Price Impact Models and Applications

Introduction to Algorithmic Trading

Kevin Webster

Spring 2023

Columbia University

Plan

Last Week
We solved the Obizhaeva and Wang (OW) model to translate alpha signals into trades.

For this Week

(a) Refresher on dealing with data.
(b) Empirical results on price impact models.
(c) Alpha model based on Order Flow Imbalance (OFI).

Next Week
Dynamic liquidity models and no price manipulation results.

Last Week's 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

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

Introduction

Two types of trading datasets

The public trading tape
Includes all public fills.

(a) the dataset is massive and ideal for non-parametric fitting.
(b) it doesn't identify traders, orders, or their alphas.
(c) dominated by tiny, low alpha orders.

Proprietary order data
Captures proprietary information: when an order starts and ends, and maybe its predicted alpha.

(a) may not have enough data to fit high-dimensional models.
(b) required to estimate trading costs for sizable or high alpha orders.
(c) tackles common trading biases more effectively (see module 3).

Non-Exhaustive List of Datasets and Studies (1/3)

Proprietary, order-level data
E.g., Almgren et al. (2005), Bershova and Rhaklin (2013), Bouchaud et al. (2015), Tóth, Eisler, and Bouchaud (2017), Frazzini, Israel, and Moskowitz (2018)

Public, fill-level data
E.g., Bouchaud et al. (2004), Cont, Kukanov, and Stoikov (2013), Chen, Horst, and Hai Tran (2019)

Non-Exhaustive List of Datasets and Studies (2/3)

Most asset classes

(a) US stocks (e.g., Bouchaud et al. (2016), Cont, Cucuringu, and Zhang (2021))

(b) foreign stocks (e.g., Bouchaud et al. 2004 , Zhou (2012))

(c) fixed income (e.g., CGFS (2016), Schneider and Lillo (2019), Tomas, Matromatteo, and Benzaquen (2022))

(d) options (e.g., Said et al. (2021), Kaeck, van Kervel, and Seeger (2021), Tomas, Mastromatteo, and Benzaquen (2022))

(e) cryptocurrencies (e.g., Donier and Bonart (2015))

Non-Exhaustive List of Datasets and Studies (3/3)

Various timescales

(a) tick timescale (e.g., Bouchaud et al. (2004), Busseti and Lillo (2012), Eisler, Bouchaud, and Kockelkoren (2018))

(b) hour to day timescale (e.g., Lillo and Farmer (2004), Bouchaud, Kockelkoren, and Potters (2006), Bershova and Rhaklin (2013), Chen, Horst, and Hai Tran (2019))

(c) weeks to months timescale (e.g. Caccioli et al. (2012))

General Lobster Methodology

Context

Data Pre-Processing
is an essential skill set for quants and traders.

* Trading teams have clean, preprocessed data for their core functions.
* However, new research or deeper debugging, requires quants to look at raw data.

The Class Samples
are already binned. To look at sample "raw" lobster data, see https://lobsterdata.com/tradesquotesandprices.

Reference Paper
Cont, Kukanov, and Stoikov (2013) details a reproducable data pre-processing methodology.

Binning (1/2)

Bin Definition
Bin the data using a bin-size $\delta t = 10s$. Each interval has a start and end time. Is the bin *forward* or *backward* looking?

Define variables through aggregatation functions.
E.g.: first, last, sum, mean, min, max...

q_t sums the signed, traded volume over interval $[t, t + \delta t)$.

Label clearly forward and backward variables.
Databases may separate the two and forbid forward variables and leading functions (e.g. next, shift) to alpha researchers.

Binning (2/2)

Example

Time	Stock	Trade	Midprice	Spread
9:30:01	AAPL	-70	154.77	0.18
9:30:03	AAPL	100	154.74	0.16
9:30:07	AAPL	-100	154.705	0.10
9:30:12	AAPL	234	154.705	0.025
9:30:13	AAPL	-10	154.78	0.02

TimeBin	Stock	Trade	MidStart	Spread
9:30:00	AAPL	-70	154.77	0.18
9:30:10	AAPL	224	154.705	0.10

Binned Table Schema (1/3)

Start and End Variables

* Start: mid, spread, depth (best bid and ask volume), lobImb ((best bid - ask volume)/depth).
* End: midEnd.

Bins are *forward* looking: 09:30:00 corresponds to interval [09:30:00, 09:30:10).

date	time	stock	mid	midEnd	spread	lobImb	depth
1/2/2019	9:32:00	A	66.335	66.28	0.025	0	303.5714
1/2/2019	9:32:10	A	66.47	66.415	0.11	-0.97044	252.25
1/2/2019	9:32:20	A	66.39	66.385	0.09	0.818182	218.8333
1/2/2019	9:32:30	A	66.255	66.245	0.045	0.904762	262
1/2/2019	9:32:40	A	66.235	66.24	0.035	-0.11111	282
1/2/2019	9:32:50	A	66.175	66.215	0.045	0.333333	420
1/2/2019	9:33:00	A	66.2	66.235	0.07	0	242.8571

Binned Table Schema (2/3)

Sum Variables

* Signed volumes: trade, orderFlow (lob events), hidden, auction.
* Unsigned volumes: trdLiq, ofLiq (lob events).
* Counts: nbTrades, nbEvents, nbHidden.

date	time	stock	trade	orderFlow	hidden	auction	trdLiq	ofLiq	depth	nbEvents	nbHidden	nbTrades
1/2/2019	9:32:00	A	100	1113	143	0	700	4477	303.5714	7	1	2
1/2/2019	9:32:10	A	0	300	0	0	0	512	252.25	4	0	0
1/2/2019	9:32:20	A	-203	-213	0	0	403	3669	218.8333	6	0	2
1/2/2019	9:32:30	A	-200	-300	0	0	200	940	262	5	0	1
1/2/2019	9:32:40	A	-320	-630	0	0	320	2030	282	5	0	1
1/2/2019	9:32:50	A	102	100	0	0	102	1900	420	5	0	1
1/2/2019	9:33:00	A	200	500	100	0	400	2900	242.8571	7	1	2

Binned Table Schema (3/3)

Statistics Variables

* effSpread (average spread paid per trade).
* effLobImb (average order book imbalance before each trade).

date	time	stock	effSpread	effLobImb
1/2/2019	9:32:00	A	0.043571	0.114286
1/2/2019	9:32:10	A		
1/2/2019	9:32:20	A	0.077556	0
1/2/2019	9:32:30	A	0.035	0.333333
1/2/2019	9:32:40	A	0.04	0.333333
1/2/2019	9:32:50	A	0.065	-0.33333
1/2/2019	9:33:00	A	0.05625	-0.08333

Horizons (1/4)

Why different regression horizons?
Different trading strategies operate on different timescales.

* E.g. a one minute strategy focuses on one minute alpha forecasts and price impact: the first minute of the day may require a different model from the middle of the day.
* E.g. a two hour forecast will be less sensitive to the "time of the day". Strategies on two hours send larger orders: concavity of price impact for sizable orders may play a larger role.

Horizons (2/4)

Bad approach to generate horizons
Computing variables as aggregation functions over the horizon: this loops over the data multiple times!

time	return
09:30	10bps
09:35	-5bps
09:40	3bps
09:45	8bps
09:50	-1bps
09:55	-8bps
10:00	-13bps
10:05	2bps

10min returns

time	return
09:30	10bps
09:35	-5bps
09:40	3bps
09:45	8bps
09:50	-1bps
09:55	-8bps
10:00	-13bps
10:05	2bps

15min returns

time	return
09:30	10bps
09:35	-5bps
09:40	3bps
09:45	8bps
09:50	-1bps
09:55	-8bps
10:00	-13bps
10:05	2bps

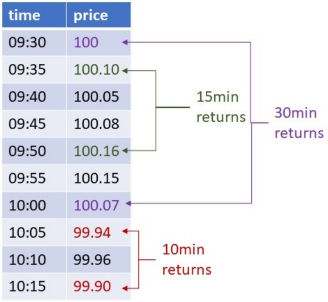
30min returns

Horizons (3/4)

Good approach to generate horizons

(a) Compute cumulative variables, e.g., prices, impact, cumulative volumes.

(b) For each horizon, compute horizon-specific difference variables.



time	price
09:30	100
09:35	100.10
09:40	100.05
09:45	100.08
09:50	100.16
09:55	100.15
10:00	100.07
10:05	99.94
10:10	99.96
10:15	99.90

Simple shift in Python Pandas

```
df[’ret10min’] = df[’price’].shift(2) % df[’price’] - 1
df[’ret30min’] = df[’price’].shift(6) % df[’price’] - 1
```

instead of expensive aggregations

```
df[’ret10min’] = df[’ret’].groupby(’10min’).aggregate(’cumulateRet’)
df[’ret30min’] = df[’ret’].groupby(’30min’).aggregate(’cumulateRet’)
```

Horizons (4/4)

```
\optional step if you aren't sure the grid is sorted by time
    . Sorting by date, stock accelerates the grouping.
tbl: `date`stock`time xasc tbl;

\xprev[n; ] shifts data by n steps. Negative n look into the
     future.

\For grids, we look ahead a fixed number of steps on the
    grid within a given (date, stock) group. This assumes
    the grid is sorted by time!
tbl: update retEod: (neg 1) + last[mid]%mid,
            ret1min: (neg 1) + xprev[-6; mid]%mid,
            ret5min: (neg 1) + xprev[-30; mid]%mid,
            ret30min: (neg 1) + xprev[-180; mid]%mid,
            ret60min: (neg 1) + xprev[-360; mid]%mid,
            by date, stock from tbl;
```

Sampling (1/3)

Why separate the data into samples?
Rigorous quantitative research requires three sample types:

(a) **Training data:** The sample used to fit the model parameters.

(b) **Testing data:** If the model has meta-parameters (e.g. regularization penalties), used to calibrate the meta-parameters.

(c) **Validation data:** Also called *out-of-sample* data.

Sampling (2/3)

Example: training data
For a **given** meta parameter λ,

$$\min_{\beta \in \mathbb{R}^n} \left\{ \sum_{t \in \mathcal{S}^{\text{train}}} (y_t - \beta x_t)^2 + \lambda ||\beta||^2 \right\}$$

fits the parameter $\beta(\lambda)$. This is a linear regression with a ridge regularization.

Example: testing
For a given trained family $\beta(\lambda)$,

$$\min_{\lambda \in \mathbb{R}_+} \left\{ \sum_{t \in \mathcal{S}^{\text{test}}} (y_t - \beta(\lambda) x_t)^2 + \lambda ||\beta(\lambda)||^2 \right\}$$

fits the meta-parameter λ.

Sampling (3/3)

Example: Validation
One estimates the out-of-sample mean-square error (MSE) and R^2 using

$$\sum_{t \in \mathcal{S}^{\text{val}}} (y_t - \beta x_t)^2$$

Regularization always reduces in-sample MSE but often increases out-of-sample MSE by reducing over-fitting.

Train	*Test*	Validate
201901	201902	201903
⋮	⋮	⋮
201910	201911	201912
β	λ	oos R^2

Review of Price Impact Models

General Model Shape (1/2)

Recurrence Equation

$$I_{t+\delta t} - I_t = f(I_t, q_t, ...)$$

allows for an efficient and general implementation.

For example, Python implements its potent exponential moving average (EMA) with a recurrence equation.

Solution to avoid

$$I_t = f(q_t, q_{t-\delta t}, ..., q_{t-n\delta t}, ...)$$

is overly memory and compute intensive. For instance, in the OW model,

$$I_t = \sum_{j \geq 0} e^{-j\beta\delta t} q_{t-j\delta t}$$

is inefficient.

General Model Shape (2/2)

Parameter Normalization
Pay attention to units. Ideally, model parameters and meta-parameters should be unit-less. Try to find normalizing variables.

Example from Almgren (2005)

$$I_t = \sigma f(q_t/ADV, ...)$$

with σ the stock's daily volatility and ADV its average daily volume, makes

* argument q/ADV is a percentage of daily volume.
* impact is measured in standard deviations of price changes.

Therefore, the model becomes comparable across stocks.

OW Model

The baseline model

$$I_{t+\delta t} - I_t = -\beta I_t \delta t + \lambda\sigma \frac{q_t}{\mathrm{ADV}}.$$

is an EMA on $\lambda\sigma q_t / ADV$.

σ, ADV can be estimated from historical data or can come from predictive models such as risk and volume models. Therefore, β, λ are the only parameters left to estimate.

References
Obizhaeva and Wang (2013), Chen, Horst, and Hai Tran (2019).

Propagagtor Model

The nonlinear Propagator model

$$I_{t+\delta t} - I_t = -\beta I_t \delta t + \lambda\sigma \cdot \mathrm{sign}(q_t)\sqrt{\frac{|q_t|}{\mathrm{ADV}}}.$$

is an EMA on $\lambda\sigma\mathrm{sign}(q_t)\sqrt{|q_t|/ADV}$.

This model is nonlinear in q but linear in λ. Therefore, one still fits the model via linear regression on the non-linear features.

References
Bouchaud et al. (2004), Busseti and Lillo (2012)

Reduced Form Model

A dynamic model
The reduced form model is a parametric form of the extended OW model of Fruth-Schoeneborn-Urusov

$$I_{t+\delta t} - I_t = -\beta I_t \delta t + \frac{\lambda \sigma q_t}{\sqrt{\mathrm{ADV} \cdot v_t}}$$

where

$$v_{t+\delta t} - v_t = -\beta v_t \delta t + |q|_t$$

is a higher frequency volume estimate.

Reference
Muhle-Karbe, Wang, and Webster (2022)

Alfonsi, Fruth, and Schied (AFS) model

The nonlinear AFS model

$$I_t = \text{sign}(J_t)\sqrt{|J_t|}$$

where

$$J_{t+\delta t} - J_t = -\beta J_t \delta t + \frac{\lambda \sigma q_t}{\text{ADV}}$$

is the linear impact in “volume space”.

Implementation Details

* Compute volume-impact via an EMA.
* Apply non-linear function as a vector operation.

Reference
Alfonsi, Fruth, and Schied (2010)

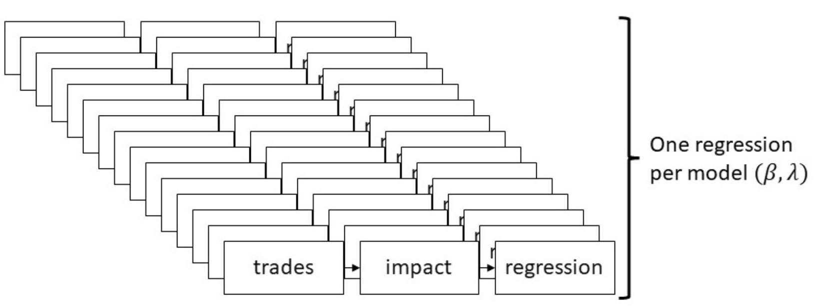
Comparison of Models

Fitting Methodology

General principle

Go through the data as few times as possible.

Naive implementation

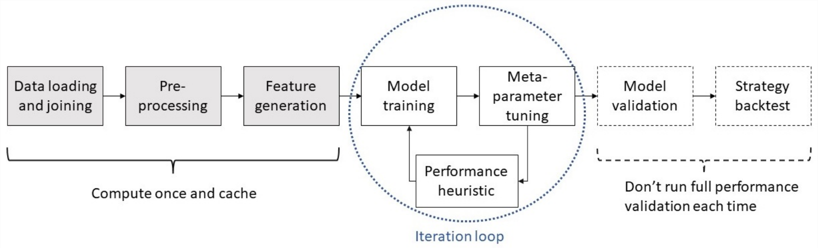


Refresher from Week 2

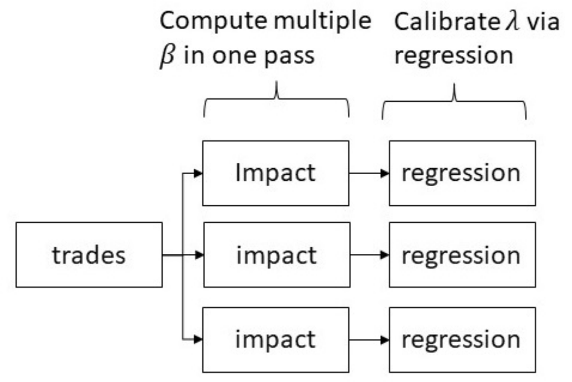
Accelerate the iteration loop.

(a) Pre-compute and cache steps: don't repeat early, expensive steps.

(b) Create intermediate tests and performance heuristics: don't repeat late, expensive validation steps.



Using pre-processing



Pre-processing implementation (1/2)

In Practice,
pre-compute impact models and fit λ, β without recomputing I.

* **Do not loop over the data for each (model, horizon) pair.**
* Compute I for multiple values of β when going through the data *once.* This is a single loop.
* Set $\lambda = 1$: the models are linear in λ.
* Compute horizon-specific increments for each prediction horizon h.
* Fit λ through linear regression: regress returns against ΔI, If available, use α as a control variable.

Pre-processing implementation (2/2)

Pre-processing step 1
Pre-compute

$$I_t(\beta = 1, 5, 10, 15, 30, 60, 120min; \lambda = 1)$$

by looping over the data $t = 1...N$ *once*.

Pre-processing step 2
Then, for each $h = 1, 15, 60$ compute

$$\Delta_h P_t = \frac{P_{t+h} - P_t}{P_t}; \quad \Delta_h I(\beta, 1) = I_{t+h}(\beta, 1) - I_t(\beta, 1)$$

Fitting step

$$\Delta_h P_t = \lambda \Delta_h I_t(\beta, 1) + \epsilon$$

estimates λ. Assume $\alpha = 0$ for the public tape and that ϵ is Gaussian and iid.

Evaluation Methodology: Sampling

Training and Validation Samples
No regularization: model is too small. Therefore, there is no testing sample, only a validation sample.

These are also called the *in-sample* (IS) and the *out-of-sample* (OOS).

Evaluation Methodology: Metrics (1/2)

In-Sample (IS) R^2

* How well the model fits the data.
* Can be computed per (stock, month): we show the average.
* Poor IS R^2 indicates an implementation bug or poor model features.

Out-Of-Sample (OOS) R^2

* How well the model works out of sample.
* Can be computed per (stock, month): we show the average.
* Poor OOS R^2 indicates overfitting compared to IS R^2.

Evaluation Methodology: Metrics (2/2)

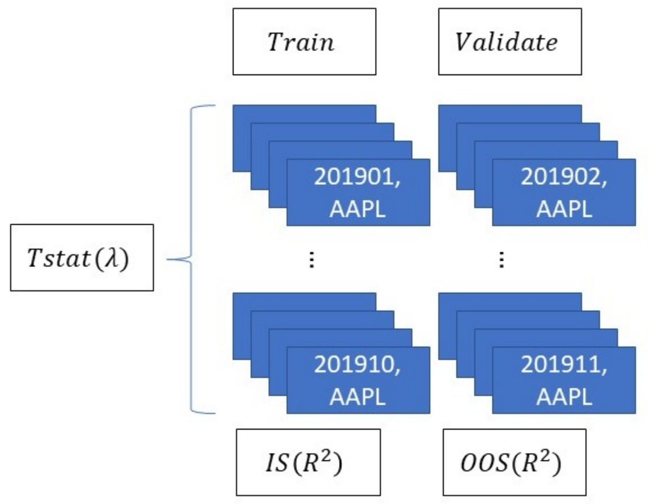
Tstat(λ)

$$\text{mean}(\lambda)/\text{sd}(\lambda)$$

where the mean and standard deviation are computed across all λ derived from different (stock, month) pairs.

* How stable the model λ is across (stock, month) pairs.
* Can be generalized over other dimensions that one wants λ stable over (e.g., time-of-day, horizons...)
* Poor tstat(λ) indicates the model is incorrectly normalized or does not fit well.

Evaluation Methodology: Summary



Timescales (1/4)

The next three slides
describe the performance of our models across various horizons h and impact halflives $\log(2)/\beta$.

Timescales (2/4)

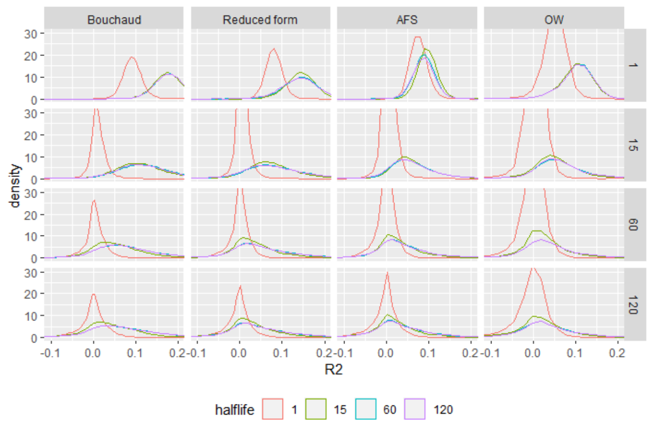


Figure 1: Distribution of R^2 across different choices of h (rows) and models (columns) for different betas (halflife in minutes).

Timescales (3/4)

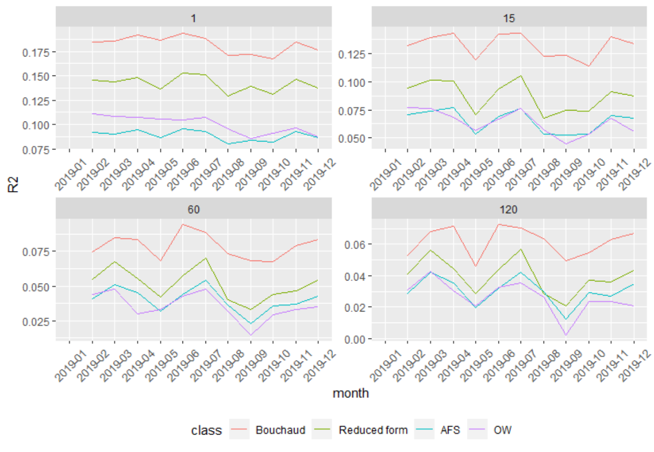


Figure 2: Average model out-of-sample R^2 across months and horizons h for $\log(2)/\beta = 60$ minutes.

Timescales (4/4)

Horizon (min)	Model	IS R^2	OOS R^2	t-stat(λ)
1	Bouchaud	19%	18%	6.2
1	Reduced-form	15%	14%	3.7
1	OW	11%	10%	2.7
1	AFS	9%	9%	4.1
15	Bouchaud	14%	13%	3.1
15	Reduced-form	11%	9%	2.1
15	OW	8%	6%	1.9
15	AFS	8%	7%	2.2
60	Bouchaud	10%	8%	2.0
60	Reduced-form	8%	5%	1.5
60	OW	7%	4%	1.3
60	AFS	7%	3%	1.2

Table 1: Model performance across horizons for $\log(2)/\beta = 60$ minutes.

Timescales Takeaways

(a) Models are not very sensitive to half-lifes within the 15min-2h range.

(b) Bouchaud's model fits the data best... but is hard to find optimal strategies with.

(c) The dynamic volume parameter in the reduced form model helps close the gap.

(d) The AFS model doesn't add much value for small orders. Better suited to larger meta orders, but no public data on these is available.

Time-of-day (1/3)

Motivation
Liquidity varies across the day: both trading volumes and price volatility are higher at the start and end of the day.

Could λ change over the day?

Diagnosis Method
Fit a separate model per time of day.

Downsides
This makes the model non-parametric: risk of overfitting due to additional model complexity.

Time-of-day (2/3)

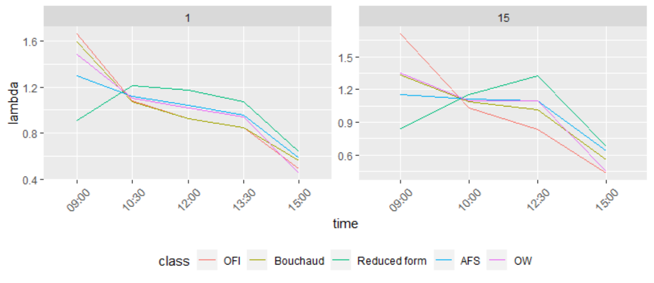


Figure 3: $\frac{\hat{\lambda}}{\text{mean}(\hat{\lambda})}$ across time of day under non-parametric fitting.

Time-of-day (3/3)

Horizon (min)	Model	IS R^2	OOS R^2	t-stat(λ)
1	Bouchaud	25%	24%	2.5
1	Reduced-form	18%	17%	2.5
1	OW	17%	14%	1.9
1	AFS	11%	10%	2.3
15	Bouchaud	20%	17%	1.5
15	Reduced-form	13%	10%	1.5
15	OW	13%	8%	1.1
15	AFS	10%	7%	1.2

Table 2: Model performance across horizons for $\log(2)/\beta = 60$ minutes under non-parametric fitting.

Time of Day Takeaways

(a) For static models, impact is under-estimated at start of the day and overestimated at the end of the day.

(b) The dynamic reduced form model corrects for this. Furthermore, it is parsimonious compared to fitting a new model every 90min.

(c) Ideally, one wants a consistent model across the day (flat line).

Universal Models (1/2)

Motivation
Our model was normalized to make λ comparable across stocks:

Could λ be the same across all stocks?

Diagnosis Method
Collect all data across stocks. Compare the stock-specialized models to a universal model.

Upsides
If the model is the same across stocks, one can pool together the data across stocks.

Therefore, one has 500x as much data.

This extra data enables a time-of-day fit,smaller sample window (e.g. daily instead of monthly), or more complex models.

Universal Models (2/2)

Model	univ.	spec.
Bouchaud	18%	19%
Red. form	13%	15%
OW	10%	12%

Table 3: In-sample R^2 for $h = 1$ minute.

Model	univ.	spec.
Bouchaud	13%	15%
Red. form	9%	11%
OW	7%	10%

Table 5: In-sample R^2 for $h = 15$ minute.

Model	univ.	spec.
Bouchaud	18%	17%
Red. form	14%	13%
OW	10%	9%

Table 4: Out-of-sample R^2 for $h = 1$ minute.

Model	univ.	spec.
Bouchaud	14%	13%
Red. form	10%	9%
OW	8%	7%

Table 6: Out-of-sample R^2 for $h = 15$ minute.

Universal Models Takeaways

(a) Universal models are less prone to overfitting: worth the effort to normalize the model across stocks.

(b) Practically appealing: one compares normalized model parameters across stocks and against the universal model.

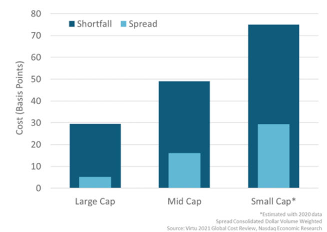
(c) Not present in table, but 500x data allows for non-parametric fitting, clustering of stocks into groups...

Summary (1/2)

Price Impact Model	$\mathbb{E}[\lvert I_T \rvert]$	$\mathbb{E}\left[\frac{\lvert I_T \rvert}{\sigma}\right]$	OOS R^2
Bouchaud	43bps	34%	18%
Reduced-form	25bps	20%	14%
OW	28bps	23%	10%

Table 7: End-of-day impact state, out-of-sample R^2 for one-minute predictions on the S&P 500 (large cap). $\mathbb{E}[\lvert I_T \rvert]$ are the trading costs of a representative daily order.

The 2022 Intern's Guide to Trading, Nasdaq (2022)



Summary (2/2)

Four "sanity checks"

(a) Depending on the prediction horizon, out-of-sample R^2 for price impact models of the aggregate order flow ranges upward of 10%.

(b) Price impact's magnitude on the aggregate flow is around 30bps on liquid stocks. Generally, price impact is five to ten times larger than the bid-ask spread across all stocks.

(c) Price impact exhibits a pronounced time of day effect: trades cause the most price impact at the start of the day.

(d) Price impact is concave for sizable orders (not shown here).

An Alpha Model

Order Flow Imbalance (OFI)

Additional events besides trades
The majority of limit order book (lob) events are *not* trades.

For example, AAPL on 2019.02.01:

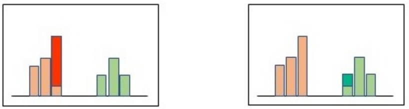
```
q)update eventPercentage: eventCount % sum eventCount from select eventCount: count i by event from tbl lj map
event        | eventCount eventPercentage
-------------| --------------------------
hiddenTrade  | 3352       0.007179929
lobCancel    | 189093     0.4050341
lobSubmission| 229197     0.4909362
trade        | 45215      0.09684978
```

Additional features for trading signals
Main idea: aggregate *all* events into one summary statistic.

Event Contributions to OFI (1/3)

Trades
The contribution of trades to order flow imbalance is the signed fill quantity $q = a - b$, where a, b fill at the ask and bid.

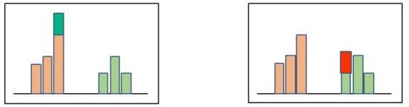
Example



Event Contributions to OFI (2/3)

Limit order submissions
The contribution of limit order submissions to order flow imbalance is $b - a$, where a, b are submissions on the ask and bid. This is a simplification: trading impacts more than posting.

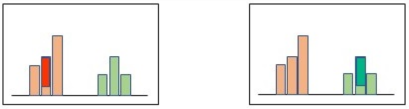
Example



Event Contributions to OFI (3/3)

Limit order cancellations
The contribution of limit order cancellations to order flow imbalance is $a - b$, where a, b are cancellations on the ask and bid.

Example



Cont, Kukanov, and Stoikov (2013)

Best practices:

(a) In your code, isolate features into single compute nodes.

(b) Explicitly map out the model's computational graph, especially those including a free parameter.

(c) This aids readability, re-usability, and backward gradient propagation.

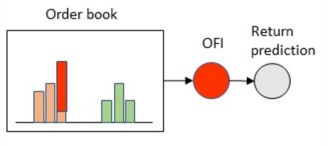


Figure 4: Single node model of Cont, Kukanov, and Stoikov (2013).

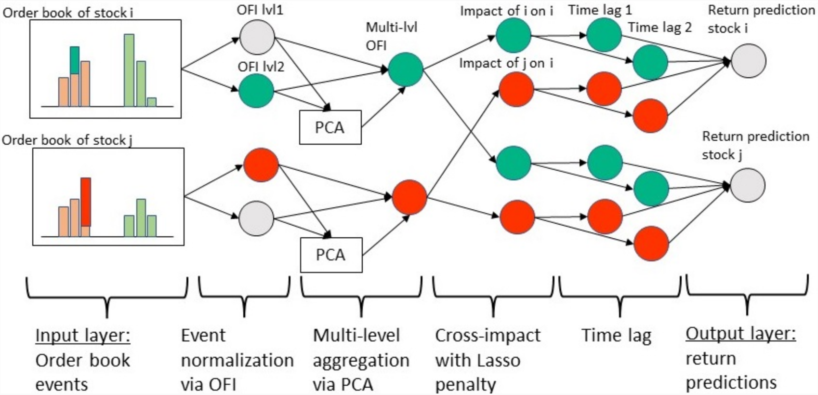


Figure 5: Model architecture in Cont, Cucuringu, and Zhang (2021) containing an input layer, four computational layers, and an output layer.

Forecasting vs Nowcasting

Cont, Cucuringu, and Zhang (2021)

> *"**Forecasting future returns.** Finally, we investigate the performance of forward-looking price-impact and cross-impact models, i.e. using OFIs to forecast future returns, which has received a lot less attention in the literature, as it is an inherently much more challenging problem than explaining contemporaneous returns." (p. 3)*

Previous impact regressions nowcasted returns.
Important to establish the true cost of trading but **distinct** from alpha research. Out of sample R^2 are drastically lower for alpha research (single digit instead of double digit).

General Machine Learning on the Order Book

Kolm, Turiel, and Westray (2022, link)
"Deep Order Flow Imbalance: Extracting Alpha at Multiple Horizons from the Limit Order Book" is becoming a trading classic:

(a) outlines standard data cleaning and feature construction (OFI variants)
(b) applies and compares *off-the-shelf* model architectures *on the same standard features*
(c) provides a robust horse-race to *benchmark* the models

Comparison to Cont, Cucuringu, and Zhang (2021)
Because the models are off-the-shelf, the architecture provides less trading intuition than the hand-crafted architecture of Cont, Cucuringu, and Zhang (2021). However, it serves as a clear benchmark to beat and still incorporates trading domain knowledge in the feature construction (OFI features).

Weekly Summary

Price impact is universal to trading
Researchers have fit price impact models across public and proprietary data, across asset classes, and across time scales with broadly consistent results.

Common sanity checks

(a) R^2 in the double digits for short time scales.
(b) Magnitude of price impact is 30bps on liquid stocks for a typical day order.
(c) Price impact exhibits a time of day effect.
(d) Price impact is concave for sizable orders.

Questions?

Next week

(a) Extension of the OW model to dynamic liquidity signals β_t, λ_t.

(b) Price manipulation concerns, and how to rule price manipulation out.

(c) The local square-root law.