Aalto University
School of Science

Master's Programme in Mathematics and Operations Research

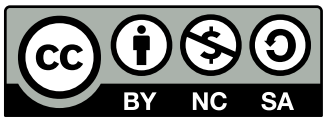
# Latent Efficient Price Recovery in Exchange-Traded Funds: A Simulation Study

**Kerkko Konola**

Master's Thesis
2026

© 2026

This work is licensed under a Creative Commons "Attribution-NonCommercial-ShareAlike 4.0 International" license.



Aalto University
School of Science

**Author** Kerkko Konola

**Title** Latent Efficient Price Recovery in Exchange-Traded Funds: A Simulation Study

**Degree programme** Mathematics and Operations Research

**Major** Systems and Operations Research

**Supervisor** Prof. Lasse Leskelä

**Advisor**

**Date** 25 May 2026 **Number of pages** 53 **Language** English

**Abstract**

Recovering the latent efficient price from limit order book data is a fundamental challenge in econometrics: the efficient price is never directly observed, only the noisy quotes and transactions that approximate it. This thesis studies whether statistical methods can reliably recover the latent price, and under what conditions one approach is likely to outperform another. The analysis is conducted in a simulation environment where the true efficient price is known by construction, allowing direct and unambiguous comparison of estimation accuracy.

The data generating process incorporates the empirically relevant features of market microstructure. A driftless Brownian motion drives the latent price, order flow follows a mean-reverting Ornstein–Uhlenbeck process with linear price impact, and microstructure noise is state-dependent and heteroskedastic. Two estimators are compared: a misspecified linear Kalman filter and a nonlinear XGBoost model. The comparison examines how the choice of estimator shapes the accuracy-robustness tradeoff across different market conditions.

In a single-ETF setting, the Kalman filter performs better when microstructure distortions are mild, reacting to directional price changes roughly 0.3 seconds faster and producing more stable estimates across 1 000 Monte Carlo replications. As distortions grow, however, XGBoost proves superior: in the high-noise regime it achieves up to 48% lower mean squared error by capturing nonlinear order flow patterns that the misspecified filter cannot represent.

The framework is then extended to a multi-ETF setting in which three funds with different liquidity profiles track the same underlying index. Adding cross-asset information improves XGBoost uniformly across all ETFs, with the least liquid instrument benefiting most. The Kalman filter responds asymmetrically. It improves for the less liquid ETFs but deteriorates for the most liquid one, a consequence of its uniform observation weighting. Notably, the two estimators converge to near-identical accuracy in the multi-ETF setting, not because linear filtering is fundamentally limited, but because the additional information reduces the penalty for misspecification.

Taken together, the results suggest that efficient price recoverability is regime- and specification-dependent, with implications for estimator selection and the design of cross-ETF statistical arbitrage strategies.

**Keywords** latent efficient price, market microstructure, Kalman filter, XGBoost, Monte Carlo simulation, exchange-traded funds, statistical arbitrage, price discovery

Aalto University
School of Science

**Tekijä** Kerkko Konola

**Työn nimi** Piilevän tehokkaan hinnan palauttaminen ETF-rahastoissa: simulaatiotutkimus

**Koulutusohjelma** Matematiikan ja operaatiotutkimuksen maisteriohjelma

**Pääaine** Systeemi- ja operaatiotutkimus

**Työn valvoja** Prof. Lasse Leskelä

**Työn ohjaaja**

**Päivämäärä** 25.5.2026 **Sivumäärä** 53 **Kieli** englanti

**Tiivistelmä**

Piilevän tehokkaan hinnan palautettavuus tarjouskirjadatasta on keskeinen haaste ekonometriassa: tehokasta hintaa ei koskaan havaita suoraan, ainoastaan sitä approksimoivat kohinaiset noteeraukset ja transaktiot. Tässä diplomityössä tutkitaan, pystyvätkö tilastolliset menetelmät luotettavasti palauttamaan tämän piilevän hinnan, ja miten eri lähestymistavat vertautuvat keskenään. Analyysi toteutetaan simulaatioympäristössä, jossa todellinen tehokas hinta on tunnettu, mikä mahdollistaa estimointitarkkuuden suoran ja yksiselitteisen vertailun.

Aineistoa generoiva stokastinen malli sisältää markkinan mikrorakenteen empiirisesti relevantit piirteet. Driftitön Brownin liike ohjaa piilevää hintaa, tilausvirta seuraa keskiarvoon palautuvaa Ornstein–Uhlenbeck-prosessia lineaarisella hintavaikutuksella, ja mikrorakennekohina on tilakohtainen ja heteroskedastinen. Vertailtavana on kaksi estimaattoria: virheellisesti määritelty lineaarinen Kalman-suodin ja epälineaarinen XGBoost-estimaattori. Näiden avulla voidaan tutkia, kuinka estimaattorin valinta muokkaa tarkkuuden ja vakauden välistä kompromissia eri markkinaolosuhteissa.

Yhden pörssinoteeratun rahaston (ETF) asetelmassa Kalman-suodin suoriutuu paremmin, kun mikrorakennehäiriöt ovat lieviä: se reagoi hinnan suunnanvaihdoksiin noin 0,3 sekuntia nopeammin ja tuottaa vakaampia estimaatteja 1 000 Monte Carlo -toiston yli. Häiriöiden kasvaessa XGBoost kuitenkin osoittautuu ylivoimaiseksi: korkean kohinan vallitessa se saavuttaa jopa 48 % pienemmän keskimääräisen neliövirheen hyödyntämällä epälineaarisia tilausvirran kuvioita, joita virheellisesti määritelty Kalman-suodin ei pysty mallintamaan.

Tutkimusta laajennetaan tämän jälkeen usean ETF:n asetelmaan, jossa kolme eri likviditeettiprofiilin rahastoa seuraa samaa kohde-etuutena olevaa indeksiä. Ristikkäisen informaation lisääminen parantaa XGBoostin suorituskykyä tasaisesti kaikissa ETF:issä, ja suurimmat hyödyt saavutetaan epälikvideimmälle instrumentille. Kalman-suodin reagoi epäsymmetrisesti: tarkkuus paranee vähemmän likvidien ETF:ien osalta, mutta heikkenee likvideimmän kohdalla yhtenäisen havaintojen painotuksen vuoksi. Huomionarvoista on, että molemmat estimaattorit suppenevat lähes identtiseen tarkkuuteen usean ETF:n asetelmassa – mutta ei siksi, että lineaarinen suodatus olisi perustavanlaatuisesti rajoittunut, vaan siksi, että lisäinformaatio pienentää määrittelyvirheen haittaa.

Yhdessä tulokset tukevat näkemystä, jonka mukaan tehokkaan hinnan palau-

tettavuus riippuu vallitsevasta markkinatilasta ja mallin määrittelystä, ja niillä on merkitystä estimaattorin valinnalle sekä ETF:ien välisen tilastoarbitraasin strategioiden suunnittelulle.

**Avainsanat** piilevä tehokas hinta, markkinoiden mikrorakenne, Kalman-suodin, XGBoost, Monte Carlo -simulaatio, ETF-rahastot, tilastoarbitraasi, hinnanmuodostus

# Contents

# Symbols and abbreviations

## Symbols

| | |
|---|---|
| $V_t$ | Latent efficient price (single-asset setting) |
| $F_t$ | Common latent factor shared across ETFs |
| $P_t$ | Observed transaction price |
| $P_{i,t}$ | Observed transaction price of ETF $i$ |
| $OF_t$ | Signed order flow |
| $X_t$ | Observable feature vector |
| $\widehat{V}_t$ | Estimated latent efficient price |
| $\widehat{F}_t$ | Estimated common latent factor |
| $W_t$ | Standard Brownian motion driving the latent price |
| $B_t$ | Standard Brownian motion driving the order flow |
| $\eta_t$ | Latent price innovation, $\eta_t \sim \mathcal{N}(0, \sigma^2 \Delta t)$ |
| $\xi_t$ | Order flow innovation, $\xi_t \sim \mathcal{N}(0, \sigma_{OF}^2)$ |
| $\varepsilon_t$ | Microstructure noise term |
| $\sigma$ | Latent price volatility parameter |
| $\sigma_0$ | Baseline microstructure noise level |
| $\sigma_{OF}$ | Std. of order flow innovations (discrete-time) |
| $\sigma_{OF}^c$ | Order flow diffusion coefficient (continuous-time OU) |
| $\theta$ | Order flow mean-reversion rate (continuous-time OU) |
| $\rho$ | Order flow persistence (discrete-time, $\rho = e^{-\theta \Delta t}$) |
| $\lambda$ | Price impact coefficient |
| $\alpha$ | Noise sensitivity to order flow intensity |
| $\Delta t$ | Simulation time step (one second) |
| $T$ | Number of time steps per simulated trading day |
| $N$ | Number of Monte Carlo replications |
| $\Delta V_t$ | First difference of the latent price, $V_t - V_{t-1}$ |
| $\mathcal{F}_t$ | Filtration generated by observable data up to time $t$ |
| $\mathbf{h}, \mathbf{R}$ | Observation vector and noise covariance (multi-ETF Kalman filter) |

## Operators

| | |
|---|---|
| $\mathbb{E}[\,\cdot\,]$ | Expectation operator |
| $\mathcal{N}(\mu, \sigma^2)$ | Normal distribution with mean $\mu$ and variance $\sigma^2$ |
| $\lvert\,\cdot\,\rvert$ | Absolute value |
| $\operatorname{sign}(\cdot)$ | Sign function |
| $\Delta$ | First-difference operator, $\Delta Z_t = Z_t - Z_{t-1}$ |

## Abbreviations

| | |
|---|---|
| ABM | Arithmetic Brownian Motion |
| AR(1) | First-order autoregressive process |
| ETF | Exchange-Traded Fund |
| FRED | Federal Reserve Economic Data |
| KF | Kalman filter |
| ML | Machine learning (XGBoost estimator) |
| MSE | Mean Squared Error |
| NAV | Net Asset Value |
| OU | Ornstein–Uhlenbeck process |
| SD | Standard Deviation |
| S&P 500 | Standard & Poor's 500 index |
| XGBoost | Extreme Gradient Boosting |

# 1 Introduction

Exchange-traded funds (ETFs) are investment vehicles whose shares trade continuously on an exchange, much like individual stocks, while the fund itself holds a portfolio of underlying assets [1]. Index-based ETFs—the focus of this thesis—aim to track a specific market index, and are built on a simple premise: the fund share price should closely track the intraday fair value of the underlying portfolio. In practice, this linkage depends on a functioning arbitrage mechanism and is subject to microstructure frictions that make it imperfect.

A key mechanism supporting the linkage between ETF prices and the value of the underlying assets is arbitrage. Authorised participants can create or redeem ETF shares in exchange for the underlying portfolio, which creates economic incentives to exploit deviations between the ETF price and its fair value [2]. In the absence of frictions, such arbitrage activity would enforce near-perfect price alignment. Empirical evidence shows that ETF prices can deviate substantially from their underlying value even in normal market conditions. Importantly, such deviations persist even after accounting for stale or noisy net asset value estimates, indicating that convergence is often delayed rather than instantaneous [3].

Such intraday deviations arise from market microstructure effects, limited liquidity, and frictions in the arbitrage mechanism. These effects can be particularly pronounced for less liquid ETFs or during periods of market stress, when order books thin and the cost of immediacy increases [4]. From a statistical perspective, the economically relevant price process is not directly observable: transaction prices reflect both the efficient price and transient microstructure distortions induced by order flow and bid-ask dynamics. Recovering the latent efficient price from these noisy observations is therefore a filtering problem.

This thesis studies the statistical recoverability of the latent efficient price from noisy limit order book observations using a fully controlled simulation environment, where the true efficient price is known to the researcher. The simulation approach is deliberate: direct evaluation of estimation accuracy is generally infeasible with empirical intraday data, where the true efficient price is unobservable. The focus is accordingly on statistical modeling and inference rather than on empirical measurement or trading strategy design.

The primary research question is:

> How accurately can the latent efficient price be recovered from noisy limit order book observations, and how does the answer depend on the degree of microstructure distortion and the choice of estimator?

Two estimators are compared: a misspecified linear Kalman filter, which serves as a structurally motivated benchmark, and a nonlinear XGBoost model, which approximates the optimal filter in a data-driven manner. The comparison is designed to identify the conditions under which the structural simplicity of the linear filter is an advantage and those under which the flexibility of the nonlinear estimator becomes decisive.

The thesis makes three contributions. It develops a structural simulation framework for studying efficient price recovery under realistic microstructure conditions: persistent order flow, linear price impact, and state-dependent heteroskedastic noise. It then quantifies the regime-dependent performance gap between linear and nonlinear filtering in a setting where ground truth is observable. Finally, it extends the framework to a multi-ETF setting in which three funds with heterogeneous liquidity profiles track the same underlying index, examining whether cross-asset price information improves efficient price recovery and whether the benefit is uniform across estimators.

The remainder of this thesis proceeds as follows. Section 2 describes the relevant market structure and background. Section 3 reviews the related literature. Section 4 introduces the model and methodology. Section 5 describes the simulation design. Section 6 reports the single-ETF results. Section 7 extends the framework to the multi-ETF setting. Section 8 reports the multi-ETF results. Section 9 discusses the findings and their implications, and Section 10 concludes.

# 2 Background and Market Structure

## 2.1 Exchange-Traded Funds: Structure and Trading

Exchange-traded funds combine features of traditional investment funds and exchange-traded securities: an ETF holds a portfolio of underlying assets, while its shares trade continuously on an exchange, much like individual stocks [1]. For index-based ETFs, the portfolio composition follows the constituents of a specified target index, held essentially fixed within a trading session and rebalanced only as the index itself is updated. This distinguishes ETFs from conventional open-end funds, whose shares are issued and redeemed at net asset value once per trading day.

ETF shares are traded in a limit order book, where prices are determined by the interaction of supply and demand among market participants. As a result, the observed ETF price at any given time reflects the state of the order book, including bid–ask spreads, available depth, and recent trades, rather than a direct calculation of the value of the underlying portfolio. Liquidity in ETF shares is therefore provided primarily by the market, not by the fund sponsor itself.

Because ETF trading is continuous, intraday prices may deviate from the contemporaneous value of the underlying assets. At short time scales, prices can be influenced by temporary order imbalances, liquidity fluctuations, and market-wide conditions affecting the cost of immediacy. These features imply that ETF prices observed at high frequency need not coincide with any notion of intraday fair value, even for ETFs tracking broad and liquid indices.

While the ETF universe includes a wide range of products, such as thematic and actively managed funds, this thesis focuses on index-based ETFs: passive funds whose holdings are constructed to replicate the constituents and weights of a specified market index, with the explicit objective of matching the index's performance. For such funds, the underlying index serves as a natural reference for the latent efficient price.

## 2.2 Net Asset Value, Intraday NAV, and Fair Value Proxies

The net asset value (NAV) of an ETF share is defined as the total market value of the fund's underlying assets minus liabilities—primarily accrued management fees and operating expenses—divided by the number of shares outstanding [5]. For index-based ETFs, the NAV is typically computed using closing prices of the underlying securities and is published once per trading day. As such, the reported NAV reflects an end-of-day valuation rather than a contemporaneous measure of the fund's value during the trading session.

For intraday analysis, the daily NAV is therefore insufficient as a benchmark. Instead, market participants rely on intraday estimates of the fund's value, commonly referred to as intraday net asset value (iNAV). The iNAV is constructed by updating the valuation of the underlying portfolio using real-time or high-frequency price information for the constituent assets, while holding the portfolio weights fixed. In practice, iNAV is disseminated at regular intervals (e.g., every 15 seconds) by exchanges or data providers [6], although its accuracy depends on the liquidity and

trading hours of the underlying assets.

While iNAV is widely used as a real-time reference for ETF valuation, it inherits microstructure noise from the underlying constituents, including bid–ask spreads, non-synchronous trading, and liquidity fluctuations [2]. An alternative reference is the index-based proxy: the published value of the target index, computed by the index provider from constituent prices using fixed aggregation rules. As summarised in Table 1, this proxy is less exposed to ETF-specific frictions and can therefore serve as a cleaner benchmark for intraday fair value. Either way, any observable proxy for the latent efficient price is subject to measurement noise.

**Table 1:** Sources of microstructure noise in intraday fair value proxies

| Feature | iNAV | Index-based proxy |
|---|---|---|
| Price source | Last trades / quotes | Rule-based aggregation |
| Update mechanism | Discrete (e.g. every 15s) | Continuous or carry-forward |
| Handling of stale prices | Implicit via last observation | Explicit smoothing rules |
| Exposure to bid–ask bounce | Direct | Limited |
| Sensitivity to order flow | Indirect via constituents | Indirect |

Table 1 summarizes the main sources of microstructure noise affecting intraday fair value proxies. While both index-based proxies and iNAV aim to reflect the same underlying portfolio value, they differ in how prices are aggregated and updated. In particular, iNAV inherits short-term fluctuations from last-trade prices and discrete update mechanisms, whereas index values are typically constructed using smoothing rules that mitigate bid–ask bounce and non-synchronous trading effects. These differences illustrate why any observable proxy for the latent efficient price is subject to measurement noise.

In market microstructure terms, the latent efficient price (also referred to as fair value) is the price that would prevail in a frictionless market: it reflects all available information about the contemporaneous value of the underlying assets while excluding the transient effects of bid–ask spreads, order imbalances, and other microstructure distortions. In practice it is never directly observed; what is observed are noisy proxies such as iNAV, index values, and transaction prices, each of which approximates the efficient price subject to measurement noise. In this thesis, fair value is therefore treated as a latent process that is unobservable in practice; the simulation environment sidesteps the proxy measurement problem by making the true efficient price known by construction.

## 2.3 Creation and Redemption Mechanism

A defining feature of exchange-traded funds is the creation and redemption mechanism, which links the ETF share price to the value of the underlying portfolio through arbitrage by authorized participants (APs). Unlike closed-end funds, ETFs allow shares outstanding to expand or contract through in-kind transactions with a limited set of institutional market participants [2].

When the ETF bid price exceeds the cost of acquiring the underlying basket at constituent ask prices, APs can profitably engage in the creation process: they purchase the constituent securities, deliver the basket to the fund sponsor in exchange for newly created ETF shares, and sell those shares in the secondary market. Conversely, when the ETF ask price falls below the proceeds from liquidating the basket at constituent bid prices, APs can acquire ETF shares in the market, redeem them for the underlying securities, and sell the constituents [2]. In frictionless markets these conditions collapse to a single price on each side, and the arbitrage mechanism is expected to enforce a tight linkage between the ETF price and the fair value of its underlying portfolio.

In practice, however, the effectiveness of the creation and redemption mechanism is limited by several frictions. These include transaction costs in the underlying securities, bid–ask spreads, market impact, inventory constraints, and operational delays associated with the creation and redemption process. Moreover, APs face intraday risk when executing arbitrage trades, as the prices of the underlying assets and the ETF may move differently during the execution window. Furthermore, these transactions occur in discrete blocks called creation units, typically comprising tens of thousands of shares, which sets a minimum threshold for arbitrage activity and limits the mechanism's responsiveness to small or short-lived deviations [2]. Taken together, these frictions imply that arbitrage is neither instantaneous nor costless, allowing short-term price deviations to persist even in normal market conditions.

The creation and redemption mechanism operates in the primary market, the channel through which APs transact directly with the fund sponsor to issue or redeem shares. Most ETF trading, by contrast, occurs in the secondary market: the exchange order book where investors buy and sell existing ETF shares without involving the fund itself. Short-lived imbalances in order flow or information arrival can therefore generate intraday price deviations that arbitrage does not immediately offset. This distinction between primary and secondary market dynamics is central to understanding intraday mispricing behavior.

The creation and redemption mechanism thus acts as a mean-reverting force on ETF mispricing, while microstructure frictions and liquidity dynamics generate short-term noise and persistence. Intraday mispricing is better understood as the outcome of this tension than as a pricing error that is swiftly eliminated.

## 2.4 Sources of Intraday Mispricing

Intraday mispricing in exchange-traded funds arises from the interaction of market microstructure effects, liquidity frictions, and limits to arbitrage. Although the creation and redemption mechanism provides a structural linkage between ETF prices and the value of the underlying portfolio, this mechanism operates with delays and costs that allow short-term price deviations to emerge and persist [3].

A primary source of intraday mispricing is liquidity imbalance in the secondary market. ETF trading is often driven by order flow associated with allocation and rebalancing decisions rather than contemporaneous information about the underlying assets [7]. Temporary imbalances between buying and selling pressure can move ETF

prices away from contemporaneous fair value, particularly when order books are thin or when liquidity providers widen spreads in response to heightened uncertainty [4].

Bid–ask bounce, discrete price updates, and non-synchronous trading in the underlying securities introduce short-term noise into both ETF prices and fair value proxies. Even when an index-based proxy is used, differences in update frequencies and price formation across assets imply that observed price discrepancies partly reflect measurement error rather than economically meaningful mispricing.

Limits to arbitrage play a central role in shaping the dynamics of intraday mispricing. Authorized participants face transaction costs, inventory constraints, and execution risk when engaging in creation and redemption activity. Moreover, arbitrage trades expose APs to intraday price risk during the execution window, particularly in volatile market conditions. As a result, arbitrage capital is deployed selectively and responds to sufficiently large and persistent deviations, rather than eliminating small or short-lived price discrepancies instantaneously [3].

Finally, intraday mispricing may exhibit temporal dependence due to volatility clustering, liquidity cycles, and predictable intraday patterns in trading activity. Periods of elevated volatility or reduced market depth can amplify both the magnitude and persistence of deviations, leading to regime-like behavior in mispricing dynamics [8].

Taken together, these mechanisms suggest that intraday mispricing is a latent, stochastic process shaped by the interplay of transitory market frictions and stabilizing arbitrage forces. Observed price deviations therefore reflect a combination of genuine mispricing dynamics and measurement noise.

# 3 Literature Review

This section reviews the existing literature related to ETF mispricing, intraday price formation, market microstructure, and limits to arbitrage. The purpose is to position the present thesis within the broader academic context and to identify gaps that motivate the modeling approach adopted in this study.

## 3.1 ETF Mispricing and Arbitrage Mechanisms

A central theme in the ETF literature is the extent to which ETF share prices deviate from the value of the underlying portfolio and the mechanisms that govern the correction of such deviations. Early empirical studies document that ETF shares frequently trade at premiums or discounts relative to their net asset value, even for funds tracking broad and liquid indices [9]. While these deviations are often small in magnitude, they can be economically meaningful, particularly for investors trading at high frequency.

The primary institutional mechanism linking ETF prices to the value of the underlying assets is the creation and redemption process. Authorized participants can exchange ETF shares for the underlying basket of securities in the primary market, creating incentives to arbitrage deviations between ETF prices and their fair value. In frictionless markets, this mechanism would imply that ETF mispricing is short-lived and tightly bounded. However, empirical evidence suggests that the effectiveness of arbitrage is limited by transaction costs, inventory constraints, execution risk, and operational frictions [2]. In addition, the authors note that convergence is not guaranteed within any finite time horizon, so arbitrage operates neither frictionlessly nor instantaneously.

A growing body of literature shows that ETF mispricing can persist even under normal market conditions. Petäjistö [3] documents systematic and persistent deviations between ETF prices and the value of the underlying assets, arguing that limits to arbitrage play a central role in explaining these patterns. Importantly, mispricing does not vanish instantaneously once it becomes detectable, suggesting that arbitrage activity responds with delays and only to sufficiently large or persistent deviations. This evidence challenges the view of ETF pricing as mechanically enforced by continuous arbitrage and highlights the role of market frictions in shaping price dynamics.

Most empirical studies in this literature rely on daily observations of ETF prices and net asset values, or on cross-sectional variation across funds and time. While such approaches are well suited for documenting the existence and average magnitude of mispricing, they provide limited insight into how mispricing evolves within the trading day. Studies that employ intraday data, such as Marshall, Nguyen and Visaltanachoti (2013) [4], typically focus on discrete arbitrage events or liquidity episodes rather than on the continuous-time dynamics of price deviations. As a result, relatively little is known about the short-term persistence, mean reversion, and state dependence of ETF mispricing at intraday horizons.

Overall, the evidence establishes that ETF mispricing is a robust and economically relevant phenomenon, driven by the interaction of arbitrage mechanisms and market

frictions. The predominant focus on daily data, however, leaves open important questions about how mispricing evolves within the trading day.

## 3.2 Intraday Pricing, iNAV, and Fair Value Measurement

Intraday valuation of exchange-traded funds presents fundamental measurement challenges. While the net asset value provides a well-defined benchmark at the end of the trading day, it is not informative about the contemporaneous value of the underlying portfolio during the trading session.

The most commonly used intraday valuation metric is the intraday net asset value (iNAV), which is constructed by updating the valuation of the ETF's underlying portfolio using real-time or high-frequency price information for the constituent securities, while keeping portfolio weights fixed. iNAV is typically disseminated at regular intervals, such as every 15 seconds, and is widely used by market participants as a reference for ETF pricing and arbitrage activity. Its appeal lies in its transparency and its close alignment with the ETF's portfolio composition.

Despite its widespread use, iNAV is subject to several sources of measurement error. Since iNAV relies on observed transaction prices or quotes of the underlying assets, it inherits microstructure noise arising from bid–ask bounce, discrete price updates, and non-synchronous trading across securities [2]. These effects can be particularly pronounced when underlying assets are illiquid or trade infrequently, or when the trading hours of the underlying market do not fully overlap with those of the ETF. Consequently, short-term deviations between ETF prices and iNAV may partly reflect noise in the valuation proxy rather than economically meaningful mispricing.

This distinction between observed prices and economically relevant value is central to the market microstructure literature. In this framework, transaction prices are commonly modeled as noisy realizations of an unobserved efficient price, with transitory components arising from bid–ask bounce, discreteness, and short-term liquidity effects [10]. From this perspective, both ETF prices and intraday valuation measures such as iNAV should be viewed as imperfect signals of an underlying latent value rather than as direct observations of fair value.

An alternative approach to intraday fair value measurement is to use the level of the underlying index tracked by the ETF, suitably scaled to the ETF share value. Index values are typically constructed using rule-based aggregation and smoothing procedures that mitigate some forms of microstructure noise, including bid–ask bounce and stale prices. While index-based proxies are not immune to measurement error, they are less directly affected by ETF-specific order flow and secondary market liquidity conditions. This makes them a potentially useful benchmark for analyzing intraday deviations in ETF prices, particularly when the objective is to study mispricing dynamics rather than the mechanical implementation of arbitrage.

The literature emphasizes that any intraday fair value measure should be interpreted as a noisy proxy for an underlying latent value. Petäjistö [3], for example, adopts a novel relative pricing approach in which the fair value of an ETF is inferred from a basket of funds tracking the same underlying index, thereby isolating a common value component from idiosyncratic pricing noise. This perspective reinforces the view that

fair value is not directly observable and that observed price discrepancies reflect a combination of true mispricing and measurement error.

The intraday fair value literature makes clear that the economically relevant price of an ETF cannot be directly observed. Any analysis of intraday pricing must therefore account for the latent nature of the efficient price, and statistical filtering is a natural response to this challenge.

## 3.3 Market Microstructure and Order Book Effects in ETFs

Unlike traditional mutual funds, ETF shares trade continuously in a limit order book, where prices are determined by the interaction of buy and sell orders submitted by market participants. As a result, intraday ETF prices are shaped not only by information about the underlying assets, but also by the state of the order book, including available liquidity, bid–ask spreads, and the distribution of depth across price levels.

Liquidity provision in ETF markets follows the same fundamental principles as in equity markets. Market makers and other liquidity suppliers post limit orders and adjust quoted spreads and depth in response to order flow, volatility, and inventory considerations. When liquidity is abundant, ETF prices can adjust smoothly to new information and arbitrage pressures. During periods of reduced liquidity or heightened uncertainty, however, spreads may widen and order book depth may thin, increasing the cost of immediacy and amplifying short-term price deviations [10].

Order flow imbalances in the secondary market can generate temporary price pressure that moves ETF prices away from contemporaneous fair value. Such effects are well documented in the market microstructure literature, where signed order flow is shown to have a transitory impact on prices before being partially reversed as liquidity is replenished [10]. In the ETF context, these imbalances may arise from portfolio rebalancing, hedging demand, or short-term trading strategies that are unrelated to changes in the value of the underlying assets.

Empirical studies provide evidence that ETF pricing is sensitive to intraday liquidity conditions and order book dynamics. Marshall, Nguyen, and Visaltanachoti [4] document that ETF spreads, depth, and trading activity vary substantially within the trading day and are closely linked to deviations between ETF prices and their underlying value. Their findings suggest that intraday mispricing is often associated with temporary liquidity shortages and elevated trading costs rather than with fundamental valuation errors.

An important distinction emphasized in the literature is that most ETF trading occurs in the secondary market, while the creation and redemption mechanism operates through the primary market. As noted by Ben-David, Franzoni, and Moussawi [2], short-term imbalances in secondary market order flow are not immediately offset by arbitrage activity, due to transaction costs, execution risk, and the discrete nature of creation units.

The microstructure literature thus highlights that intraday ETF prices reflect a dynamic interaction between liquidity provision, order flow, and arbitrage constraints,

with order book conditions playing a central role even when longer-term arbitrage forces induce mean reversion.

## 3.4 Limits to Arbitrage and Dynamic Adjustment

While arbitrage plays a central role in linking ETF prices to the value of the underlying portfolio, the literature emphasizes that arbitrage is neither costless nor instantaneous. Authorized participants and other arbitrageurs face transaction costs in both the ETF and the underlying securities, as well as bid–ask spreads, market impact, and operational frictions associated with the creation and redemption process. In addition, arbitrage trades expose participants to execution risk and intraday price risk during the period over which positions are established and unwound. These considerations imply that arbitrage capital is deployed selectively rather than continuously [2].

Inventory constraints and risk management considerations further limit the responsiveness of arbitrage activity. Market makers and authorized participants typically manage inventories subject to capital constraints and risk limits, which can reduce their willingness to absorb large or persistent imbalances in order flow. As a result, even economically detectable deviations between ETF prices and fair value may persist when the expected profits from arbitrage are insufficient to compensate for risk and transaction costs [3].

A recurring theme in the literature is that arbitrage exhibits threshold-like behavior. Small or short-lived price deviations may not trigger arbitrage activity, while larger or more persistent deviations eventually attract capital and induce corrective trading. This nonlinearity implies that price adjustment occurs with delays and that convergence toward fair value is gradual rather than immediate. Empirical evidence shows that ETF mispricing often displays mean-reverting behavior over short horizons, but with reversion speeds that depend on prevailing liquidity conditions and trading costs [3].

Limits to arbitrage thus transform arbitrage from a static pricing condition into a stabilizing force that operates over time. Rather than enforcing instantaneous price equality, arbitrage activity induces mean reversion in mispricing, counteracted by short-term price pressure from order flow and liquidity fluctuations [10].

The interaction between transient market frictions and delayed arbitrage response has important implications for intraday mispricing persistence. Mispricing can persist when liquidity is scarce or when arbitrage capacity is temporarily constrained, and it can dissipate more rapidly when trading costs are low and market depth is abundant. Mispricing therefore evolves as a friction-driven adjustment process, with reversion speeds that vary across time and market conditions rather than being uniform.

## 3.5 Order Book Modeling Approaches

The limit order book has been modeled using a variety of approaches in the market microstructure literature, reflecting the complexity of order-driven price formation. Existing models differ in their level of structural detail and in the extent to which they aim to represent individual order book events versus aggregate liquidity conditions.

A prominent class of models adopts a reduced-form perspective, in which the state of the order book is summarized by observable liquidity measures such as bid–ask spreads, available depth, and order flow imbalance. In this framework, price dynamics are modeled directly as a function of these variables, often using linear or vector autoregressive specifications [10]. Such approaches are well suited for empirical analysis, requiring modest data inputs and offering a transparent interpretation of liquidity effects on prices.

An alternative strand of the literature models order book dynamics at the event level using stochastic point processes. In these models, limit order arrivals, cancellations, and trades are treated as random events whose intensities depend on past order flow and the current state of the book. Hawkes process models have been particularly influential in this context, as they capture self-exciting behavior and clustering in order flow [11]. The key idea underlying Hawkes processes is that the occurrence of an event temporarily increases the likelihood of subsequent events, reflecting the endogenous feedback mechanisms observed in order-driven markets. While these models provide a detailed description of event-level order book dynamics, their complexity and focus on high-frequency events can make them less suitable for analyzing higher-level price deviations and mispricing dynamics.

Queueing-based models represent another approach, viewing the order book as a set of interacting queues whose depletion leads to price changes [12]. In this framework, price movements arise mechanically when the best bid or ask queue is exhausted, rather than from explicit information shocks. These models emphasize the mechanical aspects of price formation arising from order execution and cancellation dynamics, and have been used to study short-term price movements and liquidity resilience.

Finally, agent-based models generate order book dynamics endogenously by simulating the interaction of heterogeneous trading agents with simple behavioral rules. A seminal contribution in this literature is Farmer, Patelli, and Zovko [13], who show that many stylized facts of financial markets, including bid–ask spreads, volatility clustering, and price impact, can emerge even when agents follow zero-intelligence trading strategies. These models emphasize the role of market structure and order book mechanics in shaping price dynamics. However, because agent-based models rely primarily on simulation and calibration, they are typically used to study qualitative mechanisms and emergent behavior rather than for direct statistical inference on observed price processes.

In the context of this thesis, the order book is not modeled at the level of individual events. Instead, a reduced-form data generating process is used in which order flow captures the aggregate effect of trading pressure on observed prices. This approach captures the essential microstructure mechanisms identified in the literature without sacrificing tractability in a simulation setting.

## 3.6 State-Space Filtering and Machine Learning in Price Discovery

A parallel strand of the literature, rooted in market microstructure theory, models transaction prices as noisy realizations of an unobserved efficient price. The influential model of Roll [14] decomposes observed price changes into a random walk component reflecting fundamental value and a transitory component arising from bid–ask bounce. This decomposition formalizes the idea that the econometrically relevant price process is latent, and that statistical methods are required to separate it from microstructure noise.

The state-space framework provides a natural setting for efficient price estimation. In a linear Gaussian state-space model, the latent efficient price evolves according to a stochastic process and is observed only through noisy transaction prices. The Kalman filter [15] produces the minimum mean squared error estimate of the latent state given the history of observations. Its appeal in financial applications lies in its computational tractability, its recursive update structure, and its optimality under linear-Gaussian assumptions.

In practice, however, the Kalman filter is applied to settings that violate these assumptions. Transaction prices exhibit heteroskedastic noise, nonlinear price impact, and time-varying dynamics that are not captured by a linear observation equation. When the model is misspecified, the Kalman filter remains computationally convenient but is no longer optimal, and the magnitude of the resulting estimation error depends on the degree of nonlinearity and the strength of microstructure distortions.

Machine learning methods have emerged as a data-driven alternative for price discovery in high-frequency settings. Gradient-boosted decision trees, in particular the XGBoost algorithm of Chen and Guestrin [16], offer a flexible nonlinear framework that approximates complex relationships between observable features and latent values without parametric assumptions. In the microstructure context, Sirignano and Cont [17] demonstrate that deep learning models trained on limit order book data capture universal features of price formation across markets and instruments, suggesting that data-driven approaches can extract systematic patterns that elude linear models.

Despite growing interest in machine learning for financial applications, comparisons between structural linear filters and nonlinear machine learning estimators in the context of latent efficient price recovery remain limited. Most existing work evaluates machine learning models on predictive tasks using empirical data, where the true efficient price is unobservable and estimation accuracy cannot be assessed directly. This thesis addresses this gap by conducting the comparison in a simulation environment where ground truth is known, allowing the regime-dependent performance of the two approaches to be evaluated precisely.

## 3.7 Gaps in the Literature and Contribution of This Thesis

The existing literature provides extensive evidence that ETF prices can deviate from the value of the underlying portfolio and that such deviations are shaped by market microstructure frictions and limits to arbitrage. At the same time, several gaps remain

that motivate the approach taken in this thesis.

Much of the empirical literature focuses on daily observations or cross-sectional variation across funds and time. These approaches establish the existence and persistence of mispricing but provide limited insight into the short-term dynamics through which price deviations arise. In particular, the question of how accurately the latent efficient price can be recovered from noisy intraday observations has received comparatively little attention in the ETF context.

A further obstacle is that the true latent efficient price is unobservable in empirical work. Observed deviations between ETF prices and fair value proxies reflect a mixture of true mispricing and measurement noise, making it difficult to evaluate any filtering procedure directly. This identification problem limits the conclusions that can be drawn about the quality of efficient price recovery.

Finally, while the microstructure literature has compared linear state-space filters with nonlinear alternatives in equity settings, such comparisons are scarce in the ETF context, where the presence of a cross-listed underlying index introduces additional structure that multi-asset estimators may exploit.

This thesis addresses these gaps through a controlled simulation study. By constructing a data generating process in which the true latent efficient price is known by construction, the thesis enables direct evaluation of estimation accuracy under varying microstructure conditions. The comparison between a misspecified linear Kalman filter and a nonlinear XGBoost estimator quantifies the regime-dependent costs of model misspecification and the conditions under which nonlinear flexibility becomes advantageous. The multi-ETF extension further examines whether pooling information across funds tracking the same index improves efficient price recovery, and whether the benefit is uniform across estimators with different information aggregation mechanisms.

The contribution of this thesis is therefore methodological: it provides a replicable simulation framework for studying efficient price recovery under realistic microstructure noise, and produces regime-specific and specification-sensitive results that are not obtainable from empirical data alone.

# 4 Model and Methodology

## 4.1 Research Objective

The primary objective of this thesis is to investigate the statistical recoverability of a latent efficient price from noisy limit order book observations. The analysis is conducted in a controlled simulation environment where the true efficient price is known to the researcher, enabling direct evaluation of estimation accuracy. This design is deliberate: such evaluation is generally infeasible with empirical market data, where the true price is unobservable. The focus is accordingly on statistical modeling rather than on empirical measurement or trading strategy design.

## 4.2 Conceptual Framework

The model is structured as a three-layer system. The first layer is the structural truth layer: a latent efficient price that evolves stochastically and is not directly observable. The second layer is the observation layer: a limit order book that provides a noisy proxy for the true value. The third layer is the estimation layer: a statistical procedure that uses the observed order book to approximate the latent price. The mathematical specification of each layer is introduced through the Data Generating Process in the following section.

### 4.2.1 Layer 1: Structural Truth Layer

Let $V_t$ denote the latent efficient price process, representing the fundamental value of the asset that is not directly observable.

The latent price is modeled as a driftless arithmetic Brownian motion (ABM):

$$dV_t = \sigma \, dW_t$$

where $W_t$ is a standard Brownian motion and $\sigma > 0$ is a constant volatility parameter. Under this specification, price increments are additive, Gaussian, and independent of the current price level. This differs from geometric Brownian motion, under which volatility scales proportionally with price and log-prices follow a Brownian motion.

ABM is chosen over GBM on both theoretical and practical grounds. First, it keeps the latent-state dynamics linear and Gaussian, which is the natural setting for Kalman filtering and yields a clean benchmark for evaluating recoverability. Additionally, any estimation error can be attributed to the observation process and modeling assumptions rather than to nonlinearity in the state equation. Finally, at intraday horizons, the practical difference between additive and multiplicative price dynamics is typically small for reasonable calibrations, while the simpler additive specification offers ease of computation and analysis.

For simulation purposes, the continuous process is discretized as

$$V_t = V_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma^2 \Delta t)$$

Since $\eta_t$ is mean-zero and independent of past information, the latent price satisfies

$$\mathbb{E}[V_t \mid \mathcal{F}_{t-1}] = V_{t-1}$$

where $\mathcal{F}_{t-1}$ denotes the filtration generated by the latent process up to time $t-1$. Thus, the latent price is a martingale with independent increments. The estimation problem is therefore one of filtering: no predictable component is embedded in the true price dynamics, so any recoverable signal must come from the observation layer.

The volatility parameter $\sigma$ is calibrated to match empirical intraday volatility in major equity indices; the calibration procedure is described in Section 5.

#### 4.2.2 Layer 2: Observation Layer (Microstructure)

The econometrician does not observe the latent efficient price $V_t$ directly. Instead, the observable data consists of features derived from a simulated limit order book.

Let $X_t \in \mathbb{R}^k$ denote the vector of observable order book features at time $t$. The observation process depends on both $V_t$ and microstructure noise $\varepsilon_t$ through a possibly nonlinear transformation. The noise process $\varepsilon_t$ is assumed to be independent of the latent innovation $\eta_t$ and may include both additive and state-dependent components. The feature vector $X_t$ could include, for example, the midprice, bid–ask spread, order flow, traded volume, order imbalance, and a volatility proxy. The observation layer thus introduces microstructure noise, price impact effects, and potentially nonlinear distortions that make the efficient price difficult to recover directly.

Let $\mathcal{F}_t^X$ denote the filtration generated by the observable feature process $\{X_s\}_{s \le t}$. Because the observation process may be nonlinear and the noise structure may deviate from Gaussian assumptions, the optimal filter need not be linear. This motivates the comparison between classical linear state-space filtering and nonlinear machine learning methods in the estimation layer.

#### 4.2.3 Layer 3: Estimation Layer

The estimation layer aims to recover the latent efficient price $V_t$ from observable data. In principle, the object of interest is the conditional expectation

$$\mathbb{E}[V_t \mid \mathcal{F}_t^X]$$

where $\mathcal{F}_t^X$ denotes the filtration generated by the observable feature process. The task is therefore one of filtering: the latent state must be inferred from noisy and potentially distorted observations.

Two different estimation approaches are considered. The first is a linear state-space filter, which provides a simple structural benchmark. The second is a nonlinear machine learning model, which learns the relationship between observable features and the latent price directly from simulated data.

The linear benchmark is the Kalman filter. Under a correctly specified linear-Gaussian state-space model, the Kalman filter gives the minimum mean squared error estimator within that class of models. In the present setting, however, it is intentionally

misspecified. The filter treats the observed transaction price as a simple noisy proxy for the latent price, assumes a constant observation variance, and ignores the price impact term $\lambda OF_t$. Its role is therefore not to provide a fully specified optimal filter for the true data generating process, but to serve as a clear and interpretable structural benchmark.

The nonlinear approach is implemented using XGBoost. Instead of relying on an explicit state-space specification, it estimates a flexible nonlinear mapping from observable features to the latent efficient price. In the implementation used here, the model acts as a static predictor: at each time $t$, it maps the current feature vector $X_t$ to an estimate of $V_t$ without recursive updating. This makes it capable of capturing interactions and nonlinear patterns that lie outside the scope of a linear filter.

The comparison between the two approaches is meant to show whether the main limitation of the Kalman filter is its misspecification, and whether a more flexible nonlinear estimator can recover the latent price more accurately, especially across different microstructure regimes.

## 4.3 Data Generating Process

This section presents the Data Generating Process used in the simulation study. The simulation model consists of a latent efficient price, persistent order flow, linear price impact, and state-dependent microstructure noise.

### 4.3.1 Latent Efficient Price

As in the structural truth layer, the latent efficient price is modeled as a discretized driftless arithmetic Brownian motion:

$$V_t = V_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma^2 \Delta t)$$

Since $\eta_t$ is mean-zero and independent of past information, $V_t$ is a martingale with independent increments. This process defines the unobservable benchmark that the estimators seek to recover from the noisy observation layer.

### 4.3.2 Order Flow Process

Signed order flow is modeled as a stationary Ornstein–Uhlenbeck (OU) process,

$$dOF_t = -\theta\, OF_t\, dt + \sigma^c_{OF}\, dB_t, \qquad \theta > 0,$$

where $B_t$ is a standard Brownian motion independent of $W_t$. The mean-reversion rate $\theta$ controls the persistence of order flow, and the diffusion coefficient $\sigma^c_{OF}$ governs the intensity of incoming information. The process is stationary with unconditional mean zero and variance $\sigma^{c\,2}_{OF}/(2\theta)$, reverting toward zero on a time scale of $1/\theta$. Persistence (small $\theta$) reflects the empirically documented pattern of order splitting, in which an institutional investor executes a large trade through a sequence of smaller orders over time.

For simulation on a fixed grid with step $\Delta t$, the OU process admits an exact discretization. Defining $\rho = e^{-\theta\,\Delta t}$, the process satisfies

$$OF_{t+\Delta t} = \rho\, OF_t + \xi_t, \qquad \xi_t \sim \mathcal{N}\left(0,\ \frac{\sigma_{OF}^{c\,2}(1-\rho^2)}{2\theta}\right),$$

which has the form of a stationary AR(1) process with autoregressive coefficient $\rho$ and innovation variance $\sigma_{OF}^2 \equiv \sigma_{OF}^{c\,2}(1-\rho^2)/(2\theta)$. The recursion is not an Euler approximation but the exact distribution of the OU process sampled on the discrete grid, and it is the form used to advance order flow in the simulation.

### 4.3.3 Price Impact Mechanism

The observed transaction price combines the latent efficient price, linear price impact, and microstructure noise:

$$P_t = V_t + \lambda OF_t + \varepsilon_t$$

The parameter $\lambda$ measures the sensitivity of the observed price to order flow: buying pressure pushes $P_t$ above $V_t$, selling pressure below it.

### 4.3.4 Order-Flow-Dependent Microstructure Noise

Microstructure noise is modeled as conditionally Gaussian with order-flow-dependent variance:

$$\varepsilon_t \sim \mathcal{N}\Big(0,\ \sigma_0^2 + \alpha|OF_t|\Big)$$

where $\sigma_0^2 > 0$ is the baseline noise level and $\alpha \geq 0$ scales how much noise grows with order flow intensity. High trading activity thus widens the effective spread and increases price uncertainty. The linear-magnitude form $\alpha|OF_t|$ is chosen over a quadratic alternative $\alpha OF_t^2$ because it preserves a direct interpretation of $\alpha$ as the marginal noise variance per unit of absolute order flow, making the parameter straightforward to calibrate and reason about.

### 4.3.5 Observable Feature Vector

The observable feature vector used for estimation is

$$X_t = \begin{pmatrix} P_t \\ OF_t \\ |OF_t| \\ \Delta P_t \end{pmatrix}$$

where $\Delta P_t = P_t - P_{t-1}$. The four components capture the current price level, signed order flow, order flow intensity, and the most recent price change. Although $|OF_t|$ is implied by $OF_t$, it is included explicitly to help the tree-based estimator capture the symmetric effect of order flow intensity on noise variance without relying on learned sign symmetry.

### 4.3.6 Structural Properties

From an estimation standpoint, the key feature of the data generating process is that observed prices depend on order flow in two distinct ways: through the price impact term and through the variance of microstructure noise. As a result, the observation process does not match the simplified linear-Gaussian state-space model used by the Kalman filter. A nonlinear estimator with access to the full feature vector $X_t$ can in principle exploit this additional structure more effectively than the Kalman benchmark.

## 4.4 Evaluation Framework

Because the latent efficient price $V_t$ is known in simulation, estimation performance can be evaluated directly. The primary metric is mean squared error, computed on the test sample:

$$\text{MSE} = \frac{1}{T_{\text{test}}} \sum_{t \in \text{test}} (\widehat{V}_t - V_t)^2$$

A secondary metric captures response lag. A directional change in $V_t$ is defined as a reversal in the sign of the first difference $\Delta V_t = V_t - V_{t-1}$ relative to $\Delta V_{t-1}$, with both differences nonzero. For each such event, the lag measures how many steps the estimator takes to register a matching reversal in $\Delta\widehat{V}_t$, that is, a directional change whose new sign agrees with that of the true event. The reported value is the mean lag across all true events for which a matching reversal is found within a fixed look-ahead horizon; events without a matching reversal are excluded from the average. Smaller lag therefore indicates faster tracking of the latent price's turning points.

## 4.5 Methodological Contribution

The contribution of the thesis is methodological rather than empirical. It develops a controlled simulation framework for studying how well a latent efficient price can be recovered in the presence of microstructure distortions. Within this framework, the thesis compares a classical linear filtering approach with a nonlinear machine learning estimator and studies how their performance changes across different microstructure regimes and under misspecification.

The analysis proceeds in two stages. The baseline setting, described in Sections 5 and 6, considers a single ETF. Section 7 extends the setup to a multi-ETF case in which several funds track the same underlying index, and examines whether cross-asset information improves recovery of the latent price.

Although the study is simulation-based, the underlying structure is motivated by standard market microstructure mechanisms. This allows the results to be interpreted economically while retaining full control over the latent data-generating process.

# 5 Simulation Design

## 5.1 Simulation Structure

The simulation operates in discrete time at one-second frequency, so a single path spans one full trading day of $T = 23\,400$ steps. Performance estimates are based on $N = 1000$ independently simulated paths per microstructure regime, with all reported metrics averaged across replications.

## 5.2 Parameter Specification

The structural parameters are held fixed across all regimes. Table 2 reports the baseline values.

**Table 2:** Baseline simulation parameters

| Parameter | Value | Description |
|---|---|---|
| $V_0$ | 100.0 | Initial efficient price level |
| $\sigma$ | 0.004968 | Efficient price volatility (calibrated, see Section 5.3) |
| $\rho$ | 0.7 | Order flow persistence |
| $\sigma_{OF}$ | 1.0 | Std. of order flow innovations |
| $\sigma_0$ | 0.001242 | Baseline microstructure noise |

The volatility $\sigma$ is calibrated from S&P 500 daily data and scaled to the intraday frequency used in the simulation, as described in Section 5.3. The baseline noise $\sigma_0$ is set to approximately 25% of $\sigma$, so that observed prices contain meaningful microstructure noise without overwhelming the latent signal. The order flow parameters $\rho$ and $\sigma_{OF}$ are the exact one-second discretization of an Ornstein–Uhlenbeck process (Section 4.3.2); with $\Delta t = 1$ second, they correspond to mean-reversion rate $\theta = -\ln \rho / \Delta t \approx 0.357\,\mathrm{s}^{-1}$ and diffusion coefficient $\sigma_{OF}^{c} = \sigma_{OF}\sqrt{2\theta/(1-\rho^2)} \approx 1.183$.

## 5.3 Parameter Calibration

The volatility parameter $\sigma$ is calibrated from daily closing prices of the S&P 500 index, obtained from the Federal Reserve Economic Data (FRED) database [18] and covering the period from 16 March 2016 to 13 March 2026, a sample of roughly ten years of daily observations. Log returns are computed as

$$r_t = \log(P_t / P_{t-1})$$

and a 21-day rolling window is used to estimate daily return volatility. The median rolling volatility over the sample is taken as the baseline, giving $\sigma_{\text{daily}} = 0.0076$.

Because the model operates in price levels rather than returns, the daily volatility is converted to the one-step innovation parameter via

$$\sigma = \frac{V_0\,\sigma_{\text{daily}}}{\sqrt{T}} = \frac{100 \cdot 0.0076}{\sqrt{23\,400}} = 0.004968$$

## 5.4 Microstructure Regimes

The simulation varies two parameters across three regimes: the price impact coefficient $\lambda$ and the noise sensitivity $\alpha$. Table 3 reports the values. The low regime approximates a near-frictionless market, while the high regime represents conditions in which order flow has a stronger effect on observed prices and noise levels, making efficient price recovery most demanding.

**Table 3:** Microstructure regimes

| Regime | $\lambda$ | $\alpha$ | Interpretation |
|---|---|---|---|
| Low | 0.02 | 0.01 | Near-frictionless market |
| Normal | 0.05 | 0.05 | Typical trading conditions |
| High | 0.10 | 0.10 | Strong microstructure distortions |

## 5.5 Estimation Procedures

For each simulated path, the latent efficient price process $V_t$ is estimated using two alternative procedures: a Kalman filter and a nonlinear machine learning model based on XGBoost. In both cases, the object of interest is the contemporaneous latent efficient price $V_t$, not a multi-step-ahead forecast. The aim is therefore to evaluate how accurately the unobserved efficient price can be inferred from noisy observable price and order-flow data generated by the simulation model.

Each Monte Carlo replication corresponds to one independently simulated trading day observed at one-second frequency. The estimation procedure is carried out separately for each simulated day. For every replication and every regime, the same simulated path is used for both estimators.

Each path is divided into a training segment and an evaluation segment. The XGBoost model is fitted on the training segment and evaluated on the held-out segment. The Kalman filter is applied to the same simulated path, and its performance is evaluated on the same held-out segment. This ensures that both estimators are compared on a common test sample.

**Kalman Filter** The filter is based on the linear Gaussian state-space model introduced in the previous section. The latent state is the efficient price $V_t$, which evolves according to the random walk

$$V_t = V_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma^2 \Delta t)$$

The observable input to the filter is the transaction price $P_t$, which is treated as a noisy proxy for the latent efficient price through the observation equation

$$P_t = V_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_0^2)$$

This specification is intentionally misspecified relative to the true data-generating process. In the simulation model, transaction prices are also affected by order flow

through the term $\lambda OF_t$, and the observation noise variance depends on order flow through $\sigma_0^2 + \alpha|OF_t|$. By contrast, the Kalman filter assumes a simpler homoskedastic linear structure and ignores the direct effect of order flow on transaction prices. The misspecification is deliberate, reflecting the fact that empirical filtering models are typically simpler than the underlying market process.

The Kalman filter does not involve a separate training phase. For each simulated path, it is applied recursively to the full observed transaction price series, producing a filtered estimate $\widehat{V}_t^{KF}$ at every time point. Although the filter is run over the full path, its performance is evaluated only on the held-out segment in order to maintain comparability with the machine learning estimator.

**XGBoost Estimator** XGBoost is a nonlinear supervised learning model based on gradient-boosted decision trees. It uses the observable feature vector

$$X_t = (P_t, OF_t, |OF_t|, \Delta P_t)$$

to estimate the latent efficient price $V_t$ at each time point. The goal is efficient price recovery rather than forecasting: the model maps the current $X_t$ directly to an estimate of $V_t$ without recursive updating.

The model is trained on the first 70% of each simulated path using contemporaneous $(X_t, V_t)$ pairs and evaluated on the remaining 30%. A separate model is fitted for each Monte Carlo replication, preserving time ordering and ensuring out-of-sample evaluation.

The hyperparameters used in the single-ETF experiments are reported in Table 4. The values were selected manually based on standard practice for gradient-boosted models. The sensitivity of the main results to this choice is examined in Section 6.5.

**Table 4:** XGBoost hyperparameters

| Hyperparameter | Value |
|---|---|
| `n_estimators` | 300 |
| `max_depth` | 4 |
| `learning_rate` | 0.05 |
| `subsample` | 0.8 |
| `colsample_bytree` | 0.8 |
| `objective` | `reg:squarederror` |

## 5.6 Evaluation Procedure

Both estimators are evaluated on the final 30% of each simulated path. The primary metric is MSE as defined in Section 4.

**Directional Change Lag** The lag metric measures how quickly each estimator tracks turning points in $V_t$. A directional change event at time $t$ is recorded whenever

the sign of the first difference $\Delta V_t = V_t - V_{t-1}$ reverses relative to $\Delta V_{t-1}$, with both differences nonzero. For each such event with new sign $s \in \{-1, +1\}$, the procedure searches forward in the estimated series $\widehat{V}_t$ for the first directional change in $\Delta\widehat{V}_t$ whose new sign equals $s$. The search is bounded by a maximum horizon of 30 steps, corresponding to 30 seconds at the one-second simulation frequency. If a matching change is found at step $t_{\text{est}} \geq t$, its lag is recorded as $t_{\text{est}} - t$; if no matching change occurs within the horizon, the event is dropped. The reported lag is the mean over all matched events in the evaluation segment.

By construction the metric is one-sided: it captures delay rather than anticipation, and assigns no explicit penalty to events the estimator fails to register within the horizon. An estimator may therefore achieve low MSE while still reacting slowly to turning points, and the lag metric captures this distinct dimension of performance.

## 5.7 Monte Carlo Aggregation

The estimation and evaluation procedure described above is repeated for $N = 1000$ independently simulated paths within each microstructure regime. For each replication, the MSE and directional change lag are computed separately for the Kalman filter and the XGBoost estimator.

The reported regime-level results are obtained by averaging these path-level performance measures across replications within each regime. In addition to mean performance, the standard deviation of the MSE across replications is also reported to summarize the stability of each estimator under repeated simulation.

# 6 Results

This section presents the results of the Monte Carlo simulation study described in the previous chapter. The aim is to evaluate how accurately the latent efficient price can be recovered from noisy observable microstructure data under different microstructure conditions, and to compare the performance of the Kalman filter and the nonlinear XGBoost estimator.

The results reveal a clear regime-dependent pattern. In the low microstructure regime, the Kalman filter recovers the latent price substantially more accurately than XGBoost. As distortions intensify, the advantage reverses: in the normal regime the two estimators perform similarly, and in the high regime XGBoost is markedly superior. The crossover suggests that the cost of linear misspecification becomes binding only when order flow effects and noise intensity are sufficiently strong.

## 6.1 Overview of Main Findings

Table 5 reports the mean MSE and directional change lag for both estimators across all three microstructure regimes.

**Table 5:** Summary of Monte Carlo results ($N = 1000$ per regime). Each MSE entry is reported as mean ± standard deviation across the $N = 1000$ Monte Carlo replications. The standard deviation measures the replication-to-replication stability of each estimator rather than within-path variation.

| Regime | $\text{MSE}_{\text{KF}}$ | $\text{MSE}_{\text{ML}}$ | $\text{Lag}_{\text{KF}}$ (s) | $\text{Lag}_{\text{ML}}$ (s) |
|---|---|---|---|---|
| Low | 0.0107 ± 0.0003 | 0.0310 ± 0.0684 | 1.23 | 1.53 |
| Normal | 0.0547 ± 0.0015 | 0.0518 ± 0.0780 | 1.24 | 1.55 |
| High | 0.1189 ± 0.0033 | 0.0626 ± 0.0683 | 1.26 | 1.55 |

Throughout the discussion, the relative improvement of XGBoost over the Kalman filter is summarised by the quantity $\Delta\text{MSE} = (\text{MSE}_{KF} - \text{MSE}_{ML})/\text{MSE}_{KF}$, where positive values indicate a lower error for XGBoost. Its magnitude in each regime, −189% in the low regime, +5.3% in the normal regime, and +47.3% in the high regime, is revisited in the detailed discussion below.

The table reveals three consistent patterns across regimes: the Kalman filter dominates at low distortions, XGBoost gains a decisive advantage at high distortions, and the Kalman filter is faster and more stable across all regimes.

## 6.2 Estimation Accuracy Across Microstructure Regimes

In the low microstructure regime, the Kalman filter achieves a mean MSE of 0.0107, compared to 0.0310 for XGBoost. This means that the XGBoost error is about 189% higher than the Kalman filter's. When microstructure distortions are weak, transaction prices remain close to the efficient price, so the linear structure of the Kalman filter is well suited to the problem. In this regime, the misspecification of the Kalman filter

is relatively mild: the neglected order flow term $\lambda OF_t$ is small, and deviations from homoskedasticity are of limited importance. XGBoost, by contrast, is applied to a relatively smooth and close-to-linear relationship, so its additional flexibility provides little benefit.

In the normal microstructure regime, the two estimators perform similarly, with XGBoost achieving a modest improvement of 5.3% over the Kalman filter (MSE 0.0518 versus 0.0547). This regime can be interpreted as an intermediate case in which nonlinear features of the data-generating process begin to matter, but only to a limited extent.

In the high microstructure regime, XGBoost achieves a mean MSE of 0.0626 compared to 0.1189 for the Kalman filter, an improvement of 47.3%. When microstructure distortions are strong, the order flow term $\lambda OF_t$ introduces substantial systematic deviations of transaction prices from the efficient price, while the noise variance $\sigma_0^2 + \alpha|OF_t|$ creates heteroskedastic patterns that are not captured by the Kalman specification. Because XGBoost has access to order flow and related observable features, it can exploit these directly to partially correct for the distortions.

## 6.3 Directional Change Responsiveness

The Kalman filter exhibits lower directional change lag than XGBoost in all three regimes. In the low regime, the Kalman filter reacts to turning points in the latent price with an average delay of 1.23 seconds, compared to 1.53 seconds for XGBoost. The gap remains similar across regimes, with the Kalman filter consistently about 0.3 seconds faster.

The Kalman filter updates recursively at each time step, immediately incorporating new price observations into the state estimate. XGBoost, by contrast, maps $X_t$ to an estimate through a fixed function learned from training data. This function may be more accurate on average in high-distortion regimes, but it does not recursively update a previous state estimate, which limits its responsiveness to directional changes.

The lag results therefore reveal a trade-off: XGBoost recovers the level of the efficient price more accurately under strong microstructure distortions, but the Kalman filter tracks the direction of price movements more quickly across all conditions.

## 6.4 Stability Across Monte Carlo Replications

The standard deviation of MSE across replications differs substantially between the two estimators. In the low regime, the Kalman filter has a standard deviation of 0.0003, while XGBoost has a standard deviation of 0.0684, a factor of approximately 235 larger. In the normal regime the ratio is roughly 52, and in the high regime roughly 21. This narrowing ratio reflects the fact that the Kalman filter itself becomes more variable as distortions intensify. Its standard deviation rises from 0.0003 to 0.0033 across regimes, while XGBoost's absolute variability remains broadly stable.

A likely explanation is that each XGBoost model is trained on a single simulated path, making its out-of-sample performance sensitive to the particular realization of the training segment. In some replications, the training segment may not expose the

model to the full range of microstructure patterns that appear later in the evaluation segment, leading to larger estimation errors. The Kalman filter, by contrast, does not involve a path-specific training phase, which makes its performance less sensitive to this source of variation.

## 6.5 Robustness to XGBoost Hyperparameters

The main Monte Carlo results use a single fixed configuration for XGBoost (`max_depth` = 4, `n_estimators` = 300), as specified in Table 4. Here `max_depth` controls the maximum depth of each tree in the boosted ensemble, governing the complexity of feature interactions a single tree can represent, while `n_estimators` sets the number of sequential boosting rounds, that is, how many trees are added to the ensemble. Larger values of either parameter increase model capacity, at the cost of greater overfitting risk. To assess whether the regime-dependent pattern documented above reflects the structure of the estimation problem rather than an arbitrary choice of hyperparameters, a grid study is conducted over `max_depth` $\in \{3, 4, 6, 8\}$ and `n_estimators` $\in \{300, 1000\}$, giving eight configurations per regime. Each configuration is evaluated on $N = 200$ Monte Carlo replications per regime. Results are summarised in Table 6.

**Table 6:** XGBoost hyperparameter sensitivity across regimes ($N = 200$ per configuration). Best and worst configurations are selected by mean MSE within each regime. $\Delta$MSE is the relative improvement of XGBoost over the Kalman filter, defined as in the main table.

| Regime | Best configuration | $\text{MSE}_{\text{best}}$ | Worst configuration | $\text{MSE}_{\text{worst}}$ |
|---|---|---|---|---|
| Low | depth 3, 1000 trees | 0.0341 (−217%) | depth 8, 1000 trees | 0.0382 (−255%) |
| Normal | depth 3, 1000 trees | 0.0495 (+9.4%) | depth 8, 1000 trees | 0.0563 (−2.9%) |
| High | depth 3, 1000 trees | 0.0611 (+48.5%) | depth 8, 1000 trees | 0.0673 (+43.3%) |

Three observations emerge. First, the regime-dependent pattern is robust across the full grid: XGBoost loses to the Kalman filter in the low regime for every configuration, performs comparably in the normal regime, and dominates in the high regime. No hyperparameter choice reverses the qualitative ordering. Second, shallow trees dominate in all regimes. The configuration `max_depth` = 3 produces the lowest MSE in each regime, and performance degrades monotonically as depth increases. This is consistent with the interpretation that each XGBoost model is trained on a single simulated path, so deeper trees overfit rather than capture additional structure. Third, the number of boosting rounds has a small and direction-dependent effect. With shallow trees, increasing `n_estimators` from 300 to 1000 yields a minor improvement; with deeper trees, it amplifies the overfitting penalty.

The variability of XGBoost across Monte Carlo replications is also insensitive to hyperparameter choice. The standard deviation of MSE remains in the range 0.07–0.08 across all configurations and regimes, which is one to two orders of magnitude larger than the corresponding Kalman filter standard deviations reported in Table 5. The

high replication variance of XGBoost therefore reflects the path-specific nature of its training rather than a symptom of suboptimal tuning.

In summary, the baseline configuration used in the main results is itself an element of the grid, and the best grid configuration improves on it by at most two percentage points in $\Delta$MSE in any regime. The fixed-hyperparameter results presented above therefore already represent XGBoost's near-optimal performance in this environment, and the regime-dependent superiority of the Kalman filter in the low regime cannot be overturned by hyperparameter tuning alone.

## 6.6 Interpretation and Implications

The results show that efficient price recovery is strongly regime-dependent. In low-distortion environments, the linear Kalman filter is both more accurate and more responsive than XGBoost, so its structural simplicity is an advantage rather than a limitation. In high-distortion environments, the nonlinear flexibility of XGBoost becomes a clear advantage, as it can exploit the systematic relationship between order flow and transaction prices that the misspecified Kalman filter does not model.

More broadly, the results indicate that the recoverability of the latent efficient price depends strongly on the microstructure regime. For the Kalman filter, MSE increases by roughly a factor of ten from the low to the high regime, whereas for XGBoost it increases by only about a factor of two. This suggests that the estimation problem becomes substantially harder as microstructure distortions intensify, and that estimator choice matters most in the more challenging regimes.

## 6.7 Summary of Results

The Kalman filter outperforms XGBoost in the low microstructure regime, where its linear structure is well suited to the estimation problem. XGBoost gains a modest advantage in the normal regime and a clear advantage in the high regime, where it can exploit nonlinear order-flow patterns that the misspecified Kalman filter does not model explicitly. Across all regimes, the Kalman filter reacts more quickly to directional changes in the latent price and produces more stable estimates across Monte Carlo replications.

These results provide the baseline against which the multi-ETF extension in Section 7 is evaluated. The central question in that analysis is whether incorporating cross-asset information from multiple ETFs tracking the same underlying index improves on these single-asset benchmarks.

# 7 Multi-ETF Extension

This section extends the single-asset simulation framework to a multi-ETF setting. The central question is whether cross-asset price information can improve efficient price recovery when multiple exchange-traded funds track the same underlying index.

## 7.1 Motivation

In the single-asset framework, the estimation of the latent efficient price relies solely on the transaction prices and order flow of a single ETF. However, when several ETFs track the same underlying index, their prices are noisy observations of the same latent process. Intuitively, a liquid ETF with low microstructure noise should provide information about the common efficient price that can be exploited to improve estimation for a less liquid ETF.

This setting also has a natural interpretation in the context of statistical arbitrage. Petäjistö [3] proposes using the cross-section of prices among ETFs with nearly identical underlying portfolios as a real-time proxy for the true underlying value, arguing that any dispersion of an individual ETF price around the group mean is more likely attributable to mispricing than to stale pricing. He further documents that a long-short strategy exploiting cross-sectional differences in ETF premiums generates a Carhart alpha of approximately 7% per year. The present framework formalizes this intuition in a stochastic filtering setting: the common latent factor $F_t$ plays the role of Petäjistö's group-level price proxy, and the estimation problem is to recover $F_t$ from the noisy transaction prices of the individual ETFs. A position that is long the underpriced ETF and short the overpriced ETF may then be constructed based on the estimated fair value.

## 7.2 Multi-Asset Data Generating Process

Consider three ETFs that track the same underlying index. Because all three track the same portfolio, they share a single efficient price, identified here with the latent common factor $F_t$.

### 7.2.1 Common Latent Factor

The common efficient price follows a discretized driftless arithmetic Brownian motion:

$$F_t = F_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma^2 \Delta t)$$

This is the same process introduced for the single-asset latent price $V_t$ in Section 4; the symbol $F_t$ is used here to emphasize its interpretation as a factor shared across ETFs. As before, $F_t$ is a martingale with independent increments. The structural parameters are held at the values specified in Table 2, and $F_t$ is the object of estimation for every estimator in this chapter.

### 7.2.2 ETF-Specific Transaction Prices

Each ETF $i \in \{1, 2, 3\}$ has its own order flow process and microstructure parameters. The order flow for ETF $i$ follows an independent Ornstein–Uhlenbeck process with the same parameters as in Section 4.3.2, sampled on the one-second grid as

$$OF_{i,t} = \rho\, OF_{i,t-1} + \xi_{i,t}, \qquad \xi_{i,t} \sim \mathcal{N}(0, \sigma^2_{OF}).$$

The order flow processes are mutually independent across ETFs. In practice, the order flows of ETFs tracking the same index are likely correlated through arbitrage activity, but treating them as independent is a deliberate simplification: it isolates the mechanism of interest, so that any gain from the multi-ETF estimators can be attributed to the shared latent signal rather than to exploitable cross-asset order-flow patterns.

The observed transaction price of ETF $i$ is

$$P_{i,t} = F_t + \lambda_i\, OF_{i,t} + \varepsilon_{i,t}$$

where

$$\varepsilon_{i,t} \sim \mathcal{N}\Big(0,\ \sigma_0^2 + \alpha_i |OF_{i,t}|\Big)$$

The noise processes $\varepsilon_{i,t}$ are independent across ETFs and serially uncorrelated.

### 7.2.3 ETF Heterogeneity

The three ETFs are distinguished solely by their microstructure parameters $\lambda_i$ and $\alpha_i$. ETF 1 is the most liquid, with the smallest price impact and lowest noise sensitivity; ETF 3 is the least liquid, with the strongest microstructure distortions; ETF 2 sits between the two. The three profiles are calibrated to span the same low-, normal-, and high-noise range that the single-asset analysis of Section 6 explored across separate regimes, so that the cross-section of ETFs reproduces that heterogeneity within a single simulated path.

**Table 7:** ETF-specific microstructure parameters

| ETF | Liquidity profile | Price impact $\lambda_i$ | Noise sensitivity $\alpha_i$ |
|---|---|---|---|
| ETF 1 | High liquidity | 0.02 | 0.01 |
| ETF 2 | Normal liquidity | 0.05 | 0.05 |
| ETF 3 | Low liquidity | 0.10 | 0.10 |

All other parameters ($\sigma$, $\rho$, $\sigma_{OF}$, $\sigma_0$) are shared across ETFs and held at the values from Table 2. In reality, the order flow volatility $\sigma_{OF}$ would also be expected to vary with liquidity, but heterogeneity is concentrated here in $\lambda_i$ and $\alpha_i$ to keep the experiment focused on a single dimension of variation.

## 7.3 Estimators

Each ETF is analyzed with four estimators, obtained by combining two methods (Kalman filter and XGBoost) with two information sets (single-ETF and multi-ETF). The single-ETF variants use only the data of ETF $i$ and match the estimators of Section 4; the multi-ETF variants also use the transaction prices and order flows of the other two ETFs. Table 8 summarizes the four estimators.

**Table 8:** Overview of estimators per ETF

| Estimator | Data used | How information is combined |
|---|---|---|
| Single-ETF Kalman | Only ETF $i$: $P_{i,t}$ | Recursively |
| Multi-ETF Kalman | All ETFs: $P_{1,t}, P_{2,t}, P_{3,t}$ | Recursively, 3 observations at once |
| Single-ETF XGBoost | Only ETF $i$: 4 features | Learned from data |
| Multi-ETF XGBoost | All ETFs: 12 features | Learned from data |

### 7.3.1 Single-ETF Estimators (Baseline)

The single-ETF variants are identical to the estimators of Section 4, applied separately to each ETF. The Kalman filter treats $P_{i,t}$ as a noisy observation of the scalar latent state $F_t$ and is run under the same deliberate misspecification as in the baseline: the order flow term $\lambda_i OF_{i,t}$ is ignored and the observation noise is assumed homoskedastic. The XGBoost estimator uses the four-dimensional feature vector $(P_{i,t}, OF_{i,t}, |OF_{i,t}|, \Delta P_{i,t})$ introduced in Section 4, and is trained to predict $F_t$. Both serve as benchmarks for the multi-ETF variants.

### 7.3.2 Multi-ETF Kalman Filter

The multi-ETF Kalman filter keeps $F_t$ as a scalar latent state but treats the three ETF prices as a joint noisy observation of that state. The state equation is unchanged from Section 4; the observation equation becomes

$$\mathbf{P}_t = \begin{pmatrix} P_{1,t} \\ P_{2,t} \\ P_{3,t} \end{pmatrix} = \mathbf{h} F_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}),$$

with $\mathbf{h} = (1, 1, 1)^\top$ and

$$\mathbf{R} = \sigma_0^2 \, \mathbf{I}_3.$$

Here $\sigma_0^2$ is the baseline microstructure noise variance of Section 4, representing the minimum level of observation noise when order flow is zero. As in the single-asset case, this specification is deliberately misspecified: the filter ignores the order flow term $\lambda_i OF_{i,t}$ and treats the noise as homoskedastic, with the same variance assigned to every ETF. In particular, the filter does not exploit the known heterogeneity in ETF liquidity and each ETF is weighted equally at every update step, even though ETF 1 is

far cleaner than ETF 3. This is a strong simplification relative to the data generating process and is expected to limit how much the multi-ETF Kalman filter can improve on its single-ETF counterpart.

The filter is applied using the standard Kalman recursions; see Section 5 for the single-asset analogue. The resulting filtered estimate is denoted $\widehat{F}_t^{KF,\text{multi}}$.

### 7.3.3 Multi-ETF XGBoost

The multi-ETF XGBoost estimator extends the single-asset feature vector by stacking the corresponding features of all three ETFs. For each ETF $i$ the input is

$$\mathbf{X}_t^{(i)} = \left(P_{1,t},\ P_{2,t},\ P_{3,t},\ OF_{1,t},\ OF_{2,t},\ OF_{3,t},\ |OF_{1,t}|,\ |OF_{2,t}|,\ |OF_{3,t}|,\ \Delta P_{1,t},\ \Delta P_{2,t},\ \Delta P_{3,t}\right)^\top,$$

a 12-dimensional vector in place of the four features used in the single-ETF baseline. The target remains the common factor $F_t$, so that the three ETF-specific models share an estimation target but differ in how they might weight the features of a given ETF relative to those of the others. Hyperparameters and the 70/30 chronological train-test split follow Section 5.

## 7.4 Comparison Framework

The four estimators per ETF are collected in Table 9.

**Table 9:** Estimator comparison framework

| | Single-ETF | Multi-ETF |
|---|---|---|
| Kalman filter | $\widehat{F}_t^{KF,\text{single}}$ | $\widehat{F}_t^{KF,\text{multi}}$ |
| XGBoost | $\widehat{F}_t^{ML,\text{single}}$ | $\widehat{F}_t^{ML,\text{multi}}$ |

Each estimator is evaluated with the two metrics of Section 4: mean squared error, which captures point-wise accuracy, and directional change lag, which captures how quickly the estimator reacts to moves in the underlying factor. Results are reported separately for each ETF, whose distinct microstructure parameters already span the low-, normal-, and high-noise range studied in Section 6.

The comparison is aimed at two questions: whether cross-asset information improves efficient price recovery, and whether the benefit is evenly spread across ETFs. The data generating process itself suggests an uneven answer. ETF 3, whose transaction prices are the most distorted by microstructure noise, stands to gain the most from borrowing information from a cleaner reference. ETF 1 is already close to $F_t$ and has little to gain. Under the misspecified multi-ETF Kalman filter, which weights all three ETFs equally, it may even lose, being pulled toward the noisier price paths of ETF 2 and ETF 3.

# 8 Multi-ETF Results

This chapter reports the results of the multi-ETF Monte Carlo simulation over $N = 1000$ independently simulated trading days, comparing the single-ETF baseline estimators of Section 6 with their multi-ETF counterparts.

The results reveal an asymmetry. Cross-asset information benefits XGBoost uniformly and benefits the Kalman filter only for the noisier ETFs, while actively harming the estimate for the cleanest one. In addition, the regime-dependent gap between the Kalman filter and XGBoost, which dominated the single-asset analysis, nearly disappears in the multi-ETF setting: the two estimators converge to essentially the same MSE.

## 8.1 Overview

Table 10 reports the mean MSE of the single-ETF estimators per ETF, together with the relative change $\Delta$ from single- to multi-ETF. Because the multi-ETF estimators produce a single estimate of the common factor $F_t$, the multi-ETF MSE values are shared across all three ETFs and reported in the table caption.

**Table 10:** Multi-ETF Monte Carlo results ($N = 1000$). MSE entries are mean $\pm$ standard deviation across replications. The multi-ETF estimators produce a single common-factor estimate shared across all three ETFs: $0.0219 \pm 0.0005$ for the Kalman filter and $0.0218 \pm 0.0403$ for XGBoost. $\Delta$ is the relative MSE improvement from single- to multi-ETF; positive values favor the multi-ETF estimator.

| ETF | Liquidity | $\text{MSE}_{\text{KF,single}}$ | $\Delta$KF | $\text{MSE}_{\text{ML,single}}$ | $\Delta$ML |
|---|---|---|---|---|---|
| ETF 1 | High | $0.0107 \pm 0.0003$ | $-103\%$ | $0.0249 \pm 0.0447$ | $+13\%$ |
| ETF 2 | Normal | $0.0547 \pm 0.0015$ | $+60\%$ | $0.0461 \pm 0.0532$ | $+53\%$ |
| ETF 3 | Low | $0.1189 \pm 0.0034$ | $+82\%$ | $0.0648 \pm 0.0644$ | $+66\%$ |

The table quantifies the asymmetry anticipated in the introduction. XGBoost gains from cross-asset information for every ETF, by 13%, 53% and 66% moving from ETF 1 to ETF 3. The Kalman filter gains for ETF 2 and ETF 3 (+60% and +82%) but deteriorates by 103% for ETF 1. After combining the three ETF signals, the two multi-ETF estimators converge to essentially the same accuracy, 0.0218 for XGBoost against 0.0219 for the Kalman filter, in sharp contrast to the regime-dependent gap observed in the single-asset analysis of Section 6.

## 8.2 Cross-Asset Information and the Kalman Filter

The multi-ETF Kalman filter reaches a mean MSE of 0.0219. Relative to the single-ETF baseline this is an improvement of 60% for ETF 2 and 82% for ETF 3, but a deterioration of 103% for ETF 1.

The deterioration for ETF 1 follows directly from the filter misspecification. The multi-ETF Kalman filter assigns equal observation noise variance $\sigma_0^2$ to all three ETFs,

treating them as interchangeable sources of information about $F_t$. In reality, ETF 1 has much lower microstructure distortions than ETF 2 and ETF 3. Combining the clean ETF 1 signal with the noisier signals under a uniform weighting scheme, the filter dilutes the high-quality signal rather than amplifying it. Applied to ETF 1 alone, the filter achieves MSE 0.0107; forced to average in the noisier ETFs, it deteriorates to 0.0219.

For ETF 2 and ETF 3 the uniform weighting scheme works in the opposite direction. Because their own transaction prices are dominated by microstructure noise, even an equal-weighted contribution from the cleaner ETF 1 anchors the common factor estimate more accurately than the single-ETF baseline.

## 8.3 Cross-Asset Information and XGBoost

The multi-ETF XGBoost estimator reaches a mean MSE of 0.0218 and improves over the single-ETF baseline for all three ETFs, by 13% for ETF 1, 53% for ETF 2 and 66% for ETF 3.

Unlike the Kalman filter, XGBoost does not impose a fixed weighting scheme on the ETF price signals. By learning from simulated data, it can implicitly weight each ETF according to its relevance to the common factor. For ETF 1 the improvement is modest because the single-ETF baseline already exploits a low-noise signal effectively; for ETF 2 and ETF 3 the 12-dimensional feature vector gives the model access to the clean ETF 1 prices, letting it correct for the microstructure distortions that drive the single-ETF estimation error.

The monotone pattern (13%, 53%, 66%) matches the expectation set out in Section 7.4: ETF 3 gains the most because its transaction prices are the most distorted by microstructure noise.

## 8.4 Kalman Filter versus XGBoost in the Multi-ETF Setting

In the single-asset analysis, the Kalman filter and XGBoost differed sharply by regime: the Kalman filter dominated in the low-distortion regime, and XGBoost dominated in the high-distortion regime. In the multi-ETF setting this difference largely disappears. The multi-ETF Kalman filter reaches MSE 0.0219 against 0.0218 for XGBoost, a gap below 1%.

The convergence reflects an asymmetry in how much room each estimator had to improve. The Kalman filter lost to XGBoost in the high-distortion regime of Section 6 precisely because its misspecification was most costly there. The multi-ETF setting gives the filter access to the clean ETF 1 signal, which anchors the common factor estimate and corrects most of that cost. XGBoost was already closer to the attainable accuracy in the single-asset analysis and has less room to gain from cross-asset information. The two estimators therefore meet near the same point.

## 8.5 Directional Change Responsiveness and Stability

Table 11 reports the directional change lag and the MSE standard deviation across replications for the multi-ETF and single-ETF estimators.

**Table 11:** Lag and stability results. Lag is the mean directional change lag in seconds, averaged across the $N = 1000$ Monte Carlo replications. SD(MSE) is the standard deviation of the path-level MSE across those replications, restated here from Table 10 to ease side-by-side comparison across estimators.

| Estimator | Lag (s) | SD(MSE) |
|---|---|---|
| Multi-ETF KF | 1.216 | 0.0005 |
| Multi-ETF ML | 1.514 | 0.0403 |
| Single-ETF KF, ETF 1 | 1.226 | 0.0003 |
| Single-ETF ML, ETF 1 | 1.530 | 0.0447 |
| Single-ETF KF, ETF 2 | 1.241 | 0.0015 |
| Single-ETF ML, ETF 2 | 1.554 | 0.0532 |
| Single-ETF KF, ETF 3 | 1.258 | 0.0034 |
| Single-ETF ML, ETF 3 | 1.545 | 0.0644 |

The directional change lag of the multi-ETF Kalman filter (1.216 s) is slightly lower than that of any single-ETF Kalman filter variant (1.226–1.258 s). With three simultaneous observations to condition on, each state update is more precise and the filter reacts marginally faster to moves in $F_t$.

The multi-ETF XGBoost lag (1.514 s) is also below the single-ETF variants (1.530–1.554 s), by a similarly small margin. Either way, the reduction from cross-asset information is minor compared with the persistent gap between the two estimators: across all specifications, the Kalman filter reacts to directional changes about 0.3 seconds faster than XGBoost.

The MSE standard deviations in Table 11 reveal a similar pattern. The multi-ETF Kalman filter has an MSE standard deviation of 0.0005, close to the single-ETF Kalman filter variants and nearly two orders of magnitude below any XGBoost variant. The multi-ETF XGBoost has a standard deviation of 0.0403, lower than each of the single-ETF XGBoost variants (0.0447, 0.0532 and 0.0644 for ETF 1, ETF 2 and ETF 3). The added ETF signals stabilize the trained model, most visibly for the noisier ETFs where single-ETF variance is highest.

## 8.6 Summary

Cross-asset information improves efficient price recovery for both estimators, but in different ways. For the misspecified Kalman filter, combining the three ETF signals under a uniform weighting scheme helps the noisier ETFs but harms the already-accurate ETF 1. The net effect is a multi-ETF Kalman filter that is competitive but not dominant. For XGBoost, cross-asset information is uniformly beneficial: the learned feature weights adapt to the heterogeneity in ETF liquidity, and the gain is largest for the low-liquidity ETF 3.

In the multi-ETF setting the Kalman filter and XGBoost converge to near-identical MSE, in contrast to the regime-dependent gap of the single-asset analysis. The Kalman filter nevertheless continues to react more quickly to directional changes and to produce more stable estimates across Monte Carlo replications.

# 9 Discussion

## 9.1 Synthesis of Main Findings

The single-ETF experiment shows that efficient price recovery is feasible in all three microstructure regimes, but that accuracy degrades substantially as microstructure distortions intensify. The relative performance of the two estimators also reverses across regimes. In the low regime, the Kalman filter achieves mean squared error 0.0107 against XGBoost's 0.0310. In the high regime, the order flips: XGBoost reaches 0.0626 while the filter lands at 0.1189. The normal regime sits between these extremes, with the two estimators nearly equivalent. The pattern is stable across 1 000 Monte Carlo replications and reflects a fundamental tradeoff between structural efficiency and nonlinear flexibility.

The multi-ETF experiment adds a second dimension to this tradeoff. When cross-asset information is available, the benefit to each estimator depends on how it incorporates the additional observations. XGBoost improves uniformly across all three ETFs, with gains that grow monotonically with illiquidity: 13%, 53%, and 66% for ETF 1, ETF 2, and ETF 3. The misspecified Kalman filter is more selective. It is constrained to assign equal observation weights across all three ETFs, which helps the less liquid instruments but dilutes the clean ETF 1 signal with the noisier ETF 2 and ETF 3 inputs. The filter therefore improves for the less liquid ETFs but deteriorates for the most liquid one.

The net result is that the two multi-ETF estimators converge in accuracy. The multi-ETF Kalman filter achieves MSE 0.0219 and multi-ETF XGBoost 0.0218, a difference of less than 1%. This stands in sharp contrast to the regime-dependent gaps seen in the single-asset analysis. In all specifications, the Kalman filter reacts roughly 0.3 seconds faster to directional changes and produces substantially more stable estimates across Monte Carlo replications.

## 9.2 The Role of Model Misspecification

Model misspecification is a unifying theme across both experiments. The Kalman filter is intentionally misspecified in two respects. It ignores the order flow term $\lambda OF_t$ in the observation equation, and it treats the observation noise as homoskedastic. This misspecification is benign when microstructure distortions are weak: the neglected terms are small and the linear-Gaussian approximation closely tracks the true data generating process. It becomes increasingly costly as $\lambda$ and $\alpha$ grow, because the systematic component of the observation error grows relative to the signal.

The multi-ETF extension introduces a second layer of misspecification. The filter assigns equal observation noise variance to all three ETFs, ignoring the known heterogeneity in their microstructure parameters. As a consequence, it cannot exploit the cross-sectional variation in signal quality. A correctly specified multi-ETF Kalman filter would assign lower observation noise to ETF 1 and higher observation noise to ETF 3, producing a liquidity-weighted estimate of the common factor. Such a filter would likely improve on the single-ETF baseline for all three ETFs, including ETF 1.

The convergence of the two estimators in the multi-ETF setting is not a sign that linear filtering is fundamentally as powerful as a nonlinear estimator. It is the result of an uneven tradeoff. The misspecified Kalman filter gains substantially on the noisier ETFs but loses on the cleanest one, and the average happens to match the performance of XGBoost. XGBoost, by contrast, learns to weight the ETF signals by their informativeness and improves uniformly across all three. A correctly specified Kalman filter, which would assign liquidity-proportional observation weights, would likely improve on all three ETFs and could outperform XGBoost in aggregate.

The reason the misspecification is not costly on the most liquid ETF is simple: with the most liquid ETF, the nonlinearity of the observation noise is minimal. The misspecification of the Kalman filter is therefore very small. Conversely, XGBoost does not see significant help from its nonlinear mapping capabilities, and can easily be overfitted, especially since the XGBoost estimator is trained separately for each Monte Carlo replication. This implies that using XGBoost would be most useful in a situation where there are multiple illiquid ETFs tracking the same underlying security.

## 9.3 Practical Implications for ETF Pricing and Statistical Arbitrage

Although this study is simulation-based, the framework is grounded in market microstructure theory and the results have practical interpretations.

The single-ETF results suggest that the choice of estimation method should depend primarily on the liquidity profile of the ETF in question. For highly liquid ETFs with tight spreads and low price impact, a linear Kalman filter provides accurate and fast efficient price estimates at low computational cost. For less liquid ETFs, where order flow imbalances create substantial systematic deviations from the efficient price, nonlinear estimators such as XGBoost offer a meaningful accuracy advantage.

The multi-ETF results speak directly to the cross-sectional ETF arbitrage strategy described by Petäjistö [3]. In that framework, the group-level price of ETFs tracking the same index serves as a proxy for the true underlying value. Individual ETF prices that deviate from the group mean are interpreted as mispricing rather than stale pricing. The common latent factor $F_t$ in this study formalises that proxy: it is the object that all three ETFs are attempting to price, and deviations $P_{i,t} - \widehat{F}_t$ represent the estimated mispricing of each ETF. The multi-ETF results suggest that a practitioner implementing such a strategy would benefit from using a nonlinear cross-asset estimator rather than a misspecified linear filter, particularly for the less liquid instruments where the estimation error of the single-ETF baseline is largest. However, implementing a nonlinear estimation model is considerably more difficult in a real market environment, as order flow and the latent price are not as easily measurable.

In addition, the practical profitability of such a cross-ETF arbitrage strategy depends on factors beyond estimation accuracy, including transaction costs, execution latency, and the stability of the cross-sectional relationship over time. The present framework abstracts from these considerations and focuses solely on the statistical

recovery of the latent efficient price.

## 9.4 Limitations

Several limitations of the current framework should be acknowledged.

**Independent order flows.** The order flow processes of the three ETFs are assumed to be mutually independent. In practice, the order flows of ETFs tracking the same index are likely to be correlated, as arbitrageurs simultaneously trade multiple instruments to exploit pricing discrepancies. Correlated order flows would introduce additional cross-ETF predictability that neither the misspecified Kalman filter nor the current feature set of XGBoost explicitly exploits. The independence assumption is therefore a conservative simplification. It ensures that the observed benefits of cross-asset information reflect the shared price signal rather than cross-ETF order flow patterns, but it may understate the total information available to a practitioner.

**Simulation environment.** The study is conducted entirely in a controlled simulation environment. The data generating process, while structurally motivated, is a stylised representation of actual market dynamics. In particular, the linear price impact specification, the Gaussian noise assumption, and the stationarity of the order flow process are approximations that may not hold in real market data. The results should therefore be interpreted as characterising the properties of the estimators within this specific data generating process rather than as direct predictions of their performance on empirical data.

**Unobservability of the efficient price.** The simulation provides direct access to the latent efficient price, which makes estimation accuracy directly measurable. In practice, this quantity is fundamentally unobservable. Any empirical evaluation would require a proxy, such as the consolidated price across ETFs tracking the same index [3], the net asset value of the underlying basket, or the bid-ask midpoint. Each proxy carries its own measurement error and introduces circularity, since the benchmark against which estimators are compared is itself an estimate. The accuracy figures reported here should therefore be interpreted as what is achievable when ground truth is known, rather than as direct predictions of empirical performance.

**Order flow and price impact specification.** Order flow in the simulation is represented as a scalar Ornstein–Uhlenbeck process with linear price impact. Both simplifications depart from observed markets, where order flow is multivariate and price impact is nonlinear and regime-dependent. High-frequency order flow data also typically requires Level 2 market feeds that are not universally accessible, which constrains practical deployment. A richer feature set and a nonlinear impact function could alter the balance between the Kalman filter and XGBoost, particularly in settings where nonlinearity is the primary driver of the performance gap.

**XGBoost training variance.** The XGBoost estimator is trained on a single simulated path per Monte Carlo replication, making its out-of-sample performance sensitive to the particular realisation of the training segment. The resulting high standard deviation of MSE across replications (one to two orders of magnitude larger than the Kalman filter) indicates that practical deployment of an XGBoost-based estimator would require careful attention to training data selection and model stability. An alternative would be to train a single estimator on data from multiple days rather than refitting per replication. This procedure would likely reduce both estimation variance and the risk of overfitting.

**Fixed hyperparameters.** The main results use a single fixed XGBoost configuration across all regimes and ETF specifications. The sensitivity analysis in Section 6.5 shows that the baseline configuration is already close to the best in the grid: regime-specific tuning improves $\Delta$MSE by at most two percentage points in any regime, and the qualitative ordering of the two estimators across regimes is preserved for every configuration considered. Even these modest gains are specific to the simulation environment, where the heteroskedastic noise structure and its generating process are known. In a live market, the regime itself would first have to be inferred from the data, and the mapping from observable state to optimal hyperparameters would be considerably less tractable.

## 9.5 Directions for Future Research

A natural starting point for future work is a correctly specified multi-ETF Kalman filter, in which the observation noise variances are set proportional to the known microstructure parameters of each ETF. Comparing this correctly specified filter to the misspecified version and to XGBoost would isolate the cost of misspecification from the cost of using a linear model. A natural follow-up is an XGBoost estimator trained on a longer, multi-day sample, which could then be compared against both the misspecified and correctly specified Kalman filters.

A second direction is the introduction of correlated order flows across ETFs. This would require modifying the data generating process to allow for a common component in the order flow innovation, reflecting the coordinated trading activity of arbitrageurs. The multi-ETF XGBoost estimator could in principle exploit such cross-ETF order flow patterns through its 12-dimensional feature vector, potentially widening its advantage over the misspecified Kalman filter.

A third direction is the application of the framework to empirical intraday data. While the simulation environment provides full observability of the latent efficient price, empirical validation would require the construction of a proxy for the true efficient price, for example the consolidated price across all ETFs tracking the same index. Such a study would test whether the regime-dependent performance patterns documented here are also present in real market data.

Finally, the XGBoost estimator could be replaced or supplemented by sequence-aware architectures such as recurrent neural networks or transformers, which process

the full history of observations rather than a fixed feature vector at each time step. These models may be better suited to the sequential nature of the efficient price recovery problem and could reduce the training variance that limits the stability of the current XGBoost implementation.

# 10 Conclusion

This thesis investigated the statistical recoverability of a latent efficient price from noisy limit order book observations using a controlled simulation framework. The central research question asked how accurately the efficient price can be recovered, and whether a classical linear Kalman filter or a nonlinear XGBoost estimator is better suited to the task.

The single-ETF experiment demonstrated that recoverability is strongly regime-dependent. When microstructure distortions are weak, the misspecified Kalman filter achieves lower mean squared error than XGBoost and reacts more quickly to directional changes in the latent price. As distortions intensify, the Kalman filter's misspecification becomes increasingly costly, and XGBoost gains a decisive accuracy advantage by exploiting the nonlinear relationship between order flow and transaction prices. Across all regimes, the Kalman filter produces substantially more stable estimates across Monte Carlo replications, while XGBoost exhibits high variance that reflects sensitivity to the particular training path.

The multi-ETF experiment showed that cross-asset information improves efficient price recovery for both estimators, but in structurally different ways. XGBoost benefits uniformly across all three ETFs, with larger gains for less liquid instruments, because it learns to weight the ETF signals according to their informativeness. The misspecified Kalman filter, constrained to assign equal observation weights, benefits the less liquid ETFs substantially but deteriorates for the most liquid one. The net result is a convergence of the two estimators to nearly identical accuracy in the multi-ETF setting, a finding that is itself attributable to the filter's misspecification rather than to a fundamental limitation of linear filtering.

Taken together, the results support a regime-dependent and specification-sensitive view of efficient price recovery. The choice between a structural linear filter and a flexible nonlinear estimator is not universally resolved in favour of either approach: the Kalman filter is preferable when microstructure distortions are low and speed or stability is valued, while XGBoost is preferable when distortions are strong and a sufficiently representative training sample is available. In a multi-asset setting with heterogeneous ETFs, XGBoost is the more robust choice because it does not require the practitioner to specify the relative noise levels of the individual instruments.

The framework developed here provides a replicable and extensible foundation for studying efficient price recovery under microstructure noise. Natural extensions include a correctly specified multi-ETF Kalman filter with liquidity-weighted observations, correlated order flows across ETFs, and empirical validation against intraday data from ETFs tracking the same index.

# References


[1] U.S. Securities and Exchange Commission. *Concept Release on Exchange-Traded Funds*. SEC Concept Release. Accessed via Internet Archive. 2001. URL: https://www.sec.gov/rules/concept/ic-25258.htm.

[2] I. Ben-David, F. Franzoni, and R. Moussawi. "Exchange-Traded Funds". In: *Annual Review of Financial Economics* 9 (2017), pp. 169–189. DOI: 10.1146/annurev-financial-110716-032538.

[3] A. Petajisto. "Inefficiencies in the Pricing of Exchange-Traded Funds". In: *Financial Analysts Journal* 73.1 (2017), pp. 24–54. DOI: 10.2469/faj.v73.n1.7.

[4] B. R. Marshall, N. H. Nguyen, and N. Visaltanachoti. "ETF Arbitrage: Intraday Evidence". In: *Journal of Banking & Finance* 37.9 (2013), pp. 3486–3498. DOI: 10.1016/j.jbankfin.2013.05.014.

[5] Fidelity. *Understanding an ETF's NAV*. 2026. URL: https://www.fidelity.com/learning-center/investment-products/etf/etfs-nav.

[6] Fidelity. *Understanding an ETF's iNAV*. 2026. URL: https://www.fidelity.com/learning-center/investment-products/etf/etfs-inav.

[7] D. C. Brown, S. W. Davies, and M. C. Ringgenberg. "ETF Arbitrage, Non-Fundamental Demand, and Return Predictability". In: *Review of Finance* 25.4 (2021), pp. 937–972. DOI: 10.1093/rof/rfaa027.

[8] L. M. Calcagnile, F. Corsi, and S. Marmi. "Entropy and Efficiency of the ETF Market". In: *Computational Economics* 55.1 (2020), pp. 143–184. DOI: 10.1007/s10614-019-09885-z.

[9] R. F. Engle and D. Sarkar. "Premiums–Discounts and Exchange-Traded Funds". *The Journal of Derivatives* 13.4 (2006), pp. 27–45. DOI: 10.3905/jod.2006.635418.

[10] J. Hasbrouck. *Empirical Market Microstructure: Economic and Statistical Perspectives on the Dynamics of Trade in Securities Markets*. New York: Oxford University Press, 2007.

[11] E. Bacry, I. Mastromatteo, and J.-F. Muzy. "Hawkes Processes in Finance". In: *Market Microstructure and Liquidity* 1.1 (2015), p. 1550005. DOI: 10.1142/S2382626615500057.

[12] R. Cont, S. Stoikov, and R. Talreja. "A Stochastic Model for Order Book Dynamics". In: *Operations Research* 58.3 (2010), pp. 549–563. DOI: 10.1287/opre.1090.0780.

[13] J. D. Farmer, P. Patelli, and I. I. Zovko. "The Predictive Power of Zero Intelligence in Financial Markets". In: *Proceedings of the National Academy of Sciences* 102.6 (2005), pp. 2254–2259. DOI: 10.1073/pnas.0409157102.

[14] R. Roll. “A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market”. In: *Journal of Finance* 39.4 (1984), pp. 1127–1139. DOI: 10.1111/j.1540-6261.1984.tb03897.x.

[15] R. E. Kalman. “A New Approach to Linear Filtering and Prediction Problems”. In: *Journal of Basic Engineering* 82.1 (1960), pp. 35–45. DOI: 10.1115/1.3662552.

[16] T. Chen and C. Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.

[17] J. Sirignano and R. Cont. “Universal Features of Price Formation in Financial Markets: Perspectives from Deep Learning”. In: *Quantitative Finance* 19.9 (2019), pp. 1449–1459. DOI: 10.1080/14697688.2019.1622295.

[18] S&P Dow Jones Indices LLC. *S&P 500 [SP500]*. Retrieved from FRED, Federal Reserve Bank of St. Louis. 2026. URL: https://fred.stlouisfed.org/series/SP500 (visited on Mar. 14, 2026).