Price Impact Models and Applications

Introduction to Algorithmic Trading

Kevin Webster

Spring 2023

Columbia University

Plan

Last Module
Introduction to Algorithmic Trading

For this Week

(a) Mathematical setup for price impact modeling
(b) The Obhizaeva and Wang (OW) model
(c) Closed form trading strategy
(d) Trader intuition for the strategy
(e) Implications for alpha research

Next Week
Empirical estimates of price impact.

Last Module's Summary

Trades capture alpha and pay impact.

(a) Price impact captures price moves *caused* by trading.

(b) Alpha signals predict price moves *independent* of trading.

Model-driven trading algorithms (Module 2)
Quantify the trade-off between alpha and impact using a *stochastic control problem.*

Measuring Live Performance (Module 3)
Performance measurement relies on AB testing. Furthermore, teams must rule out common trading biases using *causal inference.*

Last Module's Summary

Trades capture alpha and pay impact.

(a) Price impact captures price moves *caused* by trading.

(b) Alpha signals predict price moves *independent* of trading.

Model-driven trading algorithms (Module 2)
Quantify the trade-off between alpha and impact using a *stochastic control problem.*

Measuring Live Performance (Module 3)
Performance measurement relies on AB testing. Furthermore, teams must rule out common trading biases using *causal inference.*

Introduction

Refresher on Stochastic Control

Stochastic control extends stochastic calculus
Ingredients of a control problem:

(a) A filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P}\right)$

(b) A variable x you *control*

(c) A *state variable* X that x affects, e.g.,

$$dX_t = \mu(X_t, x_t)dt + \sigma(X_t, x_t)dW_t$$

(d) An *objective function* to maximize, e.g.,

$$\sup_x \mathbb{E}\left[\int_0^T g(X_t, x_t)dt + G(X_T)\right]$$

(e) Constraints, e.g., $X_t \geq 0$

Mathematical Setup for Trading with Impact

Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P}\right)$ **be a filtered probability space.**

(a) For $t \leq T$, let $\mathcal{S}(t)$ be the space of bounded semimartingales over $[0, t]$.

(b) Let $\mathcal{S} = \mathcal{S}(T)$ for simplicity.

(c) Denote $\mathcal{S}_{1,1}$ the space of functionals

$$F : [0, T] \times \Omega \times \mathcal{S} \to \mathcal{S}$$

such that for all $t \in [0, T]$

$$(t, \omega, Z) \to F_t(\omega, Z)$$

is $\mathcal{B}([0, S]) \otimes \mathcal{F}_t \otimes \mathcal{B}(\mathcal{S}(t)) / \mathcal{B}(\mathcal{S}(t))$-measurable.

A Brief Stochastic Calculus Refresher (1/2)

Itô's formula
Let X be an Itô process with volatility σ_t,

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)\sigma_t^2 dt.$$

Example
Let

$$dX_t = \mu dt + \sigma dW_t$$

and $Y_t = e^{X_t}$. Then,

$$dY_t = Y_t dX_t + \frac{1}{2}Y_t\sigma^2 dt.$$

A Brief Stochastic Calculus Refresher (2/2)

Integration by parts
Let X, Y be two semimartingales. Then,

$$d(Y_t X_t) = Y_t dX_t + X_t dY_t + d[X, Y]_t.$$

where all the integrands use their left limit.

If X, Y are Itô processes with volatilities σ_t, θ_t and correlation ρ_t,

$$[X, Y]_t = \int_0^t \rho_u \sigma_u \theta_u du.$$

For discrete processes,

$$[X, Y]_t = \sum_{u \leq t} \Delta_u X \Delta_u Y.$$

Essential Trading Variables

The trader controls their position $Q \in \mathcal{S}$.

The fundamental price $S \in \mathcal{S}$
captures price moves *not* caused by Q. Denote the trader's alpha signal by

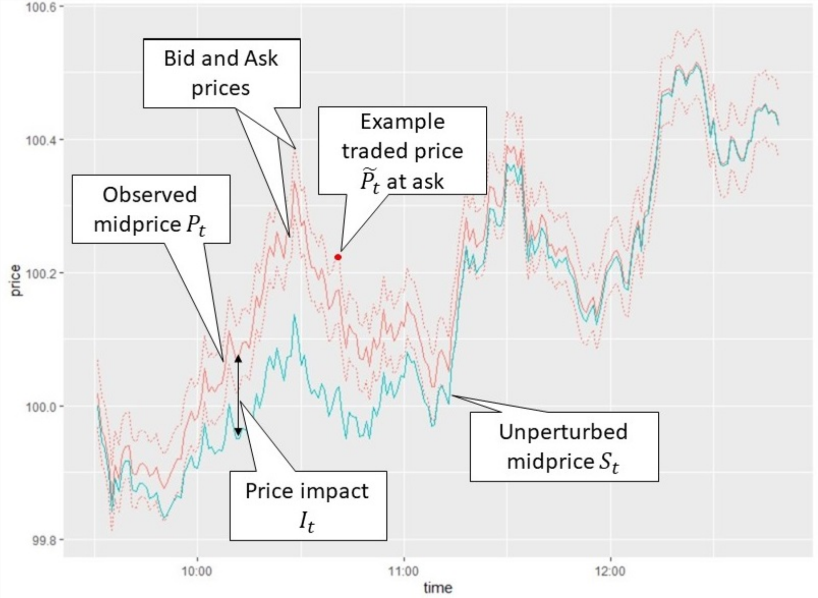
$$\alpha_t = \mathbb{E}\left[S_T - S_t \middle| \mathcal{F}_t \right].$$

We often assume α to be an Itô process with drift μ_t^α.

The observed price $P \in \mathcal{S}_{1,1}$
includes Q's causal effect. Denote the trader's price impact by

$$I(t, \omega, Q) = P(t, \omega, Q) - S(t, \omega).$$

Three Prices



Example Price Impact Models (1/2)*

The OW model
exponentially decays each trade's impact:

$$dI_t = -\beta I_t dt + \lambda dQ_t.$$

The generalized OW model
captures time-varying, stochastic liquidity parameters:

$$dI_t = -\beta_t I_t dt + \lambda_t dQ_t$$

with $\beta, \lambda \in \mathcal{S}$ and bounded away from zero.

Example Price Impact Models (2/2)*

The Alfonsi, Fruth, and Schied (AFS) model
adds a *global* nonlinearity:

$$I_t = g\left(J_t\right)$$

where

$$dJ_t = -\beta J_t dt + \lambda dQ_t$$

and g is typically a power-law (e.g. square-root).

Bouchaud's nonlinear model

$$\Delta I = -\beta I \Delta t + g(\Delta Q).$$

The Obhizaeva and Wang (OW) Model

Three Prices

Variables

(a) Q is our position and dQ are our trades.

(b) S is the price if we didn't trade, aka when $Q = 0$.

(c) $P = S + I(Q)$ is the price when we trade. $I(Q)$ is the price impact of Q.

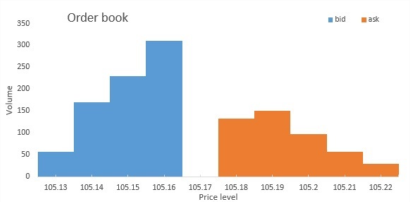
(d) $\tilde{P} = P + \frac{s}{2}\text{sign}(\Delta Q)$ is the traded price, where s is the bid-ask spread.

Refresher: The Limit Order Book

The limit order book represents instantaneous supply and demand.

(a) Impatient traders, liquidity takers, buy at the ask and sell at the bid.

(b) Patient traders, liquidity providers, buy at the bid and sell at the ask.



The spread is (ask - bid)/2.
Liquidity takers pay the spread in addition to impact costs. Spread costs average 5bps, impact costs 30bps.

The OW limit order book*

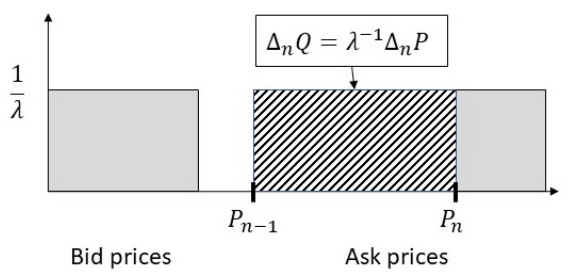


Figure 2: Fill execution under the OW limit order book.

Moving the ask by ΔP
requires a fill size $\Delta Q = \lambda^{-1}\Delta P$.

Trading Equations under the OW model

The trade ΔQ moves the price
After the fill, the price is now

$$P_n = P_{n-1} + \lambda \Delta_n Q$$

due to the limit order book's shape.

The traded price is the VWAP over the book*
The fill achieves the *average* traded price

$$\begin{aligned}\tilde{P}_{n-1} &= P_{n-1} + \frac{\lambda}{2}\Delta_n Q \\ &= \frac{1}{2}\left(P_n + P_{n-1}\right) \\ &= P_{n-1} + \frac{1}{2}\Delta_n I\end{aligned}$$

One extends this formula for general order book shapes. The OW model assumes the order book is "flat".

Exponential Decay

In between trade times, the impact state relaxes towards 0.

$$I_n = -\beta I_{n-1}\delta t + \lambda \Delta_n Q$$

where δt is the time between two trade times.

Link with continuous time exponentials
In the absence of trading,

$$I_n = (1 - \beta\delta t)^n I_0.$$

As $\delta t \to 0$, this discrete model converges to

$$I_t = e^{-\beta t} I_0$$

when there is no active trading over $[0, t]$.

Wealth Equations under the OW model (1/4)

Definition of Wealth
One defines a portfolio's *fundamental wealth* as

$$Y_N = S_N Q_N - \sum_{n \leq N} \tilde{P}_{n-1} \Delta_n Q.$$

(1) $Q_N S_N$ is a position's value marked at the fundamental price.

(2) $\tilde{P}_{n-1} \Delta_n Q$ is the cash paid (or received) from a single fill.

Mark-to-market Wealth
One defines a portfolio's *mark-to-market wealth* as

$$X_N = {\color{red} P_N} Q_N - \sum_{n \leq N} \tilde{P}_{n-1} \Delta_n Q.$$

In particular,

$$X_N = Y_N + I_N Q_N.$$

Wealth Equations under the OW model (2/4)*

Why two notions of wealth?
We'll cover this topic in depth in week 9. At a high level

(a) The fundamental P&L Y is your "true" P&L, matches your trading algorithm's arrival slippage, but carries model risk.
(b) The mark-to-market P&L X is your "perceived" P&L, matches your accounting system, and carries no model risk.
(c) They are equal once you close your position: $X = Y$ when $Q = 0$.

See Caccioli et al. (2012, link) and Kolm and Webster (2023, link) for more details.

Wealth Equations under the OW model (3/4)*

Discrete differentiation yields

$$\begin{aligned}\Delta_n Y &= S_{n-1}\Delta_n Q + Q_{n-1}\Delta_n S + \Delta_n S \Delta_n Q - \tilde{P}_{n-1}\Delta_n Q \\ &= Q_{n-1}\Delta_n S + \underbrace{\Delta_n S \Delta_n Q}_{=0} + \underbrace{\left(S_{n-1} - \tilde{P}_{n-1}\right)}_{=-I_{n-1}-\frac{1}{2}\Delta_n I}\Delta_n Q \\ &= Q_{n-1}\Delta_n S - I_{n-1}\Delta_n Q - \frac{1}{2}\Delta_n I \Delta_n Q.\end{aligned}$$

In continuous time

$$dY_t = Q_t dS_t - I_t dQ_t - \frac{1}{2} d[I, Q]_t.$$

Wealth Equations under the OW model (4/4)*

Expected wealth equation

$$\begin{aligned}\mathbb{E}\left[Y_T\right] &= \mathbb{E}\left[\int_0^T Q_t dS_t - \int_0^T I_t dQ_t - \frac{1}{2}[I, Q]_T\right] \\ &= \mathbb{E}\left[-\int_0^T Q_t d\alpha_t - \int_0^T I_t dQ_t - \frac{1}{2}[I, Q]_T\right]\end{aligned}$$

where

$$\alpha_t = \mathbb{E}\left[\, S_T - S_t \middle|\, \mathcal{F}_t\right].$$

Then, using integration by parts,

$$\mathbb{E}\left[Y_T\right] = \mathbb{E}\left[\int_0^T \left(\alpha_t - I_t\right) dQ_t - \frac{1}{2}[I, Q]_T + [\alpha, Q]_T\right].$$

OW Impact Model in Continuous Time

Differential form
Quants use

$$dI_t = -\beta I_t dt + \lambda dQ_t$$

to derive optimal strategies.

Kernel form
Quants use

$$I_t = e^{-\beta t} I_0 + \lambda \int_0^t e^{-\beta(t-s)} dQ_s$$

to compute closed-form I_t for specific Q (e.g., constant speed).

Kernel Form Proof

Variation of the constant
Let

$$\tilde{I}_t = e^{\beta t} I_t.$$

Then,

$$\begin{aligned} d\tilde{I}_t &= \beta e^{\beta t} I_t dt + e^{\beta t} dI_t \\ &= \beta e^{\beta t} I_t dt + -\beta e^{\beta t} I_t dt + \lambda e^{\beta t} dQ_t \\ &= \lambda e^{\beta t} dQ_t \end{aligned}$$

and, hence,

$$\begin{aligned} \tilde{I}_t &= I_0 + \lambda \int_0^t e^{\beta s} dQ_s \\ e^{\beta t} I_t &= I_0 + \lambda \int_0^t e^{\beta s} dQ_s. \end{aligned}$$

Numerical Implementation (1/2)

General time stamps
Quants numerically implement

$$I_{n+1} = e^{-\beta(t_{n+1}-t_n)} I_n + \lambda \Delta_n Q$$

for arbitrary time stamps t_n (e.g., fill times).

Regular time grid
On a regular grid $\delta t = t_{n+1} - t_n \ll \beta^{-1}$, one recovers

$$I_{n+1} \approx I_n - \beta I_n \delta t + \lambda \Delta_n Q.$$

Numerical Implementation (2/2)

Regular time grid in kdb/python/R
The regular time grid formula is a multiple of an *exponential moving average.* Most quant libraries have an efficient implementation

```
tbl: update impact: (lambda % beta) * ema[decay; trades] by
    stock from tbl.
```

General time stamps
The formula

$$I_{n+1} = e^{-\beta(t_{n+1}-t_n)} I_n + \lambda \Delta_n Q$$

is a *general recursive equation.* Naive implementations are inefficient. In Q, the pattern

```
x y\z
```

implements a general recursive formula at the C level. See `code.kx.com/q/ref/accumulators`.

Trading Strategies

The Control Problem

Summary of the OW setup
Given the price impact model

$$dI_t = -\beta I_t dt + \lambda dQ_t$$

a trader optimizes

$$\max_{Q} \mathbb{E}\left[\int_0^T (\alpha_t - I_t)\, dQ_t - \frac{1}{2}[I, Q]_T + [\alpha, Q]_T\right]$$

where

$$\alpha_t = \mathbb{E}\left[S_\tau - S_t \middle| \mathcal{F}_t\right]$$

is their alpha over horizon $\tau \geq T$.

In words,
each fill dQ captures alpha α_t and pays impact I_t (plus Itô terms).

Special Case: Long-Term Alpha

For a long-term signal

$$\alpha_t = \alpha$$

for a constant α representing expected price moves over $[t, \tau]$.

The optimal strategy*
is such that

$$\forall t \in (0, T), \quad I_t = \frac{1}{2}\alpha; \quad I_T = \alpha.$$

The next slides prove this relation.

Corollary
The trader pays half of their alpha in impact and captures the other half in profits.

Proof: Central Idea by Fruth et al. (2013)

Impact space
The equation

$$dI_t = -\beta I_t dt + \lambda dQ_t$$

is invertible:

$$dQ_t = \frac{1}{\lambda}\left(\beta I_t dt + dI_t\right).$$

Mathematically, it is equivalent to solve over choices of Q or I: there is a *one-to-one map*.

Trader interpretation
If you tell me your impact, I can back out your trades. Instead of optimizing a trading curve, I optimize for my impact on the market.

Proof in Impact Space

Replacing Q with I simplifies the problem.

$$
\begin{aligned}
&\frac{1}{\lambda}\int_0^T (\alpha - I_t)\left(dI_t + \beta I_t dt\right) - \frac{1}{2\lambda}[I, I]_T \\
&= \frac{1}{\lambda}\left(\int_0^T \beta(\alpha I_t - I_t^2)dt + \alpha\int_0^T dI_t - \int_0^T I_t dI_t - \frac{1}{2}[I, I]_T\right) \\
&= \frac{\beta}{\lambda}\int_0^T (\alpha I_t - I_t^2)dt + \frac{1}{\lambda}\left(\alpha I_T - \frac{1}{2}I_T^2\right).
\end{aligned}
$$

The last equality stems from integration by parts.

A myopic optimization problem

There are no dI_t terms left: one can optimize for each (t, ω) separately without taking into account the past or future. Hence,

$$\forall t \in (0, T), \quad I_t = \frac{1}{2}\alpha; \quad I_T = \alpha.$$

One-to-one map
Given the optimal impact state, one recovers the optimal holdings using

$$dQ_t^* = \frac{1}{\lambda}\left(\beta I_t^* dt + dI_t^*\right) = \frac{\beta}{2\lambda}\alpha dt.$$

Constant trading speed proportional to the long-term alpha, plus two jumps at $t = 0, T$.

Order size

$$Q_T = \frac{2 + \beta T}{2\lambda}\alpha.$$

Case of a Deterministic Alpha

Let $\alpha \in C^2$
and let α'_t, α''_t be its first two derivatives.

For example, if S has a constant drift μ over $[0, T]$,

$$\alpha_t = (T - t)\mu.$$

The optimal strategy
is such that

$$\forall t \in (0, T), \quad I_t = \frac{1}{2}\left(\alpha_t - \beta^{-1}\alpha'_t\right); \quad I_T = \alpha_T.$$

Intepretation:
Faster alpha decay $-\alpha'_t$ over the impact timescale β^{-1} requires faster trading.

Proof

Mapping to impact space

$$
\begin{aligned}
&\frac{1}{\lambda}\int_0^T (\alpha_t - I_t)\left(dI_t + \beta I_t dt\right) - \frac{1}{2\lambda}[I, I]_T \\
&= \frac{1}{\lambda}\left(\int_0^T \beta(\alpha_t I_t - I_t^2)dt + \int_0^T \alpha_t dI_t - \int_0^T I_t dI_t - \frac{1}{2}[I, I]_T\right) \\
&= \frac{\beta}{\lambda}\int_0^T (\alpha_t I_t - I_t^2)dt - \frac{1}{\lambda}\int_0^T I_t d\alpha_t + \frac{1}{\lambda}\left(\alpha_T I_T - \frac{1}{2} I_T^2\right) \\
&= \frac{\beta}{\lambda}\int_0^T (\alpha_t I_t - \beta^{-1}\alpha_t' I_t - I_t^2)dt + \frac{1}{\lambda}\left(\alpha_T I_T - \frac{1}{2} I_T^2\right).
\end{aligned}
$$

Recovering the Trading Strategy Q

One-to-one map

$$
\begin{aligned}
dQ_t^* &= \frac{1}{\lambda}\left(\beta I_t^* dt + dI_t^*\right) \\
&= \frac{1}{2\lambda}\left(\beta\alpha_t - \alpha_t' + \alpha_t' - \beta^{-1}\alpha_t''\right) dt \\
&= \frac{\beta}{2\lambda}\left(\alpha_t - \beta^{-2}\alpha_t''\right) dt.
\end{aligned}
$$

Interpretation
Trading speed is proportional to alpha, but one trades faster if the signal is concave.

Summary (1/2)

Trading algo
Trading algorithms turn impact and alpha signals into trades using a control problem

$$\max_{Q} \mathbb{E}\left[\int_0^T (\alpha_t - I_t)\, dQ_t - \frac{1}{2}[I, Q]_T + [\alpha, Q]_T\right].$$

The problem can be mapped in impact space
The following myopic problem is equivalent to the original trading algo problem:

$$\sup_{I} \mathbb{E}\left[\frac{\beta}{\lambda}\int_0^T \left(\alpha_t I_t - \beta^{-1}\alpha_t' I_t - I_t^2\right) dt + \frac{1}{\lambda}\left(\alpha_T I_T - \frac{1}{2} I_T^2\right)\right].$$

Summary (2/2)

Solution intuition
The solution is best expressed as a *target impact state:*

$$I_t = \frac{1}{2}\left(\alpha_t - \beta^{-1}\alpha_t'\right).$$

Therefore, traders aim for a linear relationship between impact, alpha and alpha decay. This relationship allows for non-parametric fitting methods, e.g., machine-learning for alpha.

Mapping the solution back to trade space
One recovers the trading strategy from the target impact using

$$Q_t = \frac{1}{\lambda} I_t + \frac{\beta}{\lambda}\int_0^t I_u du.$$

Trader Intuition

Trading as a Balancing Act

Nehren et al. (ADIA, 2020)

> *"If we find that most of the cost is due to price drift (alpha), the best option would be to accelerate the trading and pay more impact to capture more attractive prices. Conversely, if the cost is mostly driven by impact then it would behoove us to slow down our trading to minimize the impact."*

For the OW model,
this balancing act takes the form of a linear relationship between alpha, alpha decay, and price impact.

The Case Without Alpha Decay

How do I correctly size my order?

> *"As a common feature of linear-quadratic functions, the optimum is reached at the point where the quadratic penalties amount to one-half the linear term.* **This means that it is optimal to give away half of the forecast-driven gross pnl to impact cost***"*
> *(Isichenko, Bloomberg 2021)*

For small alpha decay, impact should be half the order's alpha.

We can significantly generalize this rule of thumb.
Including: decaying alpha, time-varying liquidity, and non-linear price impact.

A Simple Relation between Alpha, Decay, and Impact.

For the OW model

$$I_t^* = \frac{1}{2}\left(\alpha_t - \beta^{-1}\alpha_t'\right)$$

where $-\mu_t^\alpha$ captures the alpha's decay.

Gârleanu and Pedersen (AQR, 2013)

> *"The alpha decay is important because it determines how long the investor can enjoy high expected returns and, therefore, affects the trade-off between returns and transactions costs."*

More generally

$$I_t^* = f(\alpha_t, \alpha_t', \text{model parameters}).$$

The relation simplifies trading strategies' numerical implementation and quantifies the balance between alpha, alpha decay, and impact.

Bells and Whistles (1/3)

Square-root law
The Alfonsi-Fruth-Schied (AFS) model:

$$I_t = J_t^c$$

where

$$dJ_t = -\beta J_t dt + \lambda dQ_t.$$

Empirically, studies across all asset classes have found $c \approx 0.5$ for large trades and $c \approx 1$ for small trades.

Optimal trading strategy

$$I_t^* = \frac{1}{1+c}\left(\alpha_t - \beta^{-1}\alpha_t'\right).$$

Relationship between alpha and impact is still linear: only the relationship with Q is non-linear.

Square-root law with constant alpha
The optimal impact state targets

$$I_t^* = \frac{1}{1+0.5}\alpha = \frac{2}{3}\alpha$$

Refresher: long term price impact

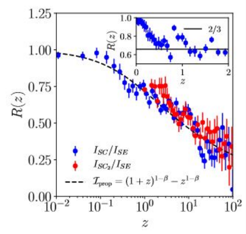


Figure 3: Multiday impact curve from Bucci et al. (2019).

Bells and Whistles (3/3)

Time-varying coefficients
The Fruth-Schoeneborn-Urusov (extended OW) model:

$$dI_t = -\beta_t I_t dt + e^{\gamma_t} dQ_t.$$

Optimal trading strategy

$$I_t^* = \frac{\beta_t + \gamma_t'}{2\beta_t + \gamma_t'}\alpha_t - \frac{1}{2\beta_t + \gamma_t'}\alpha_t'.$$

There is still a myopic, linear relationship between impact, alpha, and alpha decay. The ratios are involved but tractable.

Implications for Alpha Research

Corollary: Expected Profits

For the optimal strategy I^*,

$$\mathbb{E}\left[Y_T(I^*)\right] = \mathbb{E}\left[\int_0^T \frac{\beta}{\lambda}\left(I_t^*\right)^2 dt\right].$$

For any other strategy I,

$$\mathbb{E}\left[Y_T(I)\right] = \mathbb{E}\left[Y_T(I^*)\right] - \mathbb{E}\left[\int_0^T \frac{\beta}{\lambda}\left(I_t^* - I_t\right)^2 dt + \frac{1}{2\lambda} I_T^2\right].$$

Proof on the blackboard.

Implications for Alpha Research (1/3)

Traders want to maximize their strategy's P&L

$$\sup_{\alpha} \mathbb{E}\left[Y_T(I(\alpha))\right]$$

where $I(\alpha)$ is the optimal strategy if α is correct.

P&L regret function
Maximizing the trader's P&L is equivalent to minimizing their P&L *regret*

$$\inf_{\alpha} \left\{\mathbb{E}\left[Y_T(I(\alpha^r))\right] - \mathbb{E}\left[Y_T(I(\alpha))\right]\right\}$$

where $\alpha_t^r = S_T - S_t$ are the realized returns the alpha signal predicts.

Implications for Alpha Research (2/3)

Naive regression
applies a mean-squared error (MSE) for alpha

$$\sum_t w_t \left(\alpha_t^r - \alpha_t\right)^2$$

where w_t is a weighting scheme (e.g., equal-weighted).

P&L-aware regression minimizes P&L regret

$$\mathbb{E}\left[Y_T(I(\alpha^r))\right] - \mathbb{E}\left[Y_T(I(\alpha))\right] =$$
$$\mathbb{E}\left[\int_0^T \frac{\beta}{\lambda}\left(I_t(\alpha^r) - I_t(\alpha)\right)^2 dt + \frac{1}{2\lambda} I_T^2(\alpha)\right].$$

Therefore, maximizing P&L is the same as minimizing MSE for impact.

Implications for Alpha Research (3/3)

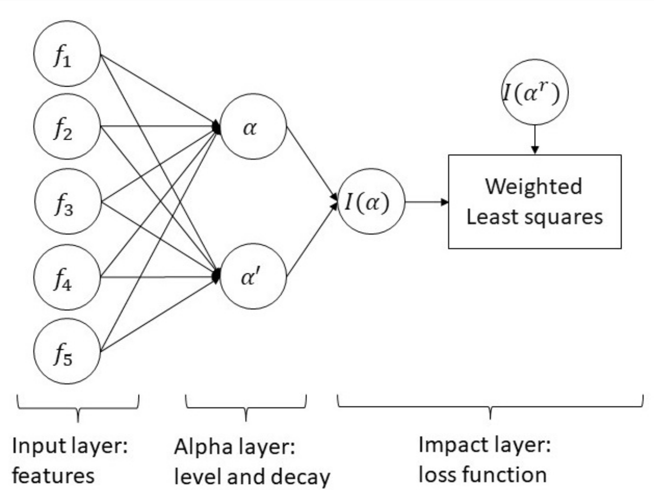


Figure 4: Implementation of the anticipated P&L regret function in a layered machine learning architecture allowing for backpropagation.

Weekly Summary

Given a trading strategy Q, there are three prices

(a) the unobserved, unperturbed price S,

(b) the observed price $P = S + I(Q)$, and

(c) the transaction price $\tilde{P} = P \pm s$

The OW price impact model is an exponential kernel of past trades.

$$dI_t = -\beta I_t dt + \lambda dQ_t$$

The optimal strategy is best expressed in impact space

$$\forall t \in (0, T) \quad I_t = \frac{1}{2}\left(\alpha_t - \beta^{-1}\alpha_t'\right); \quad I_T = \alpha_T$$

where

$$\alpha_t = \mathbb{E}\left[S_T - S_t \middle| \mathcal{F}_t\right].$$

Questions?

Next week
Empirical results on price impact models.