Short-Term Stock Price Modeling Using Observable Financial Factors

Nikolaos Halidias

Actuarial-Financial Mathematics Lab
Department of Statistics and Actuarial - Financial Mathematics

March 21, 2026

Objective

We consider the short-term stock price as a function of:

- the current time,
- past stock prices,
- and a collection of measurable market variables.

Our goal is to model the price by

$$S_t = F\left(t, S_{t_0}, S_{t_1}, \ldots, S_{t_{k-1}}, X_t\right),$$

and then approximate the unknown function F using sigmoidal functions.

Basic Modeling Assumption

Let

$$t_0 < t_1 < \cdots < t_{k-1} < t,$$

and let S_t denote the stock price at time t. We assume that

$$S_t = F\left(t, S_{t_0}, S_{t_1}, \ldots, S_{t_{k-1}}, X_t\right),$$

where $X_t \in \mathbb{R}^m$ is a vector of observable financial factors.
Thus, the current price is not determined only by its own history, but also by market microstructure and macro-financial information available at time t.

Observable Financial Factors

A typical factor vector X_t may include:

- trading volume,
- realized volatility,
- implied volatility,
- bid-ask spread,
- order imbalance,
- market index return,
- sector index return,
- short-term interest rate,
- news sentiment score,
- analyst revision score.

Hence X_t collects variables that are measurable and economically meaningful.

A More Explicit Factor Vector

For example, we may define

$$X_t = \left(V_t, \sigma_t^{\text{real}}, \sigma_t^{\text{impl}}, \text{Spread}_t, \text{OI}_t, R_t^{\text{mkt}}, R_t^{\text{sec}}, r_t, \text{Sent}_t\right),$$

where:

- V_t: trading volume,
- σ_t^{real}: realized volatility,
- σ_t^{impl}: implied volatility,
- Spread_t: bid-ask spread,
- OI_t: order imbalance,
- R_t^{mkt}: market return,
- R_t^{sec}: sector return,
- r_t: short-term rate,
- Sent_t: sentiment indicator.

Extending the Model to n Stocks

Consider now n stocks and define the price vector

$$\mathbf{S}_t = \left(S_t^{(1)}, S_t^{(2)}, \ldots, S_t^{(n)}\right)^\top.$$

We assume that the market evolution is described by

$$\mathbf{S}_t = F\left(t, \mathbf{S}_{t_0}, \mathbf{S}_{t_1}, \ldots, \mathbf{S}_{t_{k-1}}, X_t, \Sigma_t\right),$$

where:

- X_t contains observable market factors,
- Σ_t is the covariance matrix of stock returns at time t.

More explicitly,

$$\Sigma_t = \left(\sigma_{ij}(t)\right)_{1 \le i,j \le n}, \qquad \sigma_{ij}(t) = \mathrm{Cov}\left(R_t^{(i)}, R_t^{(j)}\right),$$

where $R_t^{(i)}$ denotes the return of stock i.

Thus, the model incorporates not only the dynamics of each stock, but also their comovements and mutual dependence.

Economic Interpretation of the Factors

These variables capture different mechanisms:

- **Volume** measures trading activity and information arrival.
- **Volatility** reflects uncertainty and risk perception.
- **Spread** measures liquidity and transaction cost.
- **Order imbalance** reflects short-run pressure from buyers and sellers.
- **Market and sector returns** capture systematic comovement.
- **Sentiment** captures the reaction of the market to news and opinions.

Therefore, the model combines price history with observable forces driving short-term price formation.

Feature Vector

Define the input vector

$$x_t = \left(t, S_{t_0}, S_{t_1}, \ldots, S_{t_{k-1}}, X_t\right) \in \mathbb{R}^d.$$

Then the model becomes simply

$$S_t = F(x_t).$$

Given data pairs (x_t, S_t), the problem is to estimate the unknown nonlinear map F.

Why a Nonlinear Model?

A linear specification is often too restrictive in finance, because:

- the impact of volume may saturate,
- volatility may affect price differently across regimes,
- spread effects may become important only beyond a threshold,
- sentiment shocks may have asymmetric effects.

Hence a nonlinear approximation is more appropriate for short-term market dynamics.

Sigmoidal Approximation

We approximate F by

$$\widehat{F}(x) = c + \sum_{j=1}^{M} \alpha_j \, \sigma\left(w_j^\top x + b_j\right),$$

where

$$\sigma(u) = \frac{1}{1 + e^{-u}}$$

is the sigmoid function.
Here:

- $w_j \in \mathbb{R}^d$ are weight vectors,
- $b_j \in \mathbb{R}$ are biases,
- $\alpha_j \in \mathbb{R}$ are output coefficients.

Financial Meaning of Sigmoidal Units

Each unit

$$\sigma(w_j^\top x + b_j)$$

can be viewed as a detector of a market regime or pattern.
For instance, a sigmoidal unit may respond strongly when:

- volume is unusually high,
- volatility crosses a threshold,
- spread widens significantly,
- order imbalance becomes strongly positive or negative,
- market sentiment turns sharply bullish or bearish.

Thus, the model can represent threshold-type financial effects in a smooth way.

Training Data

Suppose we observe

$$\{(x_i, y_i)\}_{i=1}^{n}, \qquad y_i = S_{t_i}.$$

The data may be collected at:

- intraday frequency,
- daily frequency,
- or another short-term horizon.

For each observation time t_i, we construct:

- lagged prices,
- contemporaneous observable factors,
- and the target value S_{t_i}.

Parameter Estimation

We estimate the parameters by minimizing a loss function such as

$$\mathcal{L}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \widehat{F}_\theta(x_i) \right)^2 .$$

Optionally, we add regularization:

$$\mathcal{L}_\lambda(\theta) = \mathcal{L}(\theta) + \lambda \|\theta\|^2.$$

This helps reduce overfitting, which is particularly important in noisy financial data.

Practical Data Pipeline

A possible workflow is:

1. collect price and market-factor data,
2. align all variables at a common time scale,
3. clean and normalize the variables,
4. construct lagged-price features,
5. fit the sigmoidal approximation,
6. evaluate out-of-sample performance.

Typical evaluation criteria include:

- RMSE,
- MAE,
- directional accuracy,
- sign prediction of short-term returns.

Example of a Concrete Specification

A concrete model may take the form

$$S_t = F\Big(t, S_{t-1}, S_{t-2}, S_{t-3}, V_t, \sigma_t^{\mathrm{real}}, \mathrm{Spread}_t, \mathrm{OI}_t, R_t^{\mathrm{mkt}}, \mathrm{Sent}_t\Big).$$

Then we approximate it by

$$\widehat{S}_t = c + \sum_{j=1}^{M} \alpha_j\, \sigma\Big(w_j^\top x_t + b_j\Big).$$

This already provides a flexible short-term predictive structure.

Remarks and Limitations

Some important issues remain:

- short-term prices are highly noisy,
- the relevance of factors may change over time,
- market regimes may shift suddenly,
- omitted variables may bias the approximation,
- prediction quality may deteriorate in stressed markets.

Therefore, rolling estimation and robust validation are often necessary.

Conclusion

We model the stock price as

$$S_t = F\left(t, S_{t_0}, \ldots, S_{t_{k-1}}, X_t\right),$$

where X_t contains measurable financial factors such as volume, volatility, spread, order imbalance, and sentiment.
Using data, we approximate the unknown function F by

$$\widehat{F}(x) = c + \sum_{j=1}^{M} \alpha_j \, \sigma(w_j^\top x + b_j).$$

This framework combines:

- financial observability,
- nonlinear approximation,
- and practical data-driven estimation.

References

N. Halidias,
Feasible Dynamic Hedging and Data-Driven Valuation of Options in Discrete Time,
Preprint, March 2026.
DOI: 10.13140/RG.2.2.21337.58728.
ResearchGate link.