GETTING MORE FOR LESS - BETTER A/B TESTING VIA CAUSAL REGULARIZATION

KEVIN WEBSTER

Imperial College London

NICHOLAS WESTRAY

Head of Execution Research, AllianceBernstein Multi-Asset Solutions & Financial Machine Learning Researcher, Courant Institute of Mathematical Sciences, NYU

Abstract. Causal regularization solves several practical problems in live trading applications: estimating price impact when alpha is unknown and estimating alpha when price impact is unknown. In addition, causal regularization increases the value of small A/B tests: one draws more robust conclusions from smaller live trading experiments than traditional econometric methods. Requiring less A/B test data, trading teams can run more live trading experiments and improve the performance of more trading algorithms. Using a realistic order simulator, we quantify these benefits for a canonical A/B trading experiment.

1. Introduction

The interplay of information and trading is critical to trade execution. Practitioners refer to price moves caused by their trading as *price impact* and price moves independent of their trading as *alpha*. As noted by [18]

> "Large scale trading will often occur in the presence of market drift (alpha) and the realized execution cost is a combination of alpha and the price impact" (p. 313, [18]).

An essential corollary is that trading causes price moves that otherwise would not have happened. Successful investment strategies across all asset classes trade to minimize the price impact and maximize the alpha during trading.

Date: July 12, 2022.

Key words and phrases. Algorithmic trading; A/B Testing; Best execution; Optimal execution; Trading; Transaction cost analysis.

The vast literature on modelling market impact proposes functional forms and describes statistical challenges for modeling price impact, including (amongst many others)

- Lillo, Farmer, and Mantegna (2003)[11], Bouchaud et al. (2004)[4], and Cont, Kukanov, and Stoikov (2014)[7] who use public trading data to fit price impact models in both US and non-US equities.
- Almgren et al. (2005)[1] and Bershova and Rhakhlin (2013)[2] who leverage proprietary orders from Citigroup US equity trading desks and AllianceBernstein to apply price impact models to Transaction Cost Analysis (TCA).
- Donier and Bonart (2015)[8] and Tomas, Matromatteo, and Benzaquen (2021)[16, 17] who estimate standard price impact models on bitcoin, fixed income, and derivatives products.

The literature highlights a crucial challenge when estimating impact: alpha signals cause trades:

> "The larger the volume Q of a metaorder, the more likely it is to originate from a stronger prediction signal." ([5] p. 238)

Bouchaud et al. (2018)[5] refer to this bias as "Prediction bias" (p. 238). In Section 2 we provide a detailed instance of this bias and how it leads to sub-optimal trading and lower P&L.

The industry standard for addressing this bias is through controlled live trading experiments, such as A/B tests that randomize decisions. For example, Bouchaud (2021)[3] leverages a year-long live trading experiment to identify price impact without bias. A/B tests address trading biases but present three downsides: First, one discards the bulk of their trading data. Second, there are far fewer trades without alpha than with alpha, making it challenging to estimate high dimensional models using machine learning. Finally, the submission of alpha-less trades leads to additional trading costs.

The present article uses causal regularization, introduced by Janzing [9], to address these shortcomings. This method improves upon traditional A/B testing by analyzing *both* the unbiased A/B testing and biased trading data. Figure 1 demonstrates this: in the left-hand panel, one uses the considerable trading data to fit an impact model leading to a biased estimation with small uncertainty. In the middle panel, one only uses unbiased data, which naturally gives an unbiased estimate but with large uncertainty. Finally, in the right-hand panel, causal inference blends both data, providing an unbiased estimate with small uncertainty, giving the *best of both.*

Five further sections structure the article. Section 2 provides our motivating example, Section 3 introduces causal regularization, Section 4 describes the simulations and experiments, Section 5 contains the results, and Section 6 concludes. Finally, Appendix A outlines the application of causal regularization to alpha research.

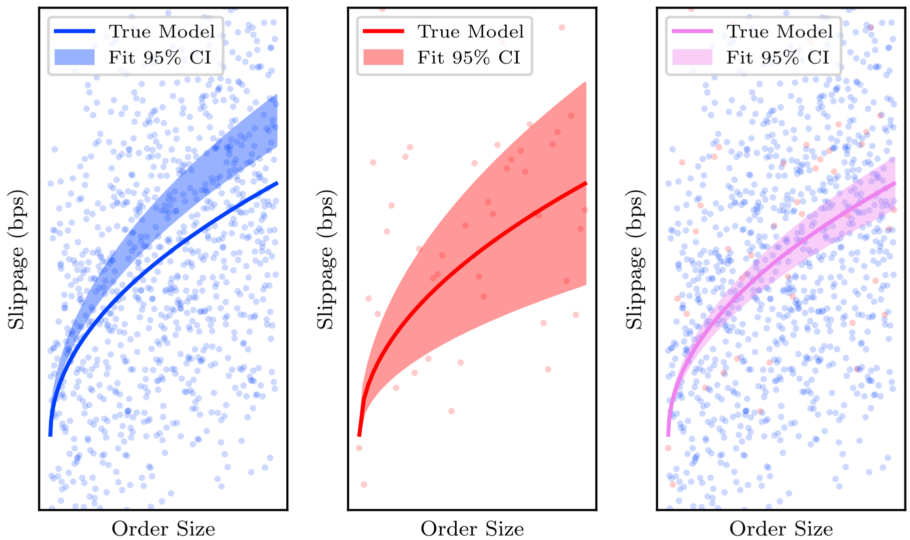

FIGURE 1. Illustration of the bias and variance trade-off between data with alpha (blue) and data without alpha (red). The last panel combines both data through *causal regularization*.

2. A MOTIVATING EXAMPLE

As a stylized example, consider a trading strategy on the Russell 3000 universe. Each day the strategy uses an alpha model to predict future returns across a subset of $N < 3000$ stocks, leading to a distribution $\alpha_i \sim N(\mu, \sigma_\alpha^2)$ of alpha signals. The trading strategy applies a portfolio optimization model. The optimization considers alpha signals α_i to submit orders of size x_i. The equation

$$x_i = \gamma\alpha_i + \nu_i \tag{1}$$

models the order sizes, with $\nu_i \sim N(0, \sigma_\nu^2)$ and $\gamma > 0$.

The strategy trades orders of size x_i over the day, for example, with a VWAP algorithm, and the realized returns r_i take the form

$$r_i = \alpha_i + \sigma \operatorname{sign}(x_i)\sqrt{|x_i|} + \varepsilon_i \tag{2}$$

where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, and the square root term represents the market impact of trading an order of size x_i.

The above structural model describes how alpha causes trades, and how both alpha and trades cause price moves, in line with Bouchaud's definition of prediction bias. In the language of causal inference, this is a causal model, which we illustrate in Figure 2.[1]

Two researchers study the trading strategy from opposing angles.

[1] See Section 2.2 "the causal discovery framework" (p. 43) of Pearl (2009)[12], for mathematical definitions of causal structures and models.

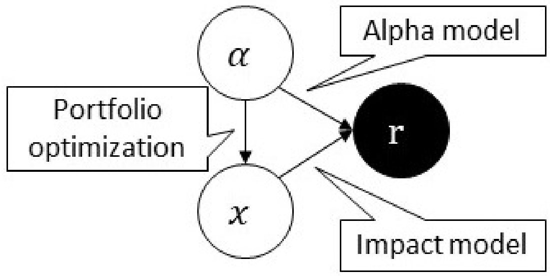

FIGURE 2. Causal structure for trading involving an alpha α causing an order of size x. Both α and x affect the returns r.

(a) An execution researcher, charged with estimating the price impact of trading an order of size x, runs the following regression without intercept

$$r_i * \mathrm{sign}(x_i) \sim \sqrt{|x_i|}.$$

The regression follows the industry practice of fitting realized signed returns against the square root of the absolute order size. However, the alpha biases the regression coefficient upwards due to its correlation with the order's size. This bias leads the execution researcher to slow down trading and capture less P&L for the strategy.

(b) An alpha researcher, charged with estimating the actual value of their alpha signals, runs the regression without intercept

$$r_i \sim \alpha_i.$$

Price impact biases the regression coefficient on α upward, leading the alpha researcher to oversize their order and capture less P&L for the strategy.

The problem faced by the execution researcher is due to a *causal bias*: the hidden alpha α confounds the estimation of price impact, as it both causes trades and price moves. This bias is only present because the execution researcher knows the alpha exists but does not observe it.[2] Mitigating this causal bias in estimating price impact is the main application of our causal regularization method. Appendix A describes the problem the alpha researcher faces.

[2]If the researcher observed the alpha, they could co-fit alpha and impact, leading to a bias-free estimate of both. However, the execution researcher's client is unlikely to share the trade's alpha.

3. A/B testing and Causal regularization

Live trading experiments tackle trading biases, such as the prediction bias from Section 2. For example, Bouchaud (2021)[3] leverages randomized trades to estimate price impact without bias.

> "We have actually shown that the short term impact of CFM's trades is indistinguishable [...] from purely random trades that were studied at CFM during a specifically designed experimental campaign in 2010-2011." (p. 4)

The scale and length of CFM's live trading experiment are impressive: submitting randomized trades is costly and maintaining a controlled experiment for a whole year requires patience. Sotiropoulos and Battle (2017)[15] observe that for electronic brokers, such as Deutsche Bank, A/B experiments are small to allow for "routine changes"(p. 5) and to control the live trading experiment's cost.

In a recent article, Janzing (2019)[9] observes a correspondence between formulas for a finite-sample bias and a causal bias in a linear regression framework. Janzing shows that regularization addresses more than finite-sample biases. Indeed, regularization does not require specific assumptions on the bias, simply that the testing data does not have the same bias as the training data. Janzing's insight [9] is that regularization reduces the regression coefficient's norm, mitigating causal bias in our model. Moreover, the regularization parameter is commonly one-dimensional: one requires significantly fewer data to calibrate it. Regularization leads to the *best of both* method defined in Algorithm 1.

Algorithm 1 Causal Regularization

Inputs:

a. Observational Data: (y_O, X_O)
b. Interventional Data: (y_I, X_I)
c. Model w/ parameter $\beta : \mathcal{M}(\beta)$
d. Fitting method with regularization parameter λ

Steps:

1. Use the fitting method to train β_λ for all values of λ on the observational data (y_O, X_O).
2. Use the interventional data (y_I, X_I) as a testing set to tune the optimal $\hat{\lambda}$.
3. Use $\mathcal{M}(\cdot, \beta_{\hat{\lambda}})$ for predictions.

With our causal model and regularization algorithm in hand, we restate the conclusions of Figure 1 using the terminology of causal machine learning. Observational data refers to alpha-driven trades, and interventional data refers to alpha-less trades, so that:

(a) The large observational data (blue data points) gives a *biased estimator with small variance* (first panel).
(b) The smaller interventional data (red data points) gives an *unbiased estimator with large variance* (second panel).
(c) Using causal regularization to train on the observational data and tune the regularization parameter on the interventional data gives the best outcome: an *unbiased estimator with small variance* (third panel).

The assumption here is that, while the interventional data may not be large enough to fit a high-dimensional model, it is large enough to *de-bias* the model trained on observational data.[3] The researcher establishes the model's shape via the training data, and the testing data removes the last degree of freedom, λ. This approach blends the practical advantages of both observational and interventional data. Notably, in trading applications, observational data is plentiful but biased, and interventional data is scarce but bias-free.

To illustrate the success of our method, we perform a simulation study using Kolm and Westray's order simulator [10]. Our large-scale simulation study shows how even tiny interventions effectively control the bias and variance of causal parameters, hence our use of *more for less.* Finally, one can run more experiments in parallel as an additional practical benefit of small interventions.

4. Experiment Description

Sotiropoulos and Battle (2017)[15] show that live trading experiments are small. Consider again a situation where a firm has a trading universe of the Russell 3000 index and suppose that they trade 1000 symbols daily, leading to 250K orders annually. A starting point that may seem statistically reasonable uses 10% of the order flow for A/B testing, yielding 25K orders in interventional data over a year. However, while the data's size is attractive, this corresponds to over a month's worth of unprofitable trades and requires traders to maintain the same live trading experiment for a year. Such a design is not realistic, so controlled live trading experiments are significantly more modest in size and duration. For example, one may more realistically allocate 3% of the orders over two months to the experiment, leading to interventional data of 1.25K orders or 0.5% of the observational data's size.

To rigorously assess our method, we need sizeable interventions to use as proper out-of-sample data and to try various sizes for the live trading experiment. Using such sizeable interventional data, we can measure our

[3] Rothenhäusler et al. (2020)[14] provide a method for dealing with the case where bias-free data is unavailable for testing: regularization is achievable if the researcher collects sufficient heterogeneous data through past experiments. Therefore, with causal regularization, a large history of heterogenous trading experiments *may* avoid the cost of implementing a dedicated experiment for a new problem.

method's reduction in bias and variance across proposed experiment sizes. Unfortunately, only a simulation can realistically achieve an intervention of that size. Therefore, we leverage Kolm and Westray's simulator framework [10], which we briefly describe in Section 4.1.

4.1. **Order Simulator.** When attempting to simulate orders, they must resemble real-life trades: we reproduce specific distributional properties and stylized facts.

(1) Order size as an ADV percentage positively correlates with participation rate as a volume percentage.
(2) Order size negatively correlates with market capitalization.
(3) The distribution of realized trading rates follows a power law.

Kolm and Westray's idea [10] is that, due to the prevalence of the Markowitz approach to portfolio management and Mean-Variance Optimization, one simulates realistic orders by first taking a set of alphas α and solving the following portfolio optimization problem

$$\begin{aligned} &\max_x \alpha^\top (w+x) - \lambda (w+x)^\top \Sigma (w+x) \\ &w + x \in \mathcal{C}_1,\ x \in \mathcal{C}_2 \end{aligned}$$

where w are the input portfolio weights, α the input alphas, Σ the stock covariance matrix, and λ the risk aversion parameter. The sets $\mathcal{C}_1 \& \mathcal{C}_2$ represent the constraints.[4] The resulting x are the required trades as a percentage of the portfolio notional. Looping over days and inserting new alphas generates a series of trades/orders. In addition, we can create additional orders by re-running the period with different alphas. As discussed above, this allows us to generate arbitrary interventions and assess our results as interventional data grows. Indeed, replicating this bootstrap would be prohibitively expensive using live trading data only.

To construct the alphas, we use the idea of *bootstrap alphas*. For a given day, for each stock indexed by i, we generate the vector of alphas

$$\alpha_i = \rho r_i + \sqrt{1-\rho^2} Z_i, \quad Z_i \sim N(0, \sigma_i^2)$$

where r_i and σ_i^2 are the stock's return and variance. We take the variance from the Northfield risk model. Choosing $\rho = 0$ provides alpha-less trades, and, more generally, ρ controls the strategy's profitability. We choose ρ such that the strategy's realized IC is around 5%. We choose the constituents of IWV, the iShares Russell 3000 ETF, as our trading universe to have the widest cross-section of stocks for our results.[5]

Because they are a mixture of actual returns and noise, the synthetic alphas have lower autocorrelation than in real-life. In addition, the objective does not consider transaction costs: this simulation over-trades compared to an actual portfolio. However, the vital observation is that, for the study

[4] We use a 5x leverage constraint and force the portfolio to be sector & delta neutral.
[5] The reader finds full details, including constituents, at iShares Russell 3000 ETF.

of price impact, we are not interested in precise portfolio characteristics, only resulting trades. [10] show that the simulated trades follow all essential stylized facts, and this property is all our study requires to assess price impact estimators.

4.2. **Methodology.** In this section, we describe three fitting methods, illustrated on the example from Section 2, using the simulated data from Section 4.

Using the order simulator, we generate:

(O) An extensive set of 2.5 million orders *with* alpha.
(I) Smaller sets of interventional alpha-less orders. Each interventional data emulates a live A/B-test limited in size and duration: they range from $n = 250$, representing 0.01% of the observational data size, to $n = 12500$, representing 0.5% of the observational data size.[6]
(V) An extensive set of 2.5 million alpha-less orders as validation data.

Data (O) emulates a deep history of observational data that a trading team may have built up over years of trading. Unfortunately, (O) contains an unknown bias due to the orders' alpha. Finally, data (V) does not have a cost-effective counterpart in real-life and simply serves as a simulation of the true out-of-sample for the price impact model.

For each order, we simulate price impact using three models from the literature: the original model proposed by Almgren et al. (2005)[1], the power-law model from Zarinelli et al. (2015)[20], and the square-root model from Bucci et al. (2018)[6]. Given a sample for (O), (I), and (V), a price impact model I, and a set of historical returns $\tilde{r}_i$, we construct synthetic returns r_i that contain the impact of the simulated orders via the formula

$$r_i = I(x_i) + \tilde{r}_i. \tag{3}$$

We define the linear regression model

$$r_i = \beta I(x_i) + \varepsilon_i. \tag{4}$$

When the fitted model perfectly recovers the impact parameter $\beta = 1$, the residuals ϵ_i match the historical returns $\tilde{r}_i$ from which we constructed our synthetic returns.[7]

We fit the model in three ways.

(1) We fit *using observational data:* a least-squares regression runs on (O).

[6]For example, while the trading team can leverage years of past (observational) trading data, A/B tests typically are only launched *after* proposing an experiment: therefore, their histories are short. The data is also limited in size by cost considerations: a well-designed controlled experiment randomizes key trading variables, leading to additional trading costs or opportunity losses.

[7]In a real-life setting, the residuals ϵ_i correspond to the impact-adjusted returns the alpha model predicts.

(2) We fit *using interventional data:* a least-square regression runs on (I).

(3) We fit *using causal regularization:* a ridge regression runs on the *observational* data (O), and the ridge meta-parameter maximizes the R^2 on the *interventional* data (I).

Remark 4.1 (Precision vs. Accuracy)**.** *The regression with* observational data only *is the most precise, given the orders of magnitude more data it has available. For example, the confidence interval for the observational data (O) is* $\sqrt{200} \approx 14$ *times tighter than for the interventional data (I) of size* $n = 12500$*, assuming the Central Limit Theorem applies.*

Due to the interventional data being bias-free, the regression on (I) is significantly less precise but more accurate. We show that causal regularization is both *precise and accurate, even for tiny sizes of the interventional data (I).*

Repeating the procedure with independent samples generates many parameter estimates and bootstraps the parameter distribution for each method. Finally, to assess the three fitting techniques, we introduce three performance metrics.

(1) *The bias* of the parameter estimate. Given that we know the actual value of the price impact parameter, we can quantify the bias of the parameter estimate for each method. The bias measures the model's *inaccuracy*.

(2) *The t-stat* of the parameter estimate. We compute the t-stat for each method from the bootstrapped distribution. The t-stat measures the model's *precision*.

(3) *The validation* R^2 of the model. Evaluating the models on data (I) unfairly biases the method based on interventional data only and produces a noisy evaluation. Instead, we leverage the validation data (V), which has the same size as the observational data and is bias-free, to compute each model's validation R^2.

One can only realistically estimate the above performance metrics in simulation: the bootstrap methodology and the validation data require an impractical number of interventions for a team to replicate in live trading. Therefore, the simulation environment from Section 4 plays a crucial role in assessing experimental designs and statistical methods.

5. Results

We first describe the simulated orders. Table 1 summarizes the distribution of the order size (ADV%), stock market capitalization, and order speed (Participation of Volume, PoV%). Figure 3 highlights the correlation between the three essential variables in the order set. Figure 4 provides the sampling distributions of the sizes and PoVs used for the calculations. First,the Mean-Variance Optimization's constraints bound the order size and speed. Second, order size and speed correlate: the larger the order, the faster the execution to attain the desired position promptly. Finally, the traded stocks' market capitalization presents a heavy tail and a negative correlation with order size: the mean-variance optimization problem submits smaller orders on more liquid names.

Quantile	ADV(%)	Mkt Cap(bil)	PoV (%)
0.0-0.2 (ADV%)	0.1	25.7	4.1
0.2-0.4 (ADV%)	0.4	14.8	4.2
0.4-0.6 (ADV%)	1.0	9.3	4.6
0.6-0.8 (ADV%)	1.8	6.2	5.2
0.8-1.0 (ADV%)	3.2	4.2	6.7
0.0-0.2 (Mkt Cap)	1.0	0.3	4.7
0.2-0.4 (Mkt Cap)	1.4	0.9	5.0
0.4-0.6 (Mkt Cap)	1.6	2.1	5.2
0.6-0.8 (Mkt Cap)	1.5	5.3	5.2
0.8-1.0 (Mkt Cap)	1.0	51.7	4.7
0.0-0.2 (PoV%)	0.9	16.6	2.2
0.2-0.4 (PoV%)	1.0	13.0	2.8
0.4-0.6 (PoV%)	1.2	11.1	3.8
0.6-0.8 (PoV%)	1.5	10.1	5.7
0.8-1.0 (PoV%)	1.9	9.5	10.2

TABLE 1. Summary statistics of the simulated orders. Each column represents the average of a metric over a given quantile. The first column of each row specifies the quantile.

The reader finds the results of the three estimation methods in Figure 5 and Table 2. The figure shows the distribution of empirical β as we evaluate different interventional data (I) and estimation techniques. In addition, the table provides the bias and t-stat of the fitted β and the R^2 on the validation data (V).

As expected, the observational data provides a precise but biased estimate of β. For small experiments, using interventional data provides a noisy but unbiased estimator. The method becomes precise once the intervention size reaches 12500 randomized trades. Figure 5 illustrates how causal regularization achieves a tighter distribution of empirical β for a given experiment size. The results are robust across all three impact models.

Causal regularization achieves unbiased, precise measurements using an order of magnitude less randomized trades. For example, the causal regularization estimator with $n = 250$ randomized trades outperforms the naive estimator with $n = 1250$ randomized trades.

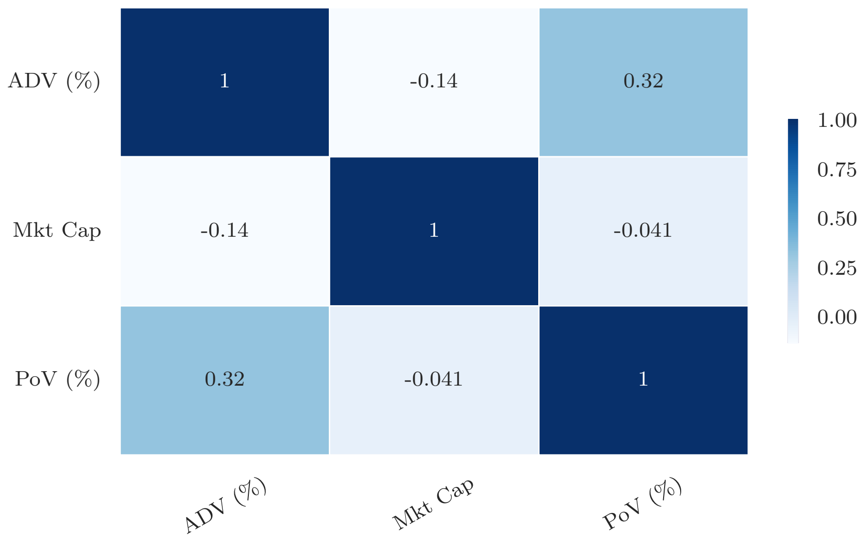

FIGURE 3. Correlations between variables in our simulated orders.

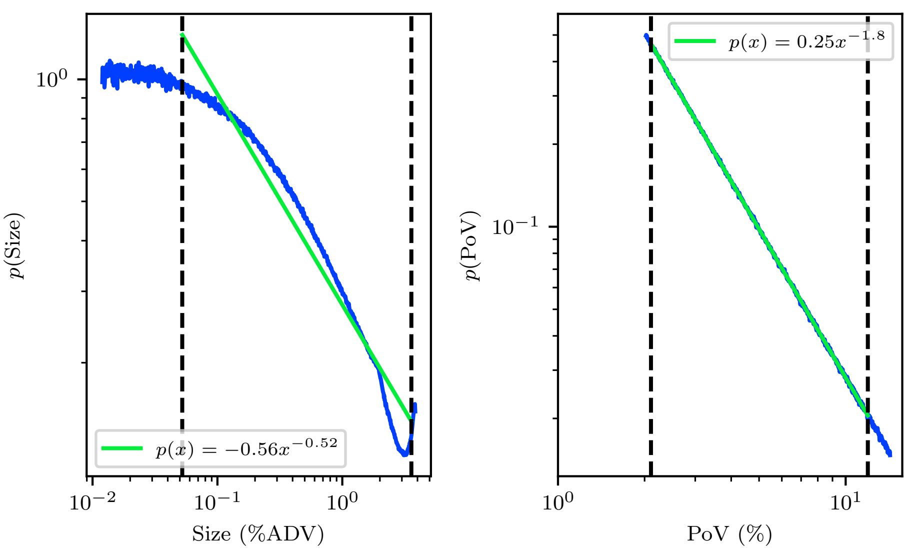

FIGURE 4. Distribution functions of the order sizes and PoVs used for the simulation.

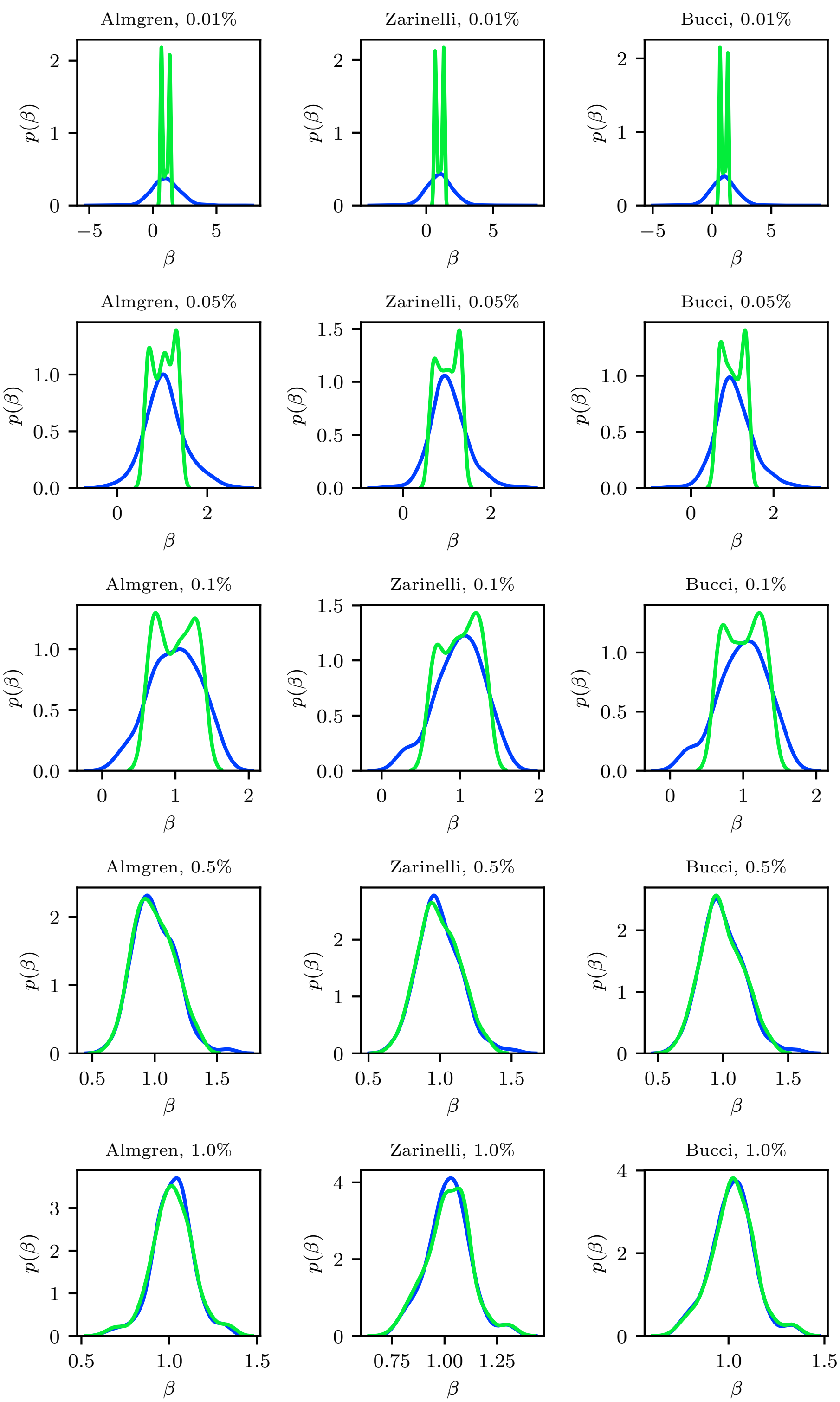

FIGURE 5. KDE plots of the simulated regression betas: the columns represent different price impact models. The rows represent assorted intervention sizes. The actual value is $\beta = 1$. Betas from fits on the interventional data alone are in blue and from the causal regularization approach in green.

Intervention size	Fitting method	Bias	t-stat	$R^2(V)$
0%	Observational data only	0.319	30.2	70bps
0.01%	Interventional data only	-0.010	0.92	-10bps
0.01%	Causal regularization	-0.010	3.32	70bps
0.05%	Interventional data only	0.05	2.40	63bps
0.05%	Causal regularization	0.01	4.12	72bps
0.5%	Interventional data only	-0.002	6.38	75bps
0.5%	Causal regularization	-0.005	6.66	75bps

Table 2. Statistical results across fitting methods and with intervention sizes ranging across 0.01% ($n = 250$ randomized trades), 0.05% ($n = 1250$ randomized trades), and 0.5% ($n = 12500$ randomized trades) of the observational data.

6. Conclusion

This article revisits A/B testing in a trading context. When designing live trading experiments, the critical trade-off is best described using the language of causal inference: observational data is plentiful but biased, while interventional data is bias-free but scarce. The authors find causal machine learning particularly well-suited for trading applications and hope that quantitative Finance follows the Technology industry in applying these methods to real-life problems. The Microsoft Research Summit of 2021[13] had over a dozen talks within its causal machine learning track, one of its seven science tracks.

> "This track focuses on emerging causal machine learning technologies and the opportunities for practical impact at the intersection of academia and industry, with contributions from researchers at Microsoft and the broader academic and industrial research communities."

Causal inference is not only well-suited to describe the trade-off inherent to A/B testing: causal machine learning methods also significantly outperform traditional econometric techniques in the data regime most common in trading. We illustrate this via a rigorous and extensive simulation experiment for trading. We show that a trading experiment with only 250 randomized trades using our causal regularization method outperforms a standard A-B test with 1250 randomized trades.

In a well-controlled simulation environment, the paper presents a concrete, robust analysis of causal regularization's benefits to the bias-free estimation of price impact. The range of applications is significantly broader, see, for instance, Appendix A on alpha research, and likely to increase with continued demand for A/B testing in a trading and best execution context.

Acknowledgements

We are grateful to Northfield for providing their risk model and Dan diBartolomeo and Steve Gaudette, in particular, for answering our data questions. In addition, the authors thank Petter Kolm and Johannes Muhle-Karbe for comments which helped improve this article.

Appendix A. A second use-case

In this appendix, we revisit the motivating example in section 2 from an alpha researcher's point of view . Recall that the returns follow the model

$$r_i = \alpha_i + \sigma \operatorname{sign}(x_i)\sqrt{|x_i|} + \varepsilon_i \tag{5}$$

where r are the observed returns, α the alpha researcher's signal, and $\sigma \operatorname{sign}(x)\sqrt{|x|}$ the price impact of the strategy's historical orders.

Imagine a scenario where an alpha researcher does not collect their own trading data but relies on a third party, for example, a broker, to capture their trading data and estimate its price impact. The alpha researcher knows that the historical regression

$$r_i \sim \alpha_i.$$

is biased due to the presence of price impact. Waelbroeck et al. (2012)[19] propose a method to remove this bias *under a given choice of price impact model*:

> "the system may estimate corrected market prices that would have been observed had price impact not existed" ([19] p. 11)

In Waelbroeck's solution, the alpha researcher takes the price impact model from their broker at face value and, for our square root impact model, runs the regression

$$r_i - \beta\sigma\,\mathrm{sign}(x_i)\sqrt{|x_i|} \sim \alpha_i.$$

An alpha researcher can implement Algorithm 2 by Waelbroeck et al. (2012)[19].

Algorithm 2 Waelbroeck's price impact adjustment algorithm

Inputs:

a. Alphas α_i
b. Observed returns r_i
c. Price impact quoted by third-party I_i

Steps:

1. Compute corrected market returns $\tilde{r}_i = r_i - I_i$
2. Regress $\tilde{r}_i \sim \alpha_i$.

But what if the alpha researcher wants to fit their alpha free of a particular price impact model? An alpha researcher using causal regularization Algorithm 3 can fit their alphas without depending on a third party's price impact model.

Algorithm 3 Causal regularization for alpha research

Inputs:

a. Alphas α_i
b. Observed returns r_i
c. A randomized set $\mathcal{I}$ of *unsubmitted* trades i

Steps:

1. Define $\mathcal{I}$ as the interventional data and its complement $\mathcal{O}$ as the observational data.
2. Apply the causal regularization algorithm 1 to the linear regression $r_i \sim \alpha_i$.

To use Algorithm 3, an alpha researcher randomly sets aside a small set of names and disables their trading strategy on those names. The alpha researcher then monitors the returns of these *unsubmitted trades* to tune their regularization penalty. The causal regularization algorithm, Algorithm 3, calibrates the correct ridge parameter for the researcher's alpha model based on this bias-free data. In conclusion, the algorithm considers the confounding effect of price impact *in a model-free way*, reducing the researcher's reliance on broker data and models.

The researcher cannot practically implement this approach without causal regularization: the opportunity cost of *not* submitting profitable trades is prohibitively high. Unfortunately, standard econometric techniques demand enormous interventional data. But *causal regularization gets more for less* and adjusts alpha for price impact in a model-free way with minimal opportunity costs.

References

[1] R. Almgren et al. Direct estimation of equity market impact. *Risk*, 18, 2005.
[2] N. Bershova and D. Rakhlin. The non-linear market impact of large trades: Evidence from buy-side order flow. *Quantitative Finance*, 13(11):pp. 1759–1778, 2013.
[3] J.P. Bouchaud. The inelastic market hypothesis: A microstructural interpretation. *Preprint*, 2021. https://arxiv.org/abs/2108.00242.
[4] J.P. Bouchaud et al. Fluctuations and response in financial markets: The subtle nature of random price changes. *Quantitative Finance*, 4(2):pp. 176–190, 2004.
[5] J.P. Bouchaud et al. *Trades, Quotes and Prices*. Cambridge University Press, 2018.
[6] F. Bucci et al. Slow decay of impact in equity markets: Insights from the ancerno database. *Market Microstructure and Liquidity*, 2018.
[7] R Cont, Kukanov A., and S. Stoikov. The price impact of order book events. *Journal of Financial Econometrics*, 12(1):pp. 47–88, 2013.
[8] J. Donier and J. Bonart. A million metaorder analysis of market impact on the bitcoin. *Market Microstructure and Liquidity*, 1(2), 2015.
[9] D. Janzing. Causal regularization. *Advances in Neural Information Processing Systems*, 2019.
[10] P. Kolm and N. Westray. Mean-variance optimization for simulation of order flow. *The Journal of Portfolio Management*, 2022.
[11] F. Lillo, J.D. Farmer, and R.N. Mantegna. Econophysics: Master curve for price-impact function. *Nature*, 421:pp. 129–130, 2003.
[12] J. Pearl. *Causality*. Cambridge University Press, 2009.
[13] Microsoft Research. *Microsoft Research Summit 2021*. https://www.microsoft.com/en-us/research/event/microsoft-research-summit-2021/, 2021. Accessed 11 March 2022.
[14] D. Rothenhäusler et al. Anchor regression: heterogeneous data meet causality. *Preprint*, 2020. https://arxiv.org/abs/1801.06229.
[15] M. Sotiropoulos and A. Battle. Extended transaction cost analysis (tca). *Deutsche Bank*, 2017.
[16] M. Tomas, I. Mastromatteo, and M. Benzaquen. How to build a cross-impact model from first principles: Theoretical requirements and empirical results. *Quantitative Finance*, 2021.
[17] M. Tomas, I. Mastromatteo, and M. Benzaquen. Cross impact in derivative markets. *Preprint*, 2022. https://arxiv.org/abs/2102.02834.

[18] R. Velu, M. Hardy, and D. Nehren. *Algorithmic Trading and Quantitative Strategies*. CRC Press, 2020.

[19] H. Waelbroeck et al. Methods and systems related to securities trading, US Patent 8,301,548 2012.

[20] E. Zarinelli et al. Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate. *Market Microstructure and Liquidity*, 2015.