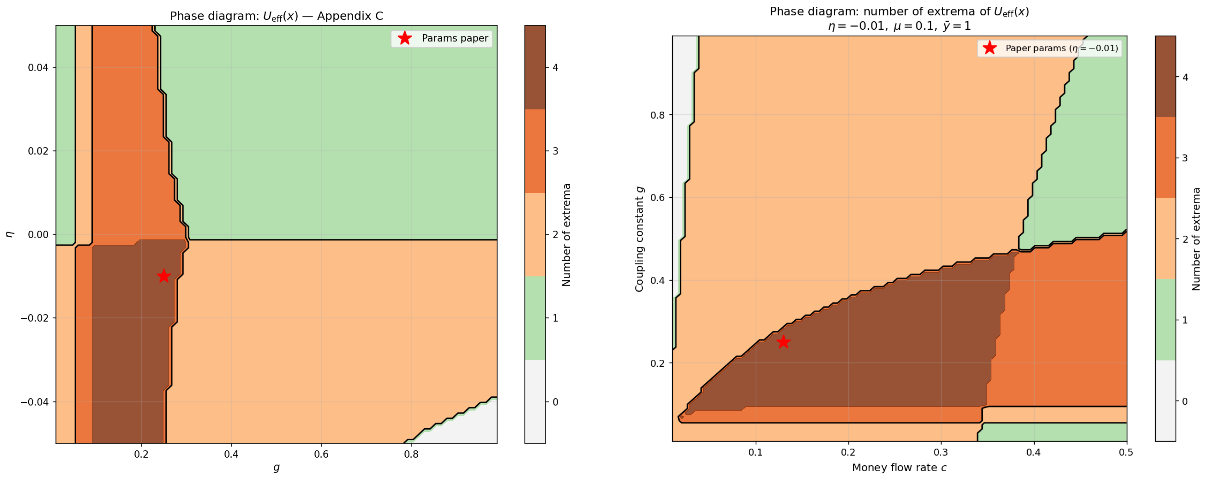

Figure 13: **Monte Carlo default analysis under qualitative parameters. Left panel:** 200 simulated paths, with red trajectories indicating paths that hit the default threshold $x = -5$ and blue trajectories indicating survivors. **Right panel:** empirical distribution of default times among the defaulting paths. In this qualitative experiment, defaults are rare and tend to occur late in the simulation horizon.

Figure 13 suggests two qualitative messages. First, default-like events are rare under this specification: only a very small fraction of simulated paths reach the extreme lower barrier. Second, when defaults do occur, they typically occur relatively late in the simulation window rather than immediately. This is consistent with the metastability interpretation of the model: the system can remain trapped for a long time in a locally stable region before a sufficiently adverse sequence of shocks pushes it across the barrier.

This figure is mainly illustrative and should be interpreted cautiously. The simulation uses qualitative parameter values rather than the final empirical calibration retained for filtering and trading. As a result, the reported frequency is best read as a conceptual demonstration that the model can generate rare barrier-crossing events, rather than as a literal estimate of default intensity for the S&P 500 or of market-implied credit spreads.

#### 4.6.3 Phase Diagrams

To further clarify when the Good/Bad/Ugly structure can arise, we compute phase diagrams for the effective one-dimensional potential $U_{\text{eff}}(x)$. Each point in parameter space is colored according to the number of extrema of $U_{\text{eff}}$: one extremum corresponds to a single-regime landscape, while three extrema correspond to the double-well structure associated with two local minima separated by an intermediate saddle-type barrier.




(a) $(g, \eta)$ space at fixed qualitative parameters.

(b) $(c, g)$ space at fixed $\eta = -0.01$.

Figure 14: **Phase diagrams for the effective potential.** Colors indicate the number of extrema of $U_{\text{eff}}(x)$. Regions with one extremum correspond to a single-regime landscape, while regions with three extrema support the qualitative Good/Bad/Ugly structure. The starred point indicates the qualitative parameter choice highlighted in the paper-style illustration.

Figure 14 shows that the three-regime structure only exists in restricted regions of parameter space. In the left panel, varying the coupling parameter $g$ and the global tilt $\eta$ reveals that the multi-extrema regime disappears when the tilt becomes too strong or when the coupling is too weak. In the right panel, varying $c$ and $g$ shows a similar phenomenon: the double-well structure requires a sufficiently strong nonlinear interaction, but only within a limited band of parameter values.

These phase diagrams are useful because they make the regime mechanism more systematic. Rather than relying on a single illustrative parameter choice, they show that the Good/Bad/Ugly landscape is not generic: it emerges only when the feedback terms are strong enough to counterbalance the global tilt of the system. In that sense, the figures provide a compact map of when the model behaves as a single-regime diffusion and when it admits the richer metastable structure emphasized throughout the paper.

As with the instanton and default illustrations, these phase diagrams should be read as qualitative structural results rather than as empirical outputs of the final calibration. Their role is to explain the geometry of the model and the conditions under which multiple regimes can appear.

### 4.7 Out-of-Sample Validation of the Particle Filter

Figure 15 evaluates whether the filtering step remains stable after the train/test split. This figure is useful because the trading signal is constructed from filtered latent states; if the filter were to deteriorate materially out of sample, the downstream strategy results would be difficult to interpret.

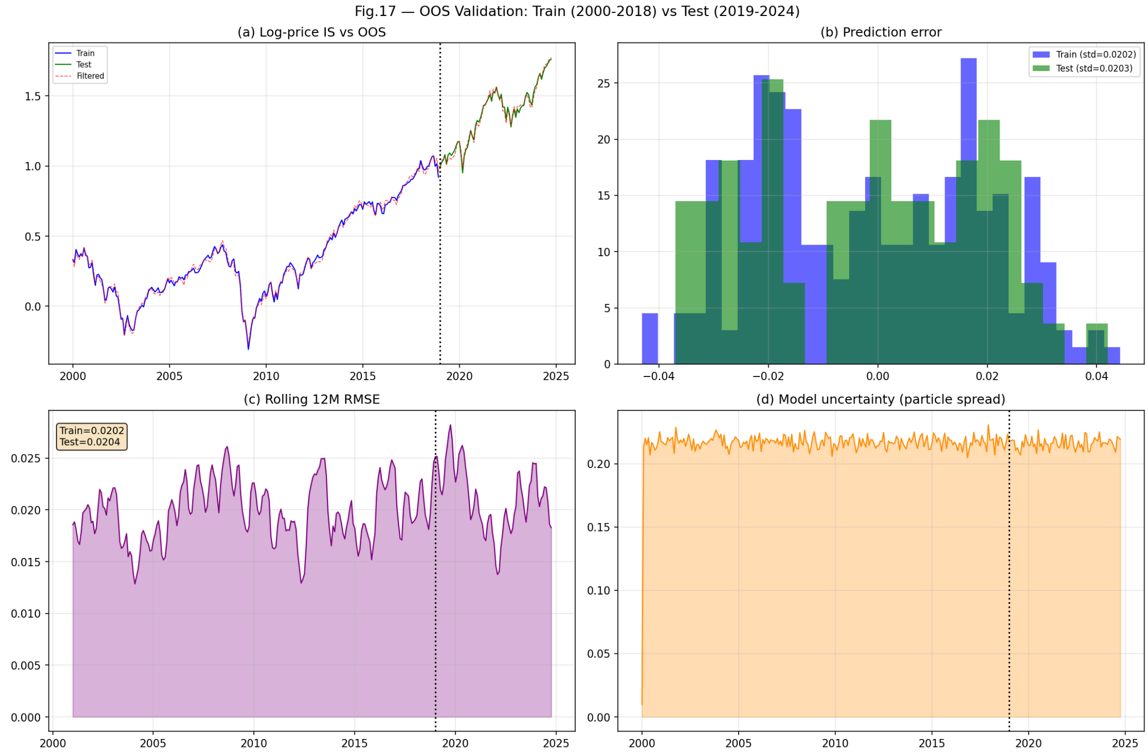



Figure 15: **Particle filter: in-sample versus out-of-sample validation.** **(a)** Observed and filtered log-price, with the train/test split marked by the dashed vertical line. **(b)** Distribution of filtering errors in-sample and out-of-sample. **(c)** Rolling 12-month RMSE over the full sample. **(d)** Particle spread, used here as a proxy for model uncertainty. The overall filtering error remains very similar across the two subsamples, with train RMSE = 0.0202 and test RMSE = 0.0204.

The main message of Figure 15 is that the particle filter remains stable out of sample. In panel (a), the filtered path continues to track the observed log-price closely after the 2019 split, just as it does in-sample. Panel (b) reinforces this point by showing that the distributions of filtering errors in the two periods are very similar. Quantitatively, the train and test RMSE values are 0.0202 and 0.0204, respectively, which corresponds to a test/train ratio very close to one.

Panel (c) shows that the rolling 12-month RMSE remains within a relatively narrow range over the full sample, without any visible break after the train/test split. This suggests that the model does not suffer from an obvious loss of tracking quality once it is applied to unseen data. Finally, panel (d) indicates that the particle spread remains well behaved over time, which is consistent with a filter that stays numerically stable and does not collapse or explode in the out-of-sample period.

Taken together, these diagnostics support the interpretation that the out-of-sample strategy results are not driven by an unstable or overfit filtering procedure. This does not prove predictive power, but it does show that the latent-state estimation step generalizes in a reasonably stable way from the calibration sample to the test sample.

## 4.8 Robustness of the Trading Signal

A natural question is whether the strategy results depend critically on the exact implementation of the trading rule. To address this, we test several variants of the long/flat signal. Out-of-sample robustness metrics should nevertheless be interpreted cautiously, since the test window contains only 70 monthly observations and coincides with a relatively strong post-2019 equity market.

The variants are as follows:

- **Threshold variations:** instead of going long when $\mu_x > 0$, we test thresholds $-0.5$, $+0.5$, and $+1.0$. The negative threshold is more aggressive, whereas the positive thresholds are more conservative.
- **1-month lag:** the position at time $t$ is based on the signal computed at $t-1$, which provides a stricter causal implementation.
- **Quarterly rebalancing:** the position is updated every three months rather than every month.
- **Noisy signal:** Gaussian noise with standard deviation equal to 30% of the drift-signal standard deviation is added before applying the trading rule, in order to test sensitivity to estimation noise.

| Variant | In-Sample Sharpe | In-Sample Max DD | Out-of-Sample Sharpe | Out-of-Sample Max DD |
|---|---|---|---|---|
| Baseline (thr = 0) | 0.084 | −39.4% | 0.864 | −20.1% |
| Threshold = −0.5 | 0.178 | −38.9% | 0.870 | −20.0% |
| Threshold = +0.5 | -0.006 | −43.6% | 0.763 | −20.1% |
| Threshold = +1.0 | -0.052 | −43.5% | 0.759 | −20.1% |
| 1-month lag | 0.112 | −41.6% | 0.698 | −20.0% |
| Quarterly rebalancing | 0.281 | −40.7% | 0.771 | −20.0% |
| Noisy signal (30%) | 0.016 | −34.7% | 0.858 | −20.0% |
| Buy & Hold | 0.212 | −52.6% | 0.831 | −24.8% |

Table 4: Robustness of the long/flat strategy under signal perturbations and implementation changes.

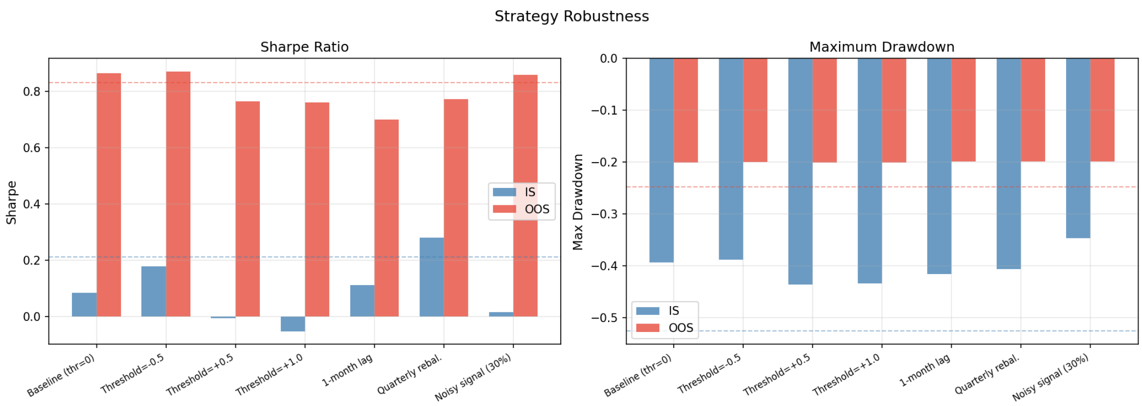



Figure 16: **Robustness tests for the long/flat strategy.** Left: Sharpe ratio for each variant (blue = in-sample, red = out-of-sample). Dashed lines indicate the corresponding buy-and-hold Sharpe ratios. Right: maximum drawdown in-sample and out-of-sample, with dashed lines indicating the buy-and-hold drawdowns.

Figure 16 and Table 4 show that the main qualitative message is fairly robust, but also more nuanced than in earlier full-sample experiments. In-sample performance remains mixed across variants: some choices, such as quarterly rebalancing or the threshold $-0.5$, modestly improve the Sharpe ratio relative to the baseline, whereas more conservative thresholds ($+0.5$ and $+1.0$)

lead to negative in-sample Sharpe ratios. This confirms that the signal is not a uniformly strong alpha source in-sample.

At the same time, the drawdown profile is consistently better than buy-and-hold across all variants. In-sample, every strategy variant produces a substantially smaller maximum drawdown than the benchmark, with values between roughly −34.7% and −43.6%, compared with −52.6% for buy-and-hold. Out-of-sample, the drawdown figures are also tightly clustered around −20%, which is slightly better than the −24.8% of buy-and-hold. This supports the interpretation that the main value of the signal lies in downside mitigation rather than in return maximization.

The out-of-sample Sharpe ratios are also relatively stable. All strategy variants remain in a band from roughly 0.70 to 0.87, with the baseline, threshold −0.5, and noisy-signal versions all slightly above the buy-and-hold Sharpe of 0.831. This suggests that the broad out-of-sample behavior is not entirely driven by a knife-edge implementation choice. However, since the out-of-sample window is short and unusually favorable to equity exposure overall, these differences should still be read as suggestive rather than conclusive.

Two specific variants are especially informative. First, the 1-month lagged version still achieves an out-of-sample Sharpe of 0.698 with a maximum drawdown of about −20.0%, showing that some defensive behavior survives even under a stricter delayed implementation. Second, the noisy-signal specification leaves the out-of-sample Sharpe almost unchanged at 0.858 while actually improving the in-sample drawdown to −34.7%, suggesting that the signal is not excessively fragile to moderate estimation noise.

Overall, the robustness analysis reinforces the main conclusion of the paper. The exact Sharpe ratio depends on implementation details and is not uniformly strong in-sample, but the signal's defensive properties persist across a broad set of perturbations. This again points to the same interpretation: the Marketron drift is better viewed as a regime-sensitive overlay for risk management than as a stable source of standalone predictive alpha.

# 5 Discussion

## 5.1 On the Nature of the Calibrated Model

An important finding of the project is that the empirically calibrated Marketron specification should be interpreted primarily as a *conditional filtering model*, not as a fully autonomous generative model for realistic long-run free simulations. In practice, the economically meaningful outputs of the framework arise when the dynamics are continuously coupled with observed data through the particle filter.

This point is illustrated by Figure 17, which shows a free forward simulation run with the calibrated parameter vector, i.e. without conditioning on observed market prices. In that setting, the simulated log-price eventually collapses to the imposed lower bound, while the memory variable becomes unstable and the signal variable converges toward its long-run level. The figure therefore shows that the calibrated specification does not by itself generate realistic unconditional market paths over long horizons.

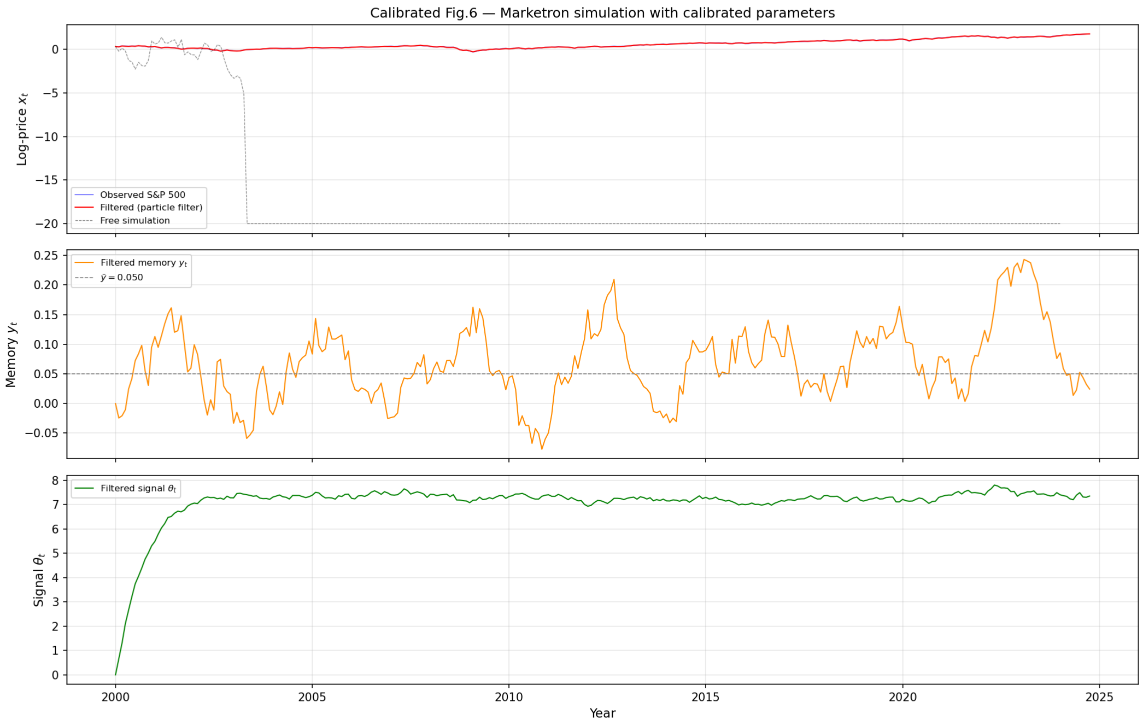



Figure 17: **Illustrative free simulation under the calibrated parameter vector (without data conditioning).** Top: simulated log-price $x_t$. Middle: simulated memory variable $y_t$. Bottom: simulated signal $\theta_t$. The path illustrates that the empirically calibrated specification is not meant to be interpreted as a realistic unconditional simulator when run freely over long horizons; its useful empirical content arises instead through filtering and conditioning on observed data.

The reason is closely related to the way the model is used in the paper. The calibration objective is built from filtered trajectories and moment targets computed on observed data, not from unconditional long-run simulations detached from the data stream. As a consequence, the retained parameter vector should be understood as one that allows the state-space model, together with the particle filter, to track the observed market and reproduce the relevant in-sample distributional features. It is not guaranteed to produce realistic free-running dynamics when the observational anchor is removed.

This does not make the framework useless; rather, it clarifies what kind of model it is. Many practically successful dynamical systems are used in exactly this conditional sense. In weather forecasting, data assimilation is essential. In engineering, Kalman-type filters are designed precisely to combine model structure with noisy measurements. In the same spirit, the relevant question here is not whether the calibrated Marketron system is a realistic autonomous simulator, but whether the *conditioned* model extracts useful latent-state information from observed prices. Our empirical results suggest that it does, at least to the extent of providing a potentially useful defensive timing overlay and a strong fit to higher-order in-sample return moments.

## 5.2 Limitations and Future Work

- **Limited out-of-sample evidence.** The strategy results are evaluated on only 70 out-of-sample monthly observations. This is enough to give an initial indication of stability, but not enough to support strong claims of predictive superiority. A longer test window, or repeated historical subperiod evaluation, would be needed for a more convincing assessment.
- **Conditional rather than autonomous modeling.** As discussed above, the calibrated

specification should be interpreted mainly as a conditional state-space model coupled with filtering, not as a realistic free-running simulator. A useful direction for future work would be to separate more clearly the parameters governing filtering performance from those governing economically meaningful unconditional simulation.

- **Monthly granularity.** The entire analysis is conducted at monthly frequency. This ignores within-month timing, execution frictions, and the possibility that the latent-state signal may behave differently at weekly or daily horizons.

- **Optimizer and local-minimum sensitivity.** Although the retained CMA-ES + L-BFGS-B pipeline performed best in our benchmark, the calibration problem remains non-convex and noisy. Different runs may still lead to materially different signal properties. Systematic multi-seed ensembling or Bayesian parameter uncertainty quantification would therefore be natural extensions.

- **Benchmark scope.** Our benchmark comparison focuses on GBM because it is the canonical baseline. This is useful pedagogically, but limited economically. A stronger empirical benchmark set would include stochastic-volatility models, regime-switching models, or simpler time-series timing rules.

- **Strategy design remains simple.** The long/flat and sized rules are intentionally minimal. Future work could combine the drift signal with volatility targeting, uncertainty-aware position sizing, or explicit regime-confidence filters. It would also be natural to study whether the signal adds value when combined with standard trend-following or drawdown-control overlays.

- **Walk-forward recalibration.** The final paper already uses an in-sample-only calibration followed by a genuine out-of-sample evaluation, which removes the earlier look-ahead issue. However, a more realistic deployment protocol would recalibrate the model sequentially over time on expanding or rolling windows, rather than fixing parameters once and for all.

# 6 Conclusion

We have implemented the Marketron model, calibrated it to S&P 500 data under a clean in-sample/out-of-sample protocol, and used its filtered latent state to construct a trading signal for equity timing. The final retained calibration pipeline is CMA-ES + L-BFGS-B on the 2000–2018 in-sample window, combined with a bootstrap particle filter with 1500 particles.

On the modeling side, the Marketron framework successfully reproduces the main non-Gaussian features of in-sample returns across horizons, including horizon-dependent mean returns, negative skewness at medium and long horizons, and positive excess kurtosis. On our in-sample moment-matching criterion, it substantially outperforms a geometric Brownian motion benchmark, with a total moment MSE of 0.032 against 8.644 for GBM. In this sense, the model's main statistical contribution is not to improve the fit of the first two moments alone, but to capture higher-order structure that a Gaussian benchmark cannot reproduce by construction.

On the trading side, the filtered model-implied drift yields a mixed but economically interpretable signal. The long/flat strategy does not outperform buy-and-hold in-sample: its Sharpe ratio is only 0.084, compared with 0.212 for the benchmark. Out-of-sample, however, it delivers a somewhat more favorable defensive profile, with Sharpe 0.864 versus 0.831 for buy-and-hold and maximum drawdown $-20.1\%$ versus $-24.8\%$. The most reasonable reading of these results is not that the Marketron drift provides a robust source of standalone alpha, but rather that it may serve as a regime-sensitive overlay that modestly improves downside behavior while preserving a large part of equity upside exposure.

The broader message of the project is therefore twofold. First, the Marketron is useful as an interpretable nonlinear state-space model capable of fitting higher-order return moments and organizing market dynamics into a regime-based framework. Second, when coupled with filtering, it can generate a trading signal whose main empirical value appears to lie in risk management rather than return maximization.

At the same time, the evidence remains limited. The out-of-sample period contains only 70 monthly observations, the strategy's Sharpe improvement is modest, and the calibrated model should be interpreted primarily as a conditional filtered system rather than as a realistic autonomous simulator. For these reasons, the results should be viewed as encouraging but preliminary. The most natural next steps would be walk-forward recalibration, comparison with stronger benchmark models, and richer strategy overlays that exploit the filtered state while accounting explicitly for uncertainty and execution constraints.

## References


I. Halperin and A. Itkin. Marketron games: Self-propelling stocks vs dumb money and metastable dynamics of the Good, Bad and Ugly markets. *arXiv:2501.12676v2*, 2025.

N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F, Radar and Signal Processing*, 140(2):107-113, 1993.

N. Hansen. The CMA evolution strategy: A comparing review. In J. A. Lozano et al., editors, *Towards a New Evolutionary Computation*, pages 75-102. Springer, 2006.

R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.