Algorithmic Aspects of Strategic Trading

Michael Kearns
University of Pennsylvania

Mirah Shi
University of Pennsylvania

Abstract

Algorithmic trading in modern financial markets is widely acknowledged to exhibit strategic, game-theoretic behaviors whose complexity can be difficult to model. A recent series of papers [8, 9, 7, 10] has made progress in the setting of trading for *position building*. Here parties wish to buy or sell a fixed number of shares in a fixed time period in the presence of both temporary and permanent market impact, resulting in exponentially large strategy spaces. While these papers primarily consider the existence and structural properties of equilibrium strategies, in this work we focus on the algorithmic aspects of the proposed model. We give an efficient algorithm for computing best responses, and show that while the temporary impact only setting yields a potential game, best response dynamics do not generally converge for the general setting, for which no fast algorithm for (Nash) equilibrium computation is known. This leads us to consider the broader notion of Coarse Correlated Equilibria (CCE), which we show can be computed efficiently via an implementation of Follow the Perturbed Leader (FTPL). While we focus on equilibrium computation, our FTPL implementation learns no-regret strategies in any (adversarial) trading environment. We illustrate the model and our results with an experimental investigation, where FTPL exhibits interesting behavior in different regimes of the relative weighting between temporary and permanent market impact.

1 Introduction

There is both a vast commercial industry and a large quantitative finance literature centered on the problem of optimally executing trades in electronic exchanges under various conditions. Many brokerages and investment banks offer trading services tracking precise benchmarks such as the volume-weighted average price (VWAP) of a stock, or the prices obtained relative to the start of trading. Such services are both informed by and influence research in algorithmic trading (see Related Work below). The overarching goal in algorithmic trading is to acquire or sell a predetermined number of shares in a specified period of time[1], while minimizing the *market impact* incurred by trading — that is, the tendency of trading to push the asset price against the interests of the trader (buying causing prices to rise, selling causing prices to fall).

Despite the fact that trading in modern electronic markets has very obvious strategic aspects, and that traders informally incorporate game-theoretic considerations in their decisions and choices, it is rare to see such considerations explicitly modeled, primarily due to sheer complexity of doing so — the possible strategies or algorithms are virtually innumerable, and there are a vast number of different exchange mechanisms and order types.

In a series of recent papers, Chriss [8, 9, 7, 10] has made significant progress with the introduction and analysis of a stylized but realistic model for competitive trading. In Chriss' model, multiple

[1] Such directives would typically come from higher-level constraints, such as the need to buy or sell shares of different stocks in order to maintain a portfolio that tracks a common index such as the S&P 500, or from a hedge fund's quantitative model that detects and acts on perceived mispricings of assets.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

traders play a game in which each player wishes to acquire either a long or short position in a common asset in a fixed time window, and each wishes to minimize their cost in doing so. As is standard in the finance literature, costs decompose into *temporary* and *permanent* market impact. Broadly speaking, temporary impact models the "mechanical" influence on prices inherent in the continuous double auction common in modern stock exchanges (due to worse prices as one eats further into the limit order books; see [20] for background), while permanent impact models the longer-term "psychological" effects of trading, such as perceptions of the value of the underlying asset. As we shall see in the model, at equilibrium, temporary impact tends to make traders want to avoid each other temporally, while permanent impact tends to make traders want to "front run" (trade before) each other. These competing forces, along with the fact that it may be beneficial to sell some shares en route to acquiring a net long position, make for a rich set of equilibrium strategies, which is the primary focus of Chriss' papers[2].

1.1 Related Work

Chriss begins by establishing the existence of Nash equilibria — since he works in a continuous time and volume model, the pure strategy spaces are infinite, and thus existence does not immediately follow from Nash's celebrated theorem. Chriss imposes continuity and boundary conditions on strategies, which together allow him to prove existence. He then proceeds to examine equilibrium structure and to consider a number of variants of the model. Chriss' model is related to earlier work on optimal trade execution in a non-strategic setting [1]. Here we consider a discrete time and volume version of Chriss' model, for which the pure strategy spaces are finite (though exponential in the time horizon) and thus mixed NE are guaranteed to exist. Since in reality trading must occur in discrete time steps and in whole shares, moving to a discrete model allows us to consider algorithmic and learning issues more precisely, which is our primary interest.

Key to Chriss' model are standard notions of (temporary and permanent) market impact, on which there is a large literature; see [16, 14, 15, 24, 4, 23] for a representative but partial sample. Broadly speaking, this literature considers various models for how trading activity influences asset prices, implications of those models for trading strategies, and empirical validation. A smaller body of work considers the algorithmic aspects of optimal trading, including machine learning approaches [12, 13, 20, 17]. Our work also focuses on algorithmic considerations, but in a game-theoretic setting.

2 Model and Results

In the problem of strategic trading, each trader wishes to build a position in a stock over a period of time. More precisely, we consider n players, where every player $i \in [n]$ wishes to distribute purchases of V_i shares of a stock over T days (or other unit of time). Every player chooses a *trading strategy* that specifies a trading schedule acquiring a target volume V_i by day T. Here negative values of V_i indicate a net short position, and positive values a net long position.

Definition 2.1 (Trading Strategy)**.** *A trading strategy $a : [T] \to \mathbb{Z}$ is a mapping from a time step t to the number of shares held by a player at time t, satisfying $a(0) = 0$ and $a(T) = V$.*

A trading strategy can be equivalently described by the number of shares bought at every time step. Given a trading strategy a, we denote by $a'(t)$ the number of shares bought at time t—i.e. $a'(t) = a(t) - a(t-1)$—so that we can equivalently write $a(t) = \sum_{s=1}^{t} a'(s)$. We will interpret negative values of $a'(t)$ as the number of shares *sold* at time t. We allow strategies that both buy and sell shares, regardless of the desired final net position V.

The action set $\mathcal{A}(V_i)$ of a player i is the set of all trading strategies a satisfying $a(0) = 0$ and $a(T) = V_i$. We will at times further restrict the action sets by implementing upper and lower trading limits—θ_U and θ_L—bounding the number of shares that can be bought at any time step. We define the action set $\mathcal{A}(V_i, \theta_L, \theta_U)$ as the set of all strategies a additionally satisfying $\theta_L \leq a'(t) \leq \theta_U$. We will simply write $\mathcal{A}_i$ when the parameters V_i, θ_L, and θ_U are not important to the discussion. We write $\boldsymbol{a} = (a_1, ..., a_n) \in \prod_{i=1}^{n} \mathcal{A}_i$ to denote an action profile and $\mathcal{A}_{-i} = \prod_{j \neq i} \mathcal{A}_j$ to denote the action space of all players excluding player i.

[2]Chriss' work is closely related to Carlin et al. [5], who study equilibria in similar competitive trading setup.

A player's cost per share purchased will take into account two basic sources of market impact — *temporary impact* and *permanent impact*.

Definition 2.2 (Temporary Impact Cost)**.** *The temporary impact cost of a trading strategy $a_i \in \mathcal{A}_i$ against strategies $a_{-i} \in \mathcal{A}_{-i}$ is:*

$$c^{temp}(a_i, a_{-i}) = \sum_{t=1}^{T} a_i'(t) \sum_{j=1}^{n} a_j'(t)$$

Definition 2.3 (Permanent Impact Cost)**.** *The permanent impact cost of a trading strategy $a_i \in \mathcal{A}_i$ against strategies $a_{-i} \in \mathcal{A}_{-i}$ is:*

$$c^{perm}(a_i, a_{-i}) = \sum_{t=1}^{T} a_i'(t) \sum_{j=1}^{n} a_j(t-1)$$

In other words, temporary impact considers the number of *instantaneous* shares bought/sold by all players at any time step, while permanent impact considers the number of shares bought/sold by all players *prior to* that time step. The cost is then formulated as a linear function of market impact. Following Chriss [8], we define a player's general cost as the sum of their temporary impact cost and their permanent impact cost. We will be able to control the relative contributions of temporary and permanent impact costs via a *market impact coefficient* κ.

Definition 2.4 (Cost of Trading)**.** *Fix $\kappa \geq 0$. The cost of a trading strategy $a_i \in \mathcal{A}_i$ against strategies $a_{-i} \in \mathcal{A}_{-i}$ is given by:*

$$c(a_i, a_{-i}) = \sum_{t=1}^{T} \left(a_i'(t) \sum_{j=1}^{n} a_j'(t) + \kappa \cdot a_i'(t) \sum_{j=1}^{n} a_j(t-1) \right)$$

It is worth noting that this model can be "explained" in terms of assumptions on the underlying limit order book dynamics that mediate all trading activity. More specifically, considering (without loss of generality) only players who wish to buy shares to obtain a long position, if we assume that (a) the distribution of share prices in the sell order book is uniform, and that (b) once consumed, shares in the sell book are never replenished by the arrival of new shares, then we recover Chriss' model. Assumption (a) corresponds to his linear temporary cost model, and assumption (b) corresponds to his linear permanent cost model. Assumption (b) can be viewed as an extreme form of permanent impact, in that any trading activity that drives the price up will never be reversed — the market always revalues the security at the current price level. It is then possible to interpret κ as a *liquidity replenishment* parameter, in that intermediate values of κ model new sell orders arriving at previous price levels at some rate.

While both of these assumptions are stylized and somewhat unrealistic in practice, they at least ground our model in assumptions about the low-level dynamics of the exchanges. They also point to more realistic variants of the model, in which we assume more natural price distributions in the order books (for instance, it is common for much more liquidity to be aggregated near the bid and ask prices, and to thin out away from them), and less extreme replenishment assumptions.

Computing a Best Response. We first prove that the algorithmic problem of computing the best-response trading schedule to the fixed actions of the other players admits an efficient dynamic programming solution.

Definition 2.5 (Best Response)**.** *Consider a player i with action set $\mathcal{A}(V_i, \theta_L, \theta_U)$ and cost function c. The best response of player i to action profile a_{-i} is the action $a_i^* = \arg\min_{a \in \mathcal{A}_i} c(a, a_{-i})$.*

Theorem 2.6 (Informal version of Theorem B.1)**.** *There is an algorithm that computes a player's best response in time $O((\theta_U - \theta_L)^2 T^2)$.*

A Decomposition of the Trading Game. We then give a decomposition of the game that will have implications for equilibrium computation (we defer formal definitions of equilibria concepts to Section A). At a high level, we show that the trading game is a mixture of a potential game and a constant-sum game. While this is the case for any game (see Section C), we show that, interestingly,

the potential game arises from trading under temporary impact only, while the constant-sum game arises essentially from trading under permanent impact only. Thus, the basic structure of the trading game differs with underlying market impact.

First, we show that if we consider temporary market impact only, then the game is a potential game and thus best-response dynamics will converge rapidly to a pure Nash equilibrium (NE).

Theorem 2.7 (Informal version of Theorems C.2 and C.4)**.** *Under only temporary impact, the trading game is a potential game. Therefore, best response dynamics converges to a pure Nash equilibrium. In particular, best response dynamics finds an ε-approximate Nash equilibrium in time $O\left(\frac{n^3\theta^4T^3}{\varepsilon}\right)$, where $\theta = \max\{|\theta_L|, |\theta_U|\}$.*

We then show that (a slight variant of) permanent impact only is a zero-sum game.

Theorem 2.8 (Informal version of Lemma C.7)**.** *The variant of permanent impact cost*

$$c^{\text{perm-avg}}(a_i, a_{-i}) := \frac{1}{2}\sum_{t=1}^{T} a_i'(t) \sum_{j=1}^{n} (a_j(t-1) + a_j(t))$$

defines a constant-sum game.

Finally, we conclude that the general game is a weighted mixture of a potential game and a zero-sum game. When $\kappa = 0$, the game is a potential game. When $\kappa = 2$, the game is constant-sum. For all other κ, the game is a weighted mixture of a potential game and a constant-sum game; the value of κ determines its proximity to either. Later on, we will see that this basic structure is reflected in interesting ways in our experimental evaluations.

Theorem 2.9 (Theorem C.8)**.** *Fix a market impact coefficient κ. Then, the cost of trading can be written as:*

$$c(a_i, a_{-i}) = \left(1 - \frac{\kappa}{2}\right) \cdot c^{\text{temp}}(a_i, a_{-i}) + \kappa \cdot c^{\text{perm-avg}}(a_i, a_{-i})$$

In particular, the classes of potential games and constant-sum games are closed under scalar multiplication, and so by Theorem C.3 and Lemma C.7, the terms in the decomposition correspond to a potential game and a constant-sum game.

Efficient Equilibria Computation. Given that the general game seems to admit no special form, and thus computing Nash equilibria may be intractable, we next turn attention to computing (approximate) coarse correlated equilibria via no-regret dyanmics. The difficulty in our setting is producing a no-regret algorithm that is computationally efficient, given that the space of trading strategies is exponentially large. We prove that despite this, there is a computationally efficient implementation of Follow the Perturbed Leader (FTPL) for our game.

In addition to their computational tractability, CCE are interesting in our setting due to the possibility of higher social welfare, and the suggestion that lightweight correlation at the exchanges themselves might render trading less costly. Beyond equilibrium computation, our implementation of FTPL can be used to learn no-regret strategies in *any* environment, where other traders could be acting adversarially.

Theorem 2.10 (Informal version of Corollary D.5)**.** *There is an instantiation of FTPL such that its empirical play in no-regret dynamics is an ε-approximate coarse correlated equilibrium after $R = O\left(\frac{n^2\theta^5T^6}{\varepsilon^2}\right)$ rounds, with total per-round running time $O(n\theta^2T^2)$. Here $\theta = \max\{|\theta_L|, |\theta_U|\}$.*

Experimental Evaluation. We conclude with an extensive experimental investigation of FTPL dynamics and CCE properties in different regimes of the relative weight on temporary and permanent impact. Despite the fact that there are no general guarantees about the CCE that FTPL will find in arbitrary games, the special structure of our aforementioned decomposition is reflected experimentally, with near-pure NE being found when temporary impact dominates and approximate mixed NE being found in the regime where permanent impact has twice the weight temporary impact. Our results and discussion can be found in Section E.

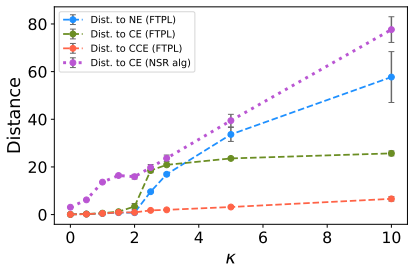


Figure 1: Distances to Nash equilibria, CE, and CCE for varying κ under FTPL dynamics. See Section E for details.

References

[1] R. Almgren and N. Chriss. 2001. Optimal execution of portfolio transactions. In *Journal of Risk*.

[2] S. Arora, E. Hazan, and S. Kale. 2012. The multiplicative weights update method: a meta-algorithm and applications. In *Theory of computing*, Vol. 8 (1), 121-164.

[3] Avrim Blum and Yishay Mansour. 2007. From External to Internal Regret. *Journal of Machine Learning Research* 8, 47, 1307–1324. http://jmlr.org/papers/v8/blum07a.html

[4] JP Bouchaud, M. Mézard, and M. Potters. 2002. Statistical properties of stock order books: empirical results and models. In *Quantitative Finance*.

[5] Bruce Ian Carlin, Miguel Sousa Lobo, and S. Viswanathan. 2007. Episodic Liquidity Crises: Cooperative and Predatory Trading. *The Journal of Finance* 62, 5 (2007), 2235–2274. https://doi.org/10.1111/j.1540-6261.2007.01274.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2007.01274.x

[6] Nicolo Cesa-Bianchi and Gabor Lugosi. 2006. *Prediction, Learning, and Games*. Cambridge University Press.

[7] Neil A. Chriss. 2024. Competitive equilibria in trading. arXiv:2410.13583 [q-fin.TR] https://arxiv.org/abs/2410.13583

[8] Neil A. Chriss. 2024. Optimal position-building strategies in competition. arXiv:2409.03586 [q-fin.TR] https://arxiv.org/abs/2409.03586

[9] Neil A. Chriss. 2024. Position-building in competition with real-world constraints. arXiv:2409.15459 [q-fin.TR] https://arxiv.org/abs/2409.15459

[10] Neil A. Chriss. 2025. Position building in competition is a game with incomplete information. arXiv:2501.01241 [q-fin.TR] https://arxiv.org/abs/2501.01241

[11] Yuval Dagan, Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. 2024. From External to Swap Regret 2.0: An Efficient Reduction for Large Action Spaces. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing* (Vancouver, BC, Canada) *(STOC 2024)*. Association for Computing Machinery, New York, NY, USA, 1216–1222. https://doi.org/10.1145/3618260.3649681

[12] E. Even-Dar, S. Kakade, M. Kearns, and Y. Mansour. 2006. (In)Stability Properties of Limit Order Dynamics. In *Proceedings of the ACM Conference on Electronic Commerce*.

[13] K. Ganchev, M. Kearns, and J. Wortman. 2010. Censored Exploration and the Dark Pool Problem. In *Communications of the ACM*.

[14] J. Gatheral. 2010. No-dynamic-arbitrage and market impact. In *Quantitative Finance*.

[15] Nikolaus Hautsch and Ruihong Huang. 2012. The market impact of a limit order. In *Journal of Economic Dynamics and Control*.

[16] J.Gatheral. 2010. Three models of market impact and data. In *Market Microstructure and High Frequency Data*.

[17] S. Kakade, M. Kearns, Y. Mansour, and L. Ortiz. 2004. Competitive Algorithms for VWAP and Limit Order Trading. In *Proceedings of the ACM Conference on Electronic Commerce*.

[18] Adam Kalai and Santosh Vempala. 2005. Efficient algorithms for online decision problems. *J. Comput. System Sci.* 71, 3 (2005), 291–307. `https://doi.org/10.1016/j.jcss.2004.10.016` Learning Theory 2003.

[19] Dov Monderer and Lloyd S. Shapley. 1996. Potential Games. *Games and Economic Behavior* 14, 1 (1996), 124–143. `https://doi.org/10.1006/game.1996.0044`

[20] Y. Nevmyvaka, M. Kearns, and Y. Feng. 2006. Reinforcement Learning for Optimized Trade Execution. In *Proceedings of the International Conference on Machine Learning*.

[21] Binghui Peng and Aviad Rubinstein. 2024. Fast Swap Regret Minimization and Applications to Approximate Correlated Equilibria. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing* (Vancouver, BC, Canada) *(STOC 2024)*. Association for Computing Machinery, New York, NY, USA, 1223–1234. `https://doi.org/10.1145/3618260.3649691`

[22] Ronnie Sadka. 2006. Momentum and Post-Earnings-Announcement Drift Anomalies: The Role of Liquidity Risk. *Journal of Financial Economics* (2006).

[23] K. Webster. 2023. *Handbook of price impact modeling*. Chapman and Hall/CRC.

[24] Elia Zarinelli, Michele Treccani, J. Doyne Farmer, and Fabrizio Lillo. 2015. Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate. In *Market Microstructure and Liquidity*.

A Model and Preliminatries

The trading game. We begin by describing the problem of strategic trading, where traders wish to build a position in a stock over a period of time. More precisely, we consider n players, where every player $i \in [n]$ wishes to distribute purchases of V_i shares of a stock over T days (or other unit of time). Every player chooses a *trading strategy* that specifies a trading schedule acquiring a target volume V_i by day T. Here negative values of V_i indicate a net short position, and positive values a net long position.

Definition A.1 (Trading Strategy)**.** *A trading strategy $a : [T] \to \mathbb{Z}$ is a mapping from a time step t to the number of shares held by a player at time t, satisfying $a(0) = 0$ and $a(T) = V$.*

A trading strategy can be equivalently described by the number of shares bought at every time step. Given a trading strategy a, we denote by $a'(t)$ the number of shares bought at time t—i.e. $a'(t) = a(t) - a(t-1)$—so that we can equivalently write $a(t) = \sum_{s=1}^{t} a'(s)$. We will interpret negative values of $a'(t)$ as the number of shares *sold* at time t. We allow strategies that both buy and sell shares, regardless of the desired final net position V.

The action set $\mathcal{A}(V_i)$ of a player i is the set of all trading strategies a satisfying $a(0) = 0$ and $a(T) = V_i$. We will at times further restrict the action sets by implementing upper and lower trading limits—θ_U and θ_L—bounding the number of shares that can be bought at any time step. We define the action set $\mathcal{A}(V_i, \theta_L, \theta_U)$ as the set of all strategies a additionally satisfying $\theta_L \leq a'(t) \leq \theta_U$. We will simply write $\mathcal{A}_i$ when the parameters V_i, θ_L, and θ_U are not important to the discussion. We write $\boldsymbol{a} = (a_1, ..., a_n) \in \prod_{i=1}^{n} \mathcal{A}_i$ to denote an action profile and $\mathcal{A}_{-i} = \prod_{j \neq i} \mathcal{A}_j$ to denote the action space of all players excluding player i.

A player's cost per share purchased will take into account two basic sources of market impact — *temporary impact* and *permanent impact*.

Definition A.2 (Temporary Impact Cost)**.** *The temporary impact cost of a trading strategy $a_i \in \mathcal{A}_i$ against strategies $a_{-i} \in \mathcal{A}_{-i}$ is:*

$$c^{temp}(a_i, a_{-i}) = \sum_{t=1}^{T} a_i'(t) \sum_{j=1}^{n} a_j'(t)$$

Definition A.3 (Permanent Impact Cost)**.** *The permanent impact cost of a trading strategy $a_i \in \mathcal{A}_i$ against strategies $a_{-i} \in \mathcal{A}_{-i}$ is:*

$$c^{perm}(a_i, a_{-i}) = \sum_{t=1}^{T} a_i'(t) \sum_{j=1}^{n} a_j(t-1)$$

In other words, temporary impact considers the number of *instantaneous* shares bought/sold by all players at any time step, while permanent impact considers the number of shares bought/sold by all players *prior to* that time step. The cost is then formulated as a linear function of market impact. Following Chriss [8], we define a player's general cost as the sum of their temporary impact cost and their permanent impact cost. We will be able to control the relative contributions of temporary and permanent impact costs via a *market impact coefficient* κ.

Definition A.4 (Cost of Trading)**.** *Fix $\kappa \geq 0$. The cost of a trading strategy $a_i \in \mathcal{A}_i$ against strategies $a_{-i} \in \mathcal{A}_{-i}$ is given by:*

$$c(a_i, a_{-i}) = \sum_{t=1}^{T} \left(a_i'(t) \sum_{j=1}^{n} a_j'(t) + \kappa \cdot a_i'(t) \sum_{j=1}^{n} a_j(t-1) \right)$$

It is worth noting that this model can be "explained" in terms of assumptions on the underlying limit order book dynamics that mediate all trading activity. More specifically, considering (without loss of generality) only players who wish to buy shares to obtain a long position, if we assume that (a) the distribution of share prices in the sell order book is uniform, and that (b) once consumed, shares in the sell book are never replenished by the arrival of new shares, then we recover Chriss' model. Assumption (a) corresponds to his linear temporary cost model, and assumption (b) corresponds to his

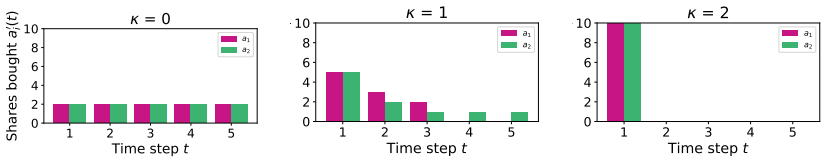


Figure 2: Pure Nash equilibria strategies (a_1, a_2) in the trading game with two players, for different κ. In this example, $T = 5$, $V_1 = V_2 = 10$, $\theta_U = 10$, and $\theta_L = 0$.

linear permanent cost model. Assumption (b) can be viewed as an extreme form of permanent impact, in that any trading activity that drives the price up will never be reversed — the market always revalues the security at the current price level. It is then possible to interpret κ as a *liquidity replenishment* parameter, in that intermediate values of κ model new sell orders arriving at previous price levels at some rate. Carlin et al. [5] attribute permanent impact to a similar phenomenon—that decreases in supply cause increases in price—but do not directly map it onto limit order book dynamics.

While both of these assumptions are stylized and somewhat unrealistic in practice, they at least ground our model in assumptions about the low-level dynamics of the exchanges. They also point to more realistic variants of the model, in which we assume more natural price distributions in the order books (for instance, it is common for much more liquidity to be aggregated near the bid and ask prices, and to thin out away from them), and less extreme replenishment assumptions.

Given a fixed action profile, the best response of player i is a strategy that minimizes cost.

Definition A.5 (Best Response)**.** *Consider a player i with action set $\mathcal{A}(V_i, \theta_L, \theta_U)$ and cost function c. The best response of player i to action profile a_{-i} is the action $a_i^* = \arg\min_{a \in \mathcal{A}_i} c(a, a_{-i})$.*

Equilibria Concepts. We will study several basic equilibria concepts, defined below in increasing generality.

Definition A.6 (Pure Nash Equilibrium)**.** *An action profile $\boldsymbol{a}$ is an ε-approximate pure Nash equilibrium if for all i, $c(a_i, a_{-i}) \leq \min_{a \in \mathcal{A}_i} c(a, a_{-i}) + \varepsilon$. When $\varepsilon = 0$, $\boldsymbol{a}$ is a pure Nash equilibrium.*

Definition A.7 (Mixed Nash Equilibrium)**.** *A profile of (independent) distributions $\mathbf{D} = (D_1 \times ... \times D_n) \in \prod_{i=1}^n \Delta\mathcal{A}_i$ is an ε-approximate mixed Nash equilibrium if for all i, $\mathbb{E}_{\boldsymbol{a} \sim \mathbf{D}}[c(a_i, a_{-i})] \leq \min_{a \in \mathcal{A}_i} \mathbb{E}_{\boldsymbol{a} \sim \mathbf{D}}[c(a, a_{-i})] + \varepsilon$. When $\varepsilon = 0$, we say $\mathbf{D}$ is a mixed Nash equilibrium.*

Definition A.8 (Correlated Equilibrium (CE))**.** *A distribution $\mathbf{D}$ over action profiles is an ε-approximate correlated equilibrium if for all i, for all swap functions $\phi_i : \mathcal{A}_i \to \mathcal{A}_i$, $\mathbb{E}_{\boldsymbol{a} \sim \mathbf{D}}[c(a_i, a_{-i})] \leq \mathbb{E}_{\boldsymbol{a} \sim \mathbf{D}}[c(\phi(a_i), a_{-i})] + \varepsilon$. When $\varepsilon = 0$, $\mathbf{D}$ is a correlated equilibrium.*

Definition A.9 (Coarse Correlated Equilibrium (CCE))**.** *A distribution $\mathbf{D}$ over action profiles is an ε-approximate coarse correlated equilibrium if for all i, $\mathbb{E}_{\boldsymbol{a} \sim \mathbf{D}}[c(a_i, a_{-i})] \leq \min_{a \in \mathcal{A}_i} \mathbb{E}_{\boldsymbol{a} \sim \mathbf{D}}[c(a, a_{-i})] + \varepsilon$. When $\varepsilon = 0$, we say $\mathbf{D}$ is a coarse correlated equilibrium.*

Examples. To provide some intuition of the game, we give some examples of equilibria strategies, under varying market impact coefficients κ. Recall that κ determines the relative contributions of temporary and permanent impact. Figure 2 shows pure Nash equilibria strategy pairs for the *buy-only* setting (i.e. $\theta_L = 0$). In this setting, we see a clear tension between temporary and permanent impact; when players pay only temporary impact cost (i.e. $\kappa = 0$), the tendency is to spread out trading activity to avoid the opponent—and themselves. When players pay only permanent impact cost, the tendency is to trade ahead (in fact, buying everything at $t = 1$ incurs 0 permanent impact cost in our model). Here, $\kappa = 2$ is large enough to induce this behavior. For the intermediary case of $\kappa = 1$, players strike a balance between the two.

Selling complicates the picture. For instance, buying upfront was previously the best *buy-only* strategy for large enough κ. When selling is allowed, players tend to want to sell immediately after their opponent buys (when costs are high) and buy immediately after their opponent sells (when costs are low). Figure 3 gives examples of best response strategies exhibiting this behavior.

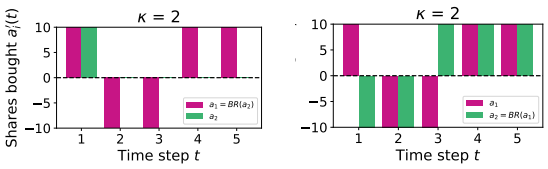


Figure 3: Examples of best response strategies under $\kappa = 2$ when allowing for selling. On the left, a_1 is a best response to a_2; on the right, a_2 is a best response to a_1. Here, $T = 5$, $V_1 = V_2 = 10$, $\theta_U = 10$, and $\theta_L = -10$.

Algorithm 1: Best response over one-step costs (BR)

Input: Target volume V, trading limits θ_L, θ_U, one-step cost function p^t
Output: Best response $a^* = \arg\min_{a \in \mathcal{A}(V,\theta_L,\theta_U)} \phi^T(a)$, where
$\phi^T(a) = \sum_{t=1}^{T} p^t(a(t-1), a'(t))$

for $s = V - \theta_U t$ **to** $V - \theta_L t$ **do**
  Initialize $\mathrm{OPT}(T, s) = p^t(V - s, s)$
  Initialize $\mathrm{BR}(T, s) = s$
for $t = T - 1$ **to** 1 **do**
  for $s = V - \theta_U t$ **to** $V - \theta_L t$ **do**
    Let $\mathrm{OPT}(t, s) = \min_{\theta_L \le k \le \theta_U} (\mathrm{OPT}(t+1, s-k) + p^t(V-s, k))$ and
    $\mathrm{BR}(t, s) = \arg\min_{\theta_L \le k \le \theta_U} (\mathrm{OPT}(t+1, s-k) + p^t(V-s, k))$
for $t = 1$ **to** T `// recover the optimal strategy` a^* **do**
  Let $a^{*\prime}(t) = \mathrm{BR}(t, V)$;
  Let $V = V - a^{*\prime}(t)$;
Return a^* ;

B Computing a Best Response

We begin by giving a dynamic programming algorithm that finds a best response to any profile of trading strategies. The dynamic programming algorithm we present in this section (Algorithm 1) will later serve as a building block in our algorithms for equilibrium computation.

The key insight is that the optimal strategy beginning at any time t depends only on the permanent impact of prior trading activity—which is determined by the number of shares held prior to t. For a strategy that must hold V shares at time T, this can be expressed as the number of remaining shares that need to be bought. That is, if m shares are held before t, then $V - m$ shares must be bought from t onwards. Thus, if we knew the number of remaining shares to be bought, we can define a subproblem solving for the optimal strategy beginning at time t with s shares remaining. In short, then, our dynamic programming algorithm solves for the best response inductively over time steps and remaining shares. The number of inductive steps will determine the computation required—which we will see depends polynomially on θ_L, θ_U, and T.

This idea lends itself to more general cost minimization problems over trading strategies; in fact, we will present and analyze a more general form of the algorithm than what is required to compute a best response in the trading game. In particular, we show how to compute $a^* = \arg\min_{a \in \mathcal{A}_i} \phi^T(a)$ for any cost function $\phi^T(a)$ that can be written as the sum of *one-step costs* p^t that depend on the "state" of trades at time t:

$$\phi^T(a) = \sum_{t=1}^{T} p^t(a(t-1), a'(t))$$

We remark that p^t is defined to take as input $a(t-1)$ rather than $a(t)$ simply for ease of presentation later on — the quantities $a(t-1)$ and $a(t)$ are essentially interchangeable, since $a(t) = a(t-1) + a'(t)$.

Our trading cost can be written in this way. Specifically, we can decompose the cost $c(a_i, a_{-i})$ of a strategy a_i into one-step costs, parameterized by a_{-i} and κ:

$$c^t(a_i(t-1), a_i'(t); a_{-i}, \kappa) = a_i'(t) \sum_{j=1}^{n} a_j'(t) + \kappa \cdot a_i'(t) \sum_{j=1}^{n} a_j(t-1)$$

so that $c(a_i, a_{-i}) = \sum_{t=1}^{T} c^t(a_i(t-1), a_i'(t); a_{-i}, \kappa)$. Thus, instantiating our algorithm with $p^t = c^t$ computes a best response in the trading game. We will later rely on this general form of the algorithm in Section D, when we reduce to a cost minimization problem that can be formulated as the sum of one-step costs.

Theorem B.1. *Given a target volume V, trading limits θ_L, θ_U, and a cost function ϕ^T such that $\phi^T(a) = \sum_{t=1}^{T} p^t(a(t-1), a'(t))$ for some function p^t, Algorithm 1 computes $a^* = \arg\min_{a \in \mathcal{A}(V, \theta_L, \theta_U)} \phi^T(a)$. Moreover, Algorithm 1 runs in time $O((\theta_U - \theta_L)^2 T^2)$.*

The algorithm is simple to describe. Let $\mathrm{OPT}(t, s)$ be the minimum cost to buy s shares beginning at time t. First, observe that at the last time step T, a strategy must buy all remaining shares s. Furthermore, if s shares remain, it must be that $V - s$ shares are held before T. Thus, $\mathrm{OPT}(T, s) = p^t(V - s, s)$. Now we work backwards. We can compute:

$$\mathrm{OPT}(t, s) = \min_k [\mathrm{OPT}(t+1, s-k) + p^t(V - s, k)]$$

Why? The one-step cost of buying k shares at time t with s shares remaining is $p^t(V - s, k)$; the minimum remaining cost is simply the minimum cost of buying $s - k$ shares beginning at the next time step, i.e. $\mathrm{OPT}(t+1, s-k)$. The cost of a best response strategy is then $\mathrm{OPT}(1, V)$; some simple bookkeeping will allow us to recover the optimal strategy.

We present these ideas in more detail below.

Proof of Theorem B.1. We first show that Algorithm 1 finds a best response. As above, let $\mathrm{OPT}(t, s)$ be the minimum cost for a strategy to buy s shares beginning at time t. The optimal cost is therefore $\mathrm{OPT}(1, V)$. Let a denote any strategy in $\mathcal{A}(V, \theta_L, \theta_U)$. Since a satisfies $\theta_L \leq a'(t) \leq \theta_U$ for all t, we have that $\theta_L t \leq a(t) \leq \theta_U t$ and thus, $V - \theta_U t \leq s \leq V - \theta_L t$ for all t.

We now proceed via induction. If s shares remain at the last time step, then it must be that $a(T-1) = V - s$ and $a'(T) = s$. Thus we can define the base case $\mathrm{OPT}(T, s)$ as the cost of buying s shares at time T, given that $V - s$ shares are held up until time T, i.e.:

$$\mathrm{OPT}(T, s) = p^t(V - s, s)$$

The inductive step rests on the following fact: for $t = 1, ..., T-1$ and $s = V - \theta_U t, ..., V - \theta_L t$,

$$\mathrm{OPT}(t, s) = \min_{\theta_L \leq k \leq \theta_U} \big(\mathrm{OPT}(t+1, s-k) + p^t(V - s, k)\big)$$

To see this, observe that if s shares remain at time t, then it must be that $V - s$ shares are held up until time t, i.e. $a(t-1) = V - s$. Suppose k shares are bought at time t—i.e. $a'(t) = k$. Then, the one-step cost incurred at time t is precisely $p^t(V - s, k)$. The optimal remaining cost is the minimum cost to buy the remaining $s - k$ shares beginning at time $t+1$—that is, $\mathrm{OPT}(t+1, s-k)$. The optimal solution to buy s shares beginning at time t buys some number of shares $k \in [\theta_L, \theta_U]$ at time step t. Since the inductive step chooses k to minimize the cost beginning at time t, it is optimal. This proves the inductive step.

Now, for every t, s pair, Algorithm 1 stores the minimizer of the previous expression:

$$\mathrm{BR}(t, s) = \arg\min_{\theta_L \leq k \leq \theta_U} \big(\mathrm{OPT}(t+1, s-k) + p^t(V - s, k)\big)$$

Thus, backtracking starting at $\mathrm{BR}(1, V)$ recovers the number of shares to buy at every time t in an optimal solution to buy V shares starting at $t = 1$, and so recovers an optimal strategy $a^*(t)$.

It remains analyze the running time of Algorithm 1. There are 3 nested iterations. The first iterates through each $t \in [T]$. For each t, it iterates through all values s between $V - \theta_U t$ and $V - \theta_L t$. Then for each value of s, it finds $\mathrm{OPT}(t, s)$ by iterating through all values k between θ_L and θ_U.

Recovering the optimal strategy takes time T. Thus, the running time is:

$$\begin{aligned}\left(\sum_{t=1}^{T}(V-\theta_L t-(V-\theta_U t))(\theta_U-\theta_L)\right)+T &= \left((\theta_U-\theta_L)\sum_{t=1}^{T}(\theta_U t-\theta_L t)\right)+T\\ &= \left((\theta_U-\theta_L)^2\sum_{t=1}^{T}t\right)+T\\ &= \left((\theta_U-\theta_L)^2\cdot\frac{T(T+1)}{2}\right)+T\\ &= O((\theta_U-\theta_L)^2T^2)\end{aligned}$$

which proves the theorem. □

C A Decomposition of the Trading Game

In this section, we give a decomposition of the trading game that will have implications for equilibrium computation. At a high level, we show that the trading game is a mixture of a potential game and a constant-sum game. While this is the case for any game [3] , we show that, interestingly, the potential game arises from trading under temporary impact only, while the constant-sum game arises essentially from trading under permanent impact only. More precisely, we will show that the trading cost decomposes into:

$$c(a_i,a_{-i})=\left(1-\frac{\kappa}{2}\right)\cdot c^{\text{temp}}(a_i,a_{-i})+\kappa\cdot c^{\text{perm-avg}}(a_i,a_{-i})$$

where c^{temp} defines a potential game and $c^{\text{perm-avg}}$ (a slight modification of permanent impact cost) defines a constant-sum game.

This decomposition shows that the basic structure of the trading game differs with underlying market impact[4]. Most saliently, when $\kappa=0$ (i.e. when players face only temporary impact), the game is a potential game, and so simple "best-response dynamics" converge to a pure Nash equilibrium. When $\kappa=2$ (i.e. the contribution of permanent impact is twice that of temporary impact), the game is constant-sum. In this case, the empirical history of "no-regret dynamics" converges to a mixed Nash equilibrium [6]. For all other κ, the game is a weighted mixture of a potential game and a constant-sum game; the value of κ determines its proximity to either. Later on, we will see that this basic structure is reflected in interesting ways in our experimental evaluations.

The remainder of the section is dedicated to substantiating these ideas. We begin by separately analyzing the terms in the decomposition — this coincides with thinking separately about the temporary impact only and permanent impact only regimes. In subsection C.1, we focus on the temporary impact only regime: we show that c^{temp} defines a potential game and provide a simple learning dynamic—best response dynamics—that converge to a pure Nash equilibrium. In subsection C.2, we focus on the permanent impact only regime: we define the variant of permanent impact cost $c^{\text{perm-avg}}$ and show that it is constant-sum. Finally, in subsection C.3, we prove the decomposition of the general trading game.

C.1 The Temporary Impact Regime

Recall that the temporary impact cost is summarized by the instantaneous number of shares bought/sold:

$$c^{\text{temp}}(a_i,a_{-i})=\sum_{t=1}^{T}a_i'(t)\sum_{j=1}^{n}a_j'(t)$$

[3] Any game is the sum of a potential game and zero-sum game (private communication from Aaron Roth): Consider an arbitrary game where player $i\in[n]$ has cost function c_i. We can decompose this into the sum of two games where every player i has cost $c_{i,1}=\frac{\sum_{i=1}^{n}c_i}{n}$ in the first, and $c_{i,2}=\frac{(n-1)c_i-\sum_{j\neq i}c_j}{n}$ in the second. In the first game, every player has the same cost, and so the sum of costs $\sum_{i=1}^{n}c_{i,1}$ is a potential function. In the second game, the sum of costs $\sum_{i=1}^{n}c_{i,2}=0$, since $(n-1)\sum_{i=1}^{n}c_i=\sum_{i=1}^{n}\sum_{j\neq i}c_j$, and so is zero-sum.

[4] Empirical studies suggest that the ratio of temporary to permanent impact varies significantly in markets [22, 5]; our result highlights how incentives can vary with this ratio.

As mentioned in the Introduction, temporary impact can be viewed as modeling the mechanical aspects of trading in the double-auction order book mechanism of modern electronic markets, in which queues or "books" of buy and sell orders are ordered by price, and (say) a buyer demanding immediately liquidity must consumer successive orders with increasing prices in the sell book.

We show that the game defined by c^{temp} is a potential game, and so, by the classical result of Monderer and Shapley [19], the simple procedure of *best response dynamics* converges to a pure Nash equilibria.

Definition C.1 (Potential Game)**.** *Consider an n-player game G where each player $i \in [n]$ chooses actions from an action set $\mathcal{A}_i$ and has cost function c_i. G is an exact potential game if there exists a potential function $\phi : \mathcal{A}_1 \times ... \times \mathcal{A}_n \to \mathbb{R}$ such that for all players $i \in [n]$ and actions $a_i, b_i \in \mathcal{A}_i$,*

$$\phi(b_i, a_{-i}) - \phi(a_i, a_{-i}) = c_i(b_i, a_{-i}) - c_i(a_i, a_{-i})$$

Theorem C.2. *[19] In any finite potential game, best response dynamics converges to a pure Nash equilibrium.*

In best response dynamics, players move sequentially to a beneficial deviation, as long as the strategy profile is not a pure Nash equilibrium. The rationale behind Theorem C.2 is simple. Any deviation strictly decreases a player's cost, and thus the potential function. Since the game is finite, the potential function must reach a minimum, at which point it must be that are no beneficial deviations—that is, players reach a Nash equilibrium. Given this fact, it suffices to produce a potential function for the temporary impact only setting.

Theorem C.3. *Consider an instance of the trading game where for every player $i \in [n]$, the cost of strategy a_i against strategies a_{-i} is given by the temporary impact cost $c^{temp}(a_i, a_{-i})$, i.e. $\kappa = 0$. This is a potential game with potential function:*

$$\phi(\boldsymbol{a}) = \sum_{t=1}^{T} \sum_{i=1}^{n} a_i'(t) \sum_{j \geq i} a_j'(t)$$

Proof. The task is to show that the change in ϕ exactly measures the change in temporary impact cost resulting from a unilateral deviation. For ease of notation, let's define $h(a_i, a_j) \coloneqq \sum_{t=1}^{T} a_i'(t) a_j'(t)$. And so we can write $c^{\text{temp}}(a_i, a_{-i}) = \sum_{j=1}^{n} h(a_i, a_j)$ and $\phi(\boldsymbol{a}) = \sum_{i=1}^{n} \sum_{j \geq i} h(a_i, a_j)$.

Suppose player k deviates from a_k to b_k. Since h is symmetric, i.e. $h(a_i, a_j) = h(a_j, a_i)$, we can write the change in potential as:

$$\begin{aligned}
\phi(b_k, a_{-k}) - \phi(a_k, a_{-k}) &= \sum_{j=1}^{n} h(b_k, a_j) + \sum_{i \neq k} \sum_{j \geq i, j \neq k} h(a_i, a_j) - \sum_{j=1}^{n} h(a_k, a_j) - \sum_{i \neq k} \sum_{j \geq i, j \neq k} h(a_i, a_j) \\
&= \sum_{j=1}^{n} h(b_k, a_j) - \sum_{j=1}^{n} h(a_k, a_j) \\
&= c^{\text{temp}}(b_k, a_{-k}) - c^{\text{temp}}(a_k, a_{-k})
\end{aligned}$$

That is, all terms not involving k cancel out, and the remaining n terms involving k exactly match the change in cost. This proves the theorem.

□

We have thus established that a pure Nash equilibrium exists in the temporary impact only setting and can be found using best response dynamics — but how quickly? To answer this, it is common to settle for an ε-approximate equilibrium. We can bound the number of rounds best response dynamics will run for, as long as each deviation leads to a large enough improve in cost—at least ε. That is, as long as the strategy profile is not an ε-approximate Nash equilibrium, players sequentially move to a strategy that lowers their cost by at least ε. Notice we can implement best response dynamics using Algorithm 1 to sequentially compute best responses, and verifying if it gives a large enough improvement.

Theorem C.4. *Best response dynamics (Algorithm 2) returns an ε-approximate Nash equilibrium in the temporary impact only setting. Moreover, it has running time bounded by $O\left(\frac{n^3 \theta^4 T^3}{\varepsilon}\right)$, where $\theta = \max\{|\theta_L|, |\theta_U|\}$.*

Algorithm 2: ε-approximate Nash equilibrium (temporary impact only)

Input: Target volumes $V_1, ..., V_n$, trading limits θ_L, θ_U
Output: ε-approximate Nash equilibrium $\boldsymbol{a}$
Initialize $\boldsymbol{a} = (a_1, ..., a_n)$ arbitrarily.
Define $c^{\text{temp, t}}(a_i(t-1), a_i'(t); a_{-i}) = a_i'(t) \sum_{j=1}^n a_j'(t)$ to be the one-step temporary cost;
for $i = 1$ **to** n **do**
  Let $\tilde{a}_i \leftarrow \mathsf{BR}(V_i, \theta_L, \theta_U, c^{\text{temp, t}}(a_i(t-1), a_i'(t); a_{-i}))$;
  If $c^{\text{temp}}(\tilde{a}_i, a_{-i}) \leq c^{\text{temp}}(a_i, a_{-i}) - \varepsilon$, set $a_i \leftarrow \tilde{a}_i$;
Return $\boldsymbol{a}$

Proof. By definition, the strategy profile found by the algorithm is an ε-approximate Nash equilibrium. Now, by Theorem C.3, we have that for any player i, $c^{\text{temp}}(a_i, a_{-i}) - c^{\text{temp}}(\tilde{a}_i, a_{-i}) = \phi(a_i, a_{-i}) - \phi(\tilde{a}_i, a_{-i})$, where $\phi(\boldsymbol{a}) = \sum_{t=1}^T \sum_{i=1}^n a_i'(t) \sum_{j \geq i} a_j'(t)$ is the potential function. Thus, for every deviation from a_i to $\tilde{a}_i$, $\phi(a_i, a_{-i}) - \phi(\tilde{a}_i, a_{-i}) \geq \varepsilon$. And so to bound the running time, it suffices to bound the magnitude of ϕ. Since $|a_i'(t)| \leq \theta$ for all players i and time steps t, we can calculate for any $\boldsymbol{a}$:

$$|\phi(\boldsymbol{a})| = \left| \sum_{t=1}^T \sum_{i=1}^n a_i'(t) \sum_{j \geq i} a_j'(t) \right| \leq \sum_{t=1}^T \sum_{i=1}^n \sum_{j \geq i} \theta^2 = \frac{n(n+1)T\theta^2}{2}$$

Therefore the algorithm halts after at most $\frac{2n(n+1)T\theta^2}{2\varepsilon} = \frac{n(n+1)T\theta^2}{\varepsilon}$ deviations. Each deviation is found using at most n calls to Algorithm 1. Thus, plugging in the guarantees of Algorithm 1 (Theorem B.1) bounds the total running time. □

We conclude this discussion by verifying that the general trading game is *not* a potential game; we show that best response dynamics can cycle in the presence of both temporary and permanent impact. The intuition is that temporary and permanent impact can create counteracting forces. As we previously discussed, temporary impact causes players to want to spread out their trades so as to avoid buying many shares at any time step. On the other hand, permanent impact causes players to want to trade ahead of everyone else. The tension between the two behavior—spreading out trades and trading ahead—can cause best response dynamics to oscillate.

Theorem C.5. *The general trading game is not a potential game.*

Proof. Theorem C.2 tells us that for any finite potential game, best response dynamics is guaranteed to converge. Thus it suffices to give an instance for which best response dynamics does not converge. Below we show an instance with $T = 5$, $\kappa = 1$, and two players, both with the action set $\mathcal{A}(V)$ for $V = 5$.

Consider a run of best response dynamics, where player 1's strategy a_1 is initialized to be:

$$a_1(1) = 2, a_1(2) = 2, a_1(3) = 1, a_1(4) = 0, a_1(5) = 0$$

and player 2's strategy a_2 is initialized to be:

$$a_2(1) = 1, a_2(2) = 1, a_2(3) = 1, a_2(4) = 1, a_2(5) = 1$$

Now, player 2 can decrease his cost against a_1 by playing a_2', where:

$$a_2'(1) = 3, a_2'(2) = 1, a_2'(3) = 0, a_2'(4) = 0, a_2'(5) = 1$$

We have that $c(a_2, a_1) = 36$ while $c(a_2', a_1) = 33$. Then, player 1 can decrease her cost against a_2' by playing a_1', where:

$$a_1'(1) = 2, a_1'(2) = 1, a_1'(3) = 1, a_1'(4) = 1, a_1'(5) = 0$$

We have that $c(a_1, a_2') = 35$ while $c(a_1', a_2') = 34$. Then, player 2 can decrease his cost against a_1' by playing a_2'', where:

$$a_2''(1) = 2, a_2''(2) = 2, a_2''(3) = 1, a_2''(4) = 0, a_2''(5) = 0$$

We have that $c(a_2', a_1') = 32$ while $c(a_2'', a_1') = 31$. Note that $a_2'' = a_1$, and so best response dynamics will cycle, i.e. it will not converge. This completes the proof. □

C.2 The Permanent Impact Regime

Next, turning to the permanent impact only setting, we show that permanent impact essentially induces a constant-sum game. More accurately, while permanent impact cost c^{perm} is *not* constant-sum, a semantic-preserving variant of c^{perm} *is*. The following example demonstrates that c^{perm} is not constant-sum.

Example C.6. *Consider the trading game with two players. Suppose both players buy all $V = V_1 = V_2$ shares at $t = 1$ and 0 shares at every step afterwards. Then the sum of permanent impact costs for both players is 0. On the other hand, suppose both players buy V/T shares at each time step. Then the sum of permanent impact costs is*

$$2\sum_{t=1}^{T}\frac{V}{T}\left(\frac{2V}{T}\cdot(t-1)\right)=\frac{4V^2}{T^2}\sum_{t=1}^{T}t-\frac{4V^2}{T}=\frac{4V^2}{T^2}\cdot\frac{T(T+1)}{2}-\frac{4V^2}{T}=2V^2-\frac{2V^2}{T}$$

which approaches $2V^2$ as T becomes large. Thus the permanent impact only setting is not constant-sum.[5]

Now, we consider a slight variant of permanent impact cost that is in fact constant-sum [6]. We define

$$c^{\text{perm-avg}}(a_i,a_{-i}) \coloneqq \frac{1}{2}\sum_{t=1}^{T}a_i'(t)\sum_{j=1}^{n}(a_j(t-1)+a_j(t))$$

to be the cost averaging the permanent impact contribution from the previous and current time step.

Lemma C.7. *The variant of permanent impact cost $c^{\text{perm-avg}}$ satisfies the following: for any action profile $\boldsymbol{a}\in\mathcal{A}(V_1)\times\ldots\times\mathcal{A}(V_n)$, we have:*

$$\sum_{i=1}^{n}c^{\text{perm-avg}}(a_i,a_{-i})=\frac{1}{2}\left(\sum_{i=1}^{n}V_i\right)^2$$

Proof. By expanding out $a_i'(t)$ for every i and rearranging the summations, we compute:

$$\begin{aligned}
\sum_{i=1}^{n}c^{\text{perm-avg}}(a_i,a_{-i}) &= \sum_{i=1}^{n}\frac{1}{2}\sum_{t=1}^{T}a_i'(t)\sum_{j=1}^{n}(a_j(t-1)+a_j(t))\\
&= \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}(a_i(t)-a_i(t-1))\sum_{j=1}^{n}(a_j(t-1)+a_j(t))\\
&= \frac{1}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}a_i(t)-\sum_{i=1}^{n}a_i(t-1)\right)\left(\sum_{j=1}^{n}(a_j(t-1)+\sum_{j=1}^{n}a_j(t)\right)\\
&= \frac{1}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}a_i(t)-\sum_{i=1}^{n}a_i(t-1)\right)\left(\sum_{i=1}^{n}(a_i(t-1)+\sum_{i=1}^{n}a_i(t)\right)\\
&= \frac{1}{2}\sum_{t=1}^{T}\left(\left(\sum_{i=1}^{n}a_i(t)\right)^2-\left(\sum_{i=1}^{n}a_i(t-1)\right)^2\right)
\end{aligned}$$

where the second-to-last step switches the indexing notation and the last step follows from the identity $(a-b)(a+b)=a^2-b^2$. Now, expanding out the telescoping sum, this quantity equals:

$$\frac{1}{2}\left(\left(\sum_{i=1}^{n}a_i(T)\right)^2-\left(\sum_{i=1}^{n}a_i(0)\right)^2\right)=\frac{1}{2}\left(\sum_{i=1}^{n}V_i\right)^2$$

[5] In fact, using the same example, we can show that the general game is not constant-sum. If both players buy all V shares upfront, the sum of temporary impact costs for both players is $4V^2$, and so the sum of temporary and permanent impact costs is $4V^2$. On the other hand, if both players buy V/T shares at each time step, then the sum of temporary costs is $2\sum_{t=1}^{T}(V/T)(2V/T)=4V^2/T$, which approaches 0 as T becomes large. So the sum of temporary and permanent impact costs approaches $\kappa\cdot 2V^2$ as T becomes large. Thus the general game is not zero-sum for $\kappa\neq 2$.

[6] Andrew Bennett, private communication.

by the boundary conditions $a_i(0) = 0$ and $a_i(T) = V_i$ for all i. □

C.3 Decomposition Theorem

Finally, we show that the general cost of trading c can be written as a weighted sum of the temporary impact cost c^{temp} and the time-averaged variant of permanent impact cost $c^{\text{perm-avg}}$. This implies that the general setting is a mixture of a potential game—coinciding with the temporary impact regime—and a constant-sum game—coinciding with (roughly) the permanent impact regime.

Theorem C.8. *Fix a market impact coefficient κ. Then, the cost of trading can be written as:*

$$c(a_i, a_{-i}) = \left(1 - \frac{\kappa}{2}\right) \cdot c^{\textit{temp}}(a_i, a_{-i}) + \kappa \cdot c^{\textit{perm-avg}}(a_i, a_{-i})$$

In particular, the classes of potential games and constant-sum games are closed under scalar multiplication, and so by Theorem C.3 and Lemma C.7, the terms in the decomposition correspond to a potential game and a constant-sum game.

Proof. We compute:

$$\begin{aligned}
c(a_i, a_{-i}) &= c^{\text{temp}}(a_i, a_{-i}) + \kappa \cdot c^{\text{perm}}(a_i, a_{-i}) \\
&= \sum_{t=1}^{T} a_i'(t) \sum_{j=1}^{n} a_j'(t) + \kappa \sum_{t=1}^{T} a_i'(t) \sum_{j=1}^{n} a_j(t-1) \\
&= \sum_{t=1}^{T} a_i'(t) \sum_{j=1}^{n} a_j'(t) \\
&\quad + \sum_{t=1}^{T} \left(\frac{\kappa}{2} a_i'(t) \sum_{j=1}^{n} a_j(t-1) + \frac{\kappa}{2} a_i'(t) \sum_{j=1}^{n} a_j'(t) - \frac{\kappa}{2} a_i'(t) \sum_{j=1}^{n} a_j'(t) + \frac{\kappa}{2} a_i'(t) \sum_{j=1}^{n} a_j(t-1) \right) \\
&= \sum_{t=1}^{T} a_i'(t) \sum_{j=1}^{n} a_j'(t) + \sum_{t=1}^{T} \left(\frac{\kappa}{2} a_i'(t) \sum_{j=1}^{n} a_j(t-1) + \frac{\kappa}{2} a_i'(t) \sum_{j=1}^{n} a_j(t) - \frac{\kappa}{2} a_i'(t) \sum_{j=1}^{n} a_j'(t) \right) \\
&= \left(1 - \frac{\kappa}{2}\right) \sum_{t=1}^{T} a_i'(t) \sum_{j=1}^{n} a_j'(t) + \frac{\kappa}{2} \sum_{t=1}^{T} a_i'(t) \sum_{j=1}^{n} (a_j(t-1) + a_j(t)) \\
&= \left(1 - \frac{\kappa}{2}\right) \cdot c^{\text{temp}}(a_i, a_{-i}) + \kappa \cdot c^{\text{perm-avg}}(a_i, a_{-i})
\end{aligned}$$

as desired. □

D Efficient Equilibria Computation in the General Game

We have just established that the general trading game admits a meaningful decomposition; however, this has no immediate implications for tractable equilibria computation. In this section, we relax our goal to the computation of CCE, a broader class of equilibria that encapsulates Nash equilibria. We show how to efficiently implement *no-regret dynamics*, for which the empirical history of play converges to a joint distribution that is an (approximate) CCE. In no-regret dynamics, each player chooses a sequence of (randomized) actions by running a regret-minimizing algorithm over R rounds.

Definition D.1 (Regret)**.** *The (average) regret of a player i who chooses a sequence of strategies $a_{i,1}, ..., a_{i,R} \in \mathcal{A}_i$ is defined as:*

$$Reg_i(R) = \max_{a_i \in \mathcal{A}_i} \frac{1}{R} \sum_{r=1}^{R} (c(a_{i,r}, a_{-i,r}) - c(a_i, a_{-i,r}))$$

where $a_{-i,r}$ is the profile of strategies chosen by players excluding i at round r.

No regret encodes the idea that if a player looks back at the history of play, they would find that no deviation to a fixed strategy would have improved their cost. It thus follows that if each player has regret bounded by ε, the time-averaged, empirical distribution of play is an ε-approximate CCE.

The difficulty in our setting is producing a no-regret algorithm that is computationally efficient, given that the space of trading strategies is exponentially large. A classic no-regret algorithm is the Multiplicative Weights algorithm (see [2] for an extensive survey), which achieves regret decreasing at a rate of $O(\sqrt{\ln|\mathcal{A}|/R})$. Thus, if all players run their own Multiplicative Weights algorithm, the empirical distribution of play converges to an ε-approximate CCE after $O(\ln|\mathcal{A}|/\varepsilon^2)$ rounds. However, running Multiplicative Weights will be computationally expensive for our problem: it maintains an explicit distribution over the strategy space and so requires per-round computation that is linear in the number of strategies—which, for our setting, is exponential in T.

We show that an instantiation of the Follow The Perturbed Leader (FTPL) algorithm [18] can be used to implement no-regret dynamics efficiently—the number of rounds required to reach an approximate equilibrium and the per-round runtime is polynomial in our problem parameters. We note that FTPL maintains no-regret guarantees in adversarial environments, so while our main focus is on the joint execution of FTPL by all players, our instantiation of FTPL can be used by a *single* player to obtain vanishing regret in *any* environment, where other players could be acting adversarially. Next we introduce FTPL and state its regret guarantees.

FTPL preliminaries. FTPL (Algorithm 3) is a no-regret algorithm for *online linear optimization* (OLO) problems, defined by a learner with strategy space $\mathcal{F} \subseteq \mathbb{R}^d$ and an adversary with strategy space $\mathcal{H} \subseteq \mathbb{R}^d$. At every round $r \in [R]$, the learner chooses a strategy $f_r \in \Delta\mathcal{F}$ and the adversary chooses a strategy $h_r \in \mathcal{H}$. The learner then observes h_r and incurs cost $\langle f_r, h_r\rangle$.

For every round r, let $H_r = \sum_{s=1}^{r-1} h_s$ be the cumulative cost observed so far. To run FTPL, the learner best responds to a noisy version of H_r. In particular, the learner optimizes over the cost $\langle f, H_r + N_r\rangle$, where N_r is chosen uniformly at random from $[0,\eta]^d$. Rephrased, the learner samples from a distribution given by the randomized best response. Crucially, this distribution is maintained only implicitly (unlike e.g. Multiplicative Weights); for low-dimensional problems, then, the brunt of the computation lies in the optimization/best response step.

Theorem D.2. *[18] Let* $D = \max_{f,f'\in\mathcal{F}} \|f - f'\|_1, M = \max_{h\in\mathcal{H}} \|h\|_1$, *and* $C = \max_{f\in\mathcal{F},h\in\mathcal{H}} |\langle f, h\rangle|$. *Against any adversary's choice of strategies* $h_1, ..., h_R$, *FTPL (Algorithm 3) with noise parameter* $\eta = \sqrt{\frac{2MCR}{D}}$ *obtains regret:*

$$\max_{f\in\mathcal{F}} \frac{1}{R}\sum_{r=1}^{R} (\mathbb{E}[\langle f_r, h_r\rangle] - \langle f, h_r\rangle) \leq 2\sqrt{\frac{DMC}{R}}$$

where the expectation is taken over the noise vectors.

An OLO formulation of the trading game. In order to implement FTPL, we must first cast our problem as an instance of OLO. We will take the perspective of player i (who corresponds to the learner) computing a no regret strategy against other players (who, in aggregate, correspond to the adversary). Although players' cost functions are *not* linear in their actions, we show how to "linearize" the problem by constructing higher-dimensional representations of the strategy spaces for the learner and adversary. Broadly speaking, we use the fact that the cost function is linear in the opponents' strategies a_{-i} and introduce dimensions to represent nonlinearities in a_i—in particular the product relationships. More specifically, the learner will play over the space $\mathcal{F}_{\mathcal{A}_i} = \{f(a_i)\}_{a_i\in\mathcal{A}_i} \subseteq \mathbb{R}^{2T}$ and the adversary will play over the space $\mathcal{H}_{\mathcal{A}_{-i}} = \{h(a_{-i})\}_{a_{-i}\in\mathcal{A}_{-i}} \subseteq \mathbb{R}^{2T}$, where h and f apply

the following transformations:

$$f(a_i) = \begin{bmatrix} a_i'(1) \\ \vdots \\ a_i'(t) \\ \vdots \\ a_i'(T) \\ a_i'(1)(a_i'(1) + \kappa a_i(0)) \\ \vdots \\ a_i'(t)(a_i'(t) + \kappa a_i(t-1)) \\ \vdots \\ a_i'(T)(a_i'(T) + \kappa a_i(T-1)) \end{bmatrix} \quad , \quad h(a_{-i}) = \begin{bmatrix} \sum_{j \neq i} a_j'(1) + \kappa a_j(0) \\ \vdots \\ \sum_{j \neq i} a_j'(t) + \kappa a_j(t-1) \\ \vdots \\ \sum_{j \neq i} a_j'(T) + \kappa a_j(T-1) \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Note that the dimension of the strategy space is only larger by a factor of 2. The transformation is cost-preserving: the learner's cost of playing $f(a_i)$ against $h(a_{-i})$ in the OLO problem matches their cost of playing a_i against a_{-i} in the trading game:

$$\begin{aligned} \langle f(a_i), h(a_{-i}) \rangle &= \sum_{t=1}^{T} a_i'(t) \sum_{j \neq i} (a_j'(t) + \kappa a_j(t-1)) + \sum_{t=1}^{T} a_i'(t)(a_i'(t) + \kappa a_i(t-1)) \\ &= \sum_{t=1}^{T} \left(a_i'(t) \sum_{j=1}^{n} a_j'(t) + \kappa a_i'(t) \sum_{j=1}^{n} a_j(t-1) \right) \\ &= c(a_i, a_{-i}) \end{aligned}$$

Algorithm 3: FTPL

for $r = 1$ **to** R **do**
- Let $H_r = \sum_{s=1}^{r-1} h_s$ be the cumulative cost so far;
- Let $N_r \sim [0, \eta]^d$ be a noise vector chosen uniformly at random;
- Choose the strategy $f_r = \arg\min_{f \in \mathcal{F}} \langle f, H_r + N_r \rangle$;
- Observe h_r;

With this instantiation in hand, we can now appeal to the guarantees of FTPL in the following corollary to Theorem D.2.

Corollary D.3. *In our instantiation, the quantities D, M, and C are polynomial in n, T, and θ, where $\theta = \max\{|\theta_L|, |\theta_U|\}$. In particular, we have that $D \leq O(\theta^2 T^2)$, $M \leq O((n-1)\theta T^2)$, and $C \leq O(n\theta^2 T^2)$. Plugging this in, FTPL obtains regret bounded by $O\left(\frac{n\theta^{5/2}T^3}{\sqrt{R}}\right)$ in our instantiation.*

It remains to consider how to solve the (randomized) best response problem of FTPL in our instantiation. Observe that there is a one-to-one correspondence between strategies $a_i \in \mathcal{A}_i$ and $f(a_i) \in \mathcal{F}$, and so we can speak interchangeably about choosing strategies a_i and $f(a_i)$. Thus, we can write the best response problem as finding:

$$a_{i,r} = \arg\min_{a_i \in \mathcal{A}_i} \langle f(a_i), H_r + N_r \rangle$$

Using the notation v^k for the k^{th} coordinate of a vector v, observe that we can write:

$$\begin{aligned} \langle f(a_i), H_r + N_r \rangle &= \sum_{k=1}^{2T} f(a_i)^k (H_r + N_r)^k \\ &= \sum_{t=1}^{T} a_i'(t)(H_r + N_r)^t + \sum_{t=1}^{T} a_i'(t)(a_i'(t) + \kappa a_i(t-1))(H_r + N_r)^{T+t} \end{aligned}$$

Notice that once we have fixed H_r and N_r, the cost at each step is solely a function of $a_i(t-1)$ and $a_i'(t)$. Thus, we can invoke Algorithm 1 as a subroutine, instantiated with the one-step cost:

$$p_r^t(a_i(t-1), a_i'(t)) \coloneqq a_i'(t)(H_r + N_r)^t + a_i'(t)(a_i'(t) + \kappa a_i(t-1))(H_r + N_r)^{T+t}$$

Then, since computing H_r and N_r at every round can be done in time $2T$, the running time of FTPL directly inherits from the guarantees of Algorithm 1.

Corollary D.4. *Our instantiation of FTPL has per-round running time $O(\theta^2 T^2)$, where $\theta = \max\{|\theta_L|, |\theta_U|\}$.*

No-regret dynamics. Finally, to implement no-regret dynamics, every player $i \in [n]$ maintains a copy of FTPL (Algorithm 3). In rounds $r \in [R]$, every player simultaneously draws a strategy $a_{i,r}$ from the distribution maintained by their copy of FTPL (therefore ensuring that each player's randomness is private). Then, every player observes the full action profile $(a_{1,r}, ..., a_{n,r})$ and updates their copy of FTPL with the cost vector $h(a_{-i,r})$.

Corollary D.5. *For every player $i \in [n]$, let $a_{i,1}, ..., a_{i,R}$ be draws from the distributions maintained by FTPL in no-regret dynamics, set with noise parameter $\eta = nT\sqrt{2\theta R}$, where $\theta = \max\{|\theta_L|, |\theta_U|\}$. Let $\mathbf{D}$ be the empirical distribution over the realized action profiles $\boldsymbol{a}_1, ..., \boldsymbol{a}_R$, where $\boldsymbol{a}_r = (a_{1,r}, ..., a_{n,r})$. Then, $\mathbf{D}$ is an ε-approximate coarse correlated equilibrium after $R = O\left(\frac{n^2\theta^5 T^6}{\varepsilon^2}\right)$ rounds of no-regret dynamics, with total per-round running time $O(n\theta^2 T^2)$.*

In Appendix A, we investigate computation of approximate CE via no-*swap*-regret dynamics.

E Experiments

In light of our theoretical results, it is natural to ask how quickly no-regret dynamics converge to an approximate CCE in actual implementation, and what the approximate equilibria look like — in particular, are they "close" to the stronger notion of Nash equilibria? We empirically investigate these questions under different regimes of market impact, which dictate how "close" or "far'" the game is from a potential game and a zero-sum game. The code can be found here.

Parameter settings. We implement no-regret dynamics between two players using our instantiation of FTPL as described in Section D. Throughout, we will fix the setting of $T = 5, V_1 = V_2 = 10, \theta_L = -5, \theta_U = 5$ (thus both players have 5 time periods in which to acquire a net long position of 10 shares, and are able to buy or (short) sell up to 5 shares at each step) while varying the market impact coefficient κ. For each setting of κ we execute 100 runs of no-regret dynamics, each consisting of 2500 rounds. Recall that FTPL takes in an additional noise parameter η. Using the theoretical guideline of $\eta \approx \sqrt{\text{number of rounds}}$, we choose $\eta = 50$.

E.1 Convergence Rate

First we examine how regret evolves as no-regret dynamics progresses, in order to evaluate the speed of convergence in our implementation. In Figures 4 and 5, we show cumulative and average/per-round regret (respectively) as a function of rounds of no-regret dynamics for different settings of κ. We find that average regret converges to 0 (and so the empirical distribution converges to a coarse correlated equilibria) more rapidly that our theory suggests; while our asymptotic convergence rates scale as $O(T^6/\varepsilon^2)$, we see that average regret (i.e. distance to coarse correlated equilibria) flattens out after 500-1000 rounds for all settings of κ.

In Figure 4, we see that regret behaves somewhat differently over the course of FTPL for different κ. Most notably, for $\kappa = 2$, we see that regret oscillates. As a whole, as κ increases, the shape of regret transitions from quickly flattening out, to oscillating, to quickly flattening out again. However, the individual trajectories at small κ are quite smooth, while behavior becomes more volatile at larger κ. These findings reflect changes in the game's underlying structure — as we saw in Section C, the game morphs from being a potential game (at $\kappa = 0$) to a constant-sum game (at $\kappa = 2$).

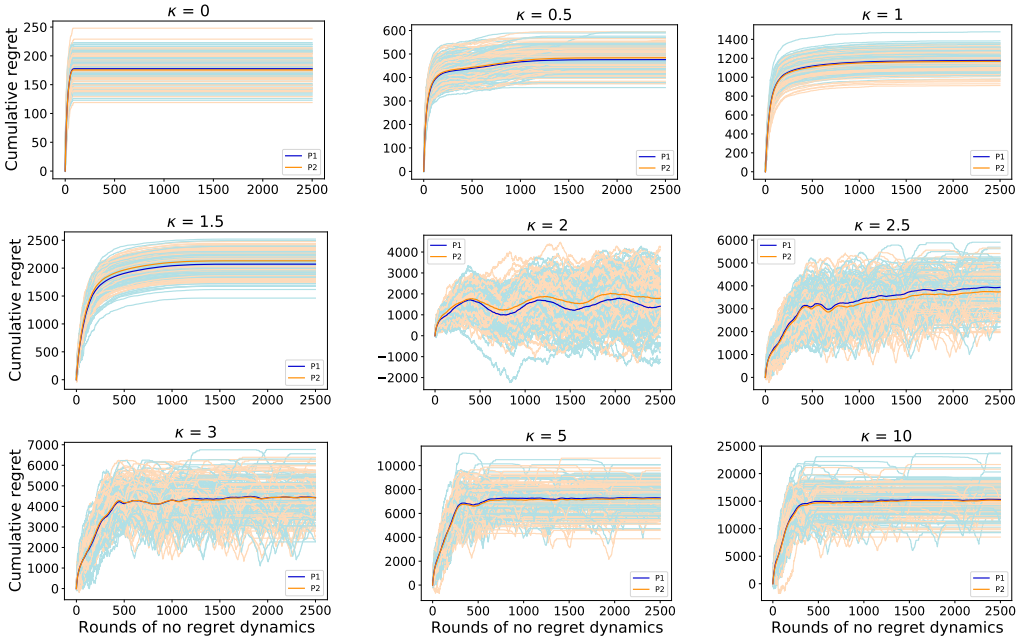


Figure 4: Cumulative regrets of players 1 and 2 for varying κ, averaged across 100 runs of no-regret dynamics. Faint lines represent individual runs, dark lines represent averages.

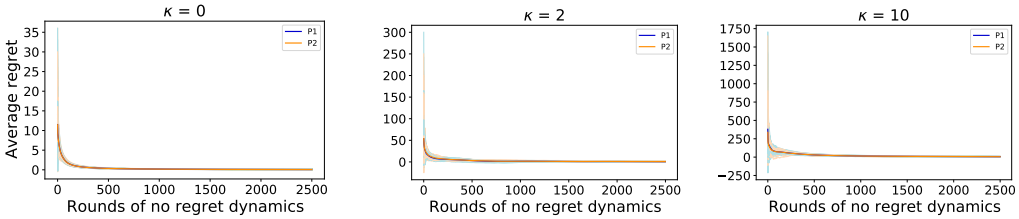


Figure 5: Average/per-round regrets of players 1 and 2 for varying κ, averaged across 100 runs of no-regret dynamics. We exclude other κ for concision; the curves (as shown in this manner) look fairly identical.

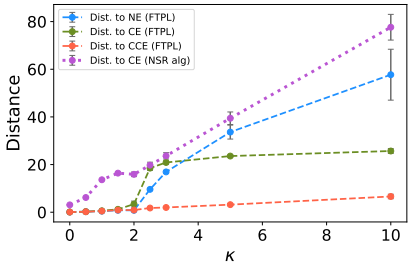
E.2 Equilibria Properties

We have established that no-regret dynamics converge to approximate coarse correlated equilibria fairly quickly. Now we ask: What do the approximate equilibria found by no-regret dynamics look like? Can we say anything more interesting or specific about these equilibria; in particular, are they close to stronger forms of equilibria like Nash equilibria?

Figure 6 answers the questions: how close is the outputted joint distribution to the stronger notion of CE, and how close are the outputted marginal distributions to a mixed Nash equilibrium? We measure these distances directly using the definitions:

$$\text{Dist. to NE} = \max_{i=1,2} \left[\mathbb{E}_{\substack{a_1 \sim D_1 \\ a_2 \sim D_2}} [c(a_i, a_{-i})] - \min_{a^* \in \mathcal{A}} \mathbb{E}_{a_{-i} \sim D_{-i}} [c(a^*, a_{-i})] \right]$$

$$\text{Dist. to CE} = \max_{i=1,2} \left[\mathbb{E}_{(a_1,a_2) \sim \mathbf{D}} [c(a_i, a_{-i})] - \min_{\phi : \mathcal{A} \to \mathcal{A}} \mathbb{E}_{(a_1,a_2) \sim \mathbf{D}} [c(\phi(a_i), a_{-i})] \right]$$



Our results show that when the game interpolates between a potential game and zero-sum game ($\kappa \leq 2$), FTPL finds almost-exact CE and mixed Nash equilibria; in fact the distances closely mirror the distances to CCE, which is theoretically guaranteed to be low. When the game is zero-

Figure 6: Distances to Nash equilibria and CE for varying κ. We use distance

sum (i.e. when $\kappa = 2$), no-regret dynamics is guaranteed to converge to a mixed Nash equilibrium [6]. Thus for $\kappa \geq 2$, the figure presents an intuitive relationship: as the game becomes less like a zero-sum game (i.e. as κ increases beyond 2), the distance to NE found by FTPL dynamics also increases. Meanwhile, for larger κ, the distance to CE increases sharply then plateaus.

Figure 6 also shows the distance to to CE of the joint distribution outputted by a recent no-*swap*-regret algorithm [11, 21], which guarantees convergence to an approximate CE. We see that although FTPL dynamics only guarantees convergence to approximate CCE, its joint play is in fact closer to CE than an implementation of no-swap-regret dynamics for all κ. We formally present the no-swap-regret algorithm and give implementation details in Appendix A.

Next we measure how "correlated" the joint history of play is. Recall that a CCE is a mixed Nash equilibria if its joint distribution can be written as a product distribution—that is, each player's actions can be drawn independently from their own marginal distributions. Since the strategy spaces are not numeric but discrete, combinatorial objects, we cannot measure correlations between player actions in the standard way, but instead will examine the total variation (TV) distance between the joint distribution returned by no-regret dynamics and the product of each player's marginal distribution. More specifically, let $\mathbf{D}$ be the empirical joint distribution over the realized action pairs $(a_{1,1}, a_{2,1}), ..., (a_{1,R}, a_{2,R})$. Let D_1 be the marginal distribution over the first player's actions and D_2 be the marginal distribution over the second player's actions. We compute the TV distance between $\mathbf{D}$ and $D_1 \times D_2$ as follows:

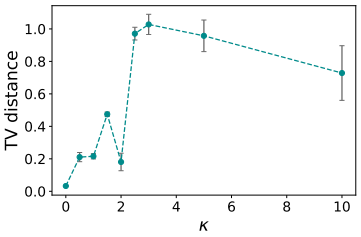


Figure 7: TV distances between outputted joint distribution and product of marginal distributions, for varying κ. The plot shows means and std. deviations over 100 runs of no-regret dynamics.

$$TV(\mathbf{D}, D_1 \times D_2) = \sum_{(a_1,a_2) \in \text{supp}(\mathbf{D})} \left| \Pr_{\mathbf{D}}[(a_1, a_2)] - \Pr_{D_1}[a_1] \cdot \Pr_{D_2}[a_2] \right|$$

In the sequel, we will refer to this distance informally as "correlation" between player strategies. Figure 7 shows TV distances for varying κ. For each setting of κ, we report the average TV distance computed over 100 runs of no-regret dynamics. As expected, TV distance is low for the special case of $\kappa = 2$ (no-regret dynamics are known to converge to Nash equilibria in zero/constant-sum games). The TV distance is particularly low for $\kappa = 0$ — in our subsequent results, we see that this can be explained by the fact that when $\kappa = 0$, no-regret dynamics finds a *pure* Nash equilibrium fairly quickly (recall that pure Nash equilibria are guaranteed to exist in this regime). For large κ, the approximate coarse correlated equilibria found by no-regret dynamics exhibit high correlation.

In Figure 8, we take a closer look at the strategy pairs played by FTPL over one run of no-regret dynamics (we note that although we show the outputs of just one run, the behavior is typical across runs). These visualizations help explain some phenomena we find above. First, for small κ, the equilibria are in fact "close" to a *pure* Nash equilibrium. Specifically, for $\kappa = 0, 0.5$, and 1.5, no-regret dynamics converges to consistently plays a pure Nash equilibrium after some number of rounds. For $\kappa = 0$, this strategy pair is reached fairly quickly. This helps explain why regret stabilizes rapidly and TV distances are low for small κ. For $\kappa = 2$, players oscillate between playing, still, a small subset of actions. For higher κ, we see oscillatory behavior, albeit longer-lived. We note that for $\kappa = 5$ and 10, the strategy pairs found by the end of 2500 rounds are *not* pure Nash equilibria (even though FTPL might appear to have stabilized), indicating that players might continue to cycle between strategy pairs as no-regret dynamics progresses.

Finally, we touch on the *social welfare* of the equilibria found by no-regret dynamics. Recall that expected welfare is defined as $\sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{a} \in \mathbf{D}}[c(a_i, a_{-i})]$ (since we are talking about costs, *lower* welfare is better). Figure 9 shows the welfare of the approximate coarse correlated equilibria found by no-regret dynamics for varying κ. In general, we would expect costs — and thus welfare — to increase as κ increases. Interestingly, we see an inflection point at $\kappa = 2$; welfare increases at a

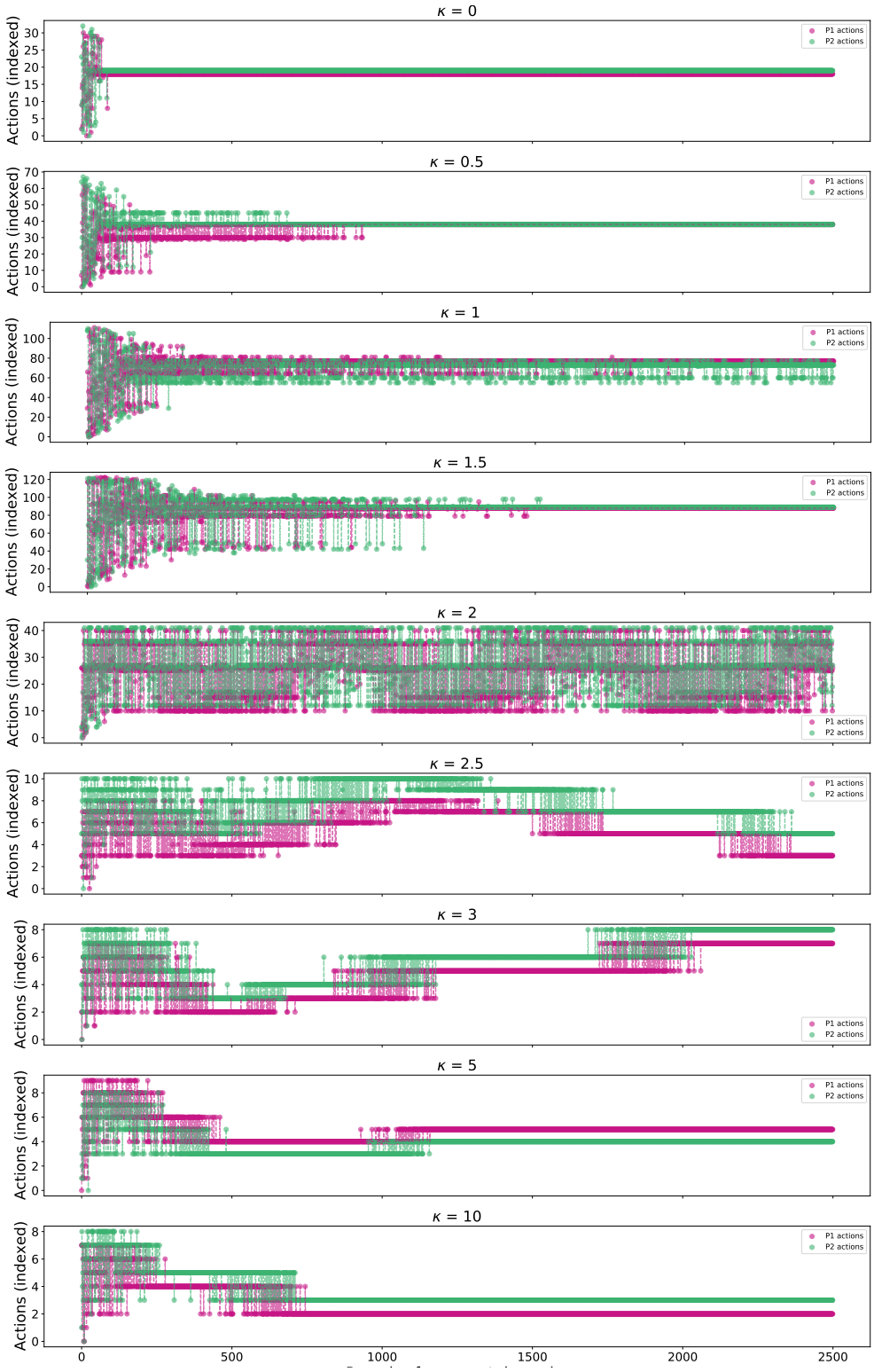


Figure 8: Actions played by players 1 and 2 over one run of no-regret dynamics for varying κ (note: actions are indexed in no particular order and indices might differ from plot to plot; the purpose is to show the progression of actions played)

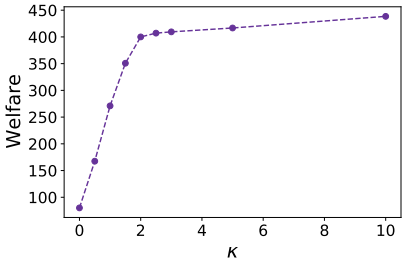


Figure 9: Welfare of approximate CCE for varying κ. For each κ, we plot the mean welfare computed over 100 runs of no-regret dynamics (std deviation bars are too small to appear on plot).

slower rate for large κ. Previously, we observed that for large κ, no-regret dynamics returns equilibria exhibiting high correlation (i.e. high TV distances). We can thus interpret this as preliminary evidence showing that allowing for correlation might improve social welfare in markets where permanent impact dominates. Indeed, we can think of correlation as a form of *collusion*. Our findings suggest that as κ increases—intuitively, this is also as we move farther from a zero-sum game—the potential benefits of collusion increase. Moreover, this type of collusive behavior can emerge organically from simple learning dynamics.

F Limitations and Future Work

Our work leaves open several avenues for future work. In this work, we considered static strategies. Although we can view our static strategies as baselines or planned trajectories for more sophisticated strategies, future research could explore competitive trading with dynamic strategies that adapt to market activity. Other interesting directions include more general cost functions, such as non-linear functions and those modeling decaying permanent impact, and incorporating different types of (possibly unknown) players with objectives beyond pure position acquisition. Finally, in our game, the parameter κ governs how close the game is to a potential game and a zero-sum game. Given that every game can be decomposed into a potential and a zero-sum game, it is broadly relevant to investigate how equilibria properties and computation evolve with shifting game structure, independent of our specific trading context.

G Acknowledgements

We give warm thanks to Neil Chriss, Yuriy Nevmyvaka, Andrew Bennett and Anderson Schneider for helpful discussions.

A Correlated Equilibria Computation

In Section D, we gave an efficient algorithm to compute approximate coarse correlated equilibria via no-regret dynamics. Here, we investigate efficient computation of correlated equilibria (CE), a stronger equilibrium concept than CCE.

Just as coarse correlated equilibrium is tied to the notion of regret (otherwise referred to as *external regret*), correlated equilibrium is tied to the notion of *swap regret*.

Definition A.1 (Swap Regret)**.** *Let $\Phi_i = \{\phi : \mathcal{A}_i \to \mathcal{A}_i\}$ be the collection of all functions mapping actions to actions. The (average) swap regret of a player i who chooses a sequence of actions $a_{i,1}, ..., a_{i,R} \in \mathcal{A}_i$ is defined as:*

$$SReg_i(R) = \max_{\phi \in \Phi_i} \frac{1}{R} \sum_{r=1}^{R} (c(a_{i,r}, a_{-i,r}) - c(\phi(a_{i,r}), a_{-i,r}))$$

where $a_{-i,r}$ is the profile of strategies chosen by players excluding i at round r.

No swap regret is a stronger guarantee than no (external) regret; it asks that given the history of play, a player has no incentive to deviate to a fixed strategy *conditioned on the strategy they chose*. By definition, if every player at swap regret at most ε, then the empirical joint distribution of play is an ε-approximate correlated equilibrium.

To implement no-*swap*-regret dynamics, we use an existing reduction transforming any no-regret algorithm to a no-swap-regret algorithm. The classical reduction of Blum and Mansour [3] guarantees, against any sequence of opponent actions, $SReg(R) \leq |\mathcal{A}| \cdot Reg(R)$ given an algorithm that obtains regret bounded by $Reg(R)$ (for certain no-regret algorithms, a tighter analysis improves the dependence on $|\mathcal{A}|$ by a factor of $\sqrt{|\mathcal{A}|}$). To handle a large action space, we use recent reductions of Dagan et al. [11] and Peng and Rubinstein [21] that guarantee vanishing swap regret at a rate depending only on the external regret guarantee of the no-regret algorithm, avoiding any dependence on the number of actions (however this comes at an exponential cost in the approximation parameter). We state its guarantees below but defer details of the reduction to Dagan et al. [11] and Peng and Rubinstein [21].

Theorem A.2 (Theorem 3.1 of Dagan et al. [11])**.** *Fix an action set $\mathcal{A}$. Fix $M, d, R \in \mathbb{N}$ such that $M^{d-1} \leq R \leq M^d$. Given an algorithm that, against any adversarial sequence of actions, guarantees regret at most $Reg(R)$ after R rounds, there is an algorithm producing randomized actions $p_1, ..., p_R \in \Delta\mathcal{A}$ such that (in expectation over the randomized actions):*

$$SReg(R) \leq Reg(M) + \frac{3}{d}$$

Moreover, if the per-round running time of the no-regret algorithm is C, then the per-round amortized running time of this algorithm is $O(C)$.

Importing our instantiation of FTPL, we obtain the following guarantee on the joint history given by no-swap-regret dynamics. The number of rounds needed to reach an ε-approximate correlated equilibrium is polynomial in the parameters of the game, but depends exponentially on $1/\varepsilon$.

Corollary A.3. *Fix a player i in the trading game with action set $\mathcal{A}_i$. Using the instantiation of FTPL given by Corollary D.3, there is an algorithm producing randomized actions $p_1, ..., p_R \in \Delta\mathcal{A}_i$ such that, in expectation over the randomized actions, $SReg_i(R) \leq \varepsilon$ after $R = O\left(\left(\frac{n^2\theta^5T^6}{\varepsilon^2}\right)^{\frac{1}{\varepsilon}}\right)$ rounds. Moreover, the per-round amortized running time is $O(\theta^2T^2)$, where $\theta = \max\{|\theta_L|, |\theta_U|\}$. If every player i runs a copy of this algorithm, the empirical distribution $\mathbf{D}$ is an ε-approximate correlated equilibrium after $R = O\left(\left(\frac{n^2\theta^5T^6}{\varepsilon^2}\right)^{\frac{1}{\varepsilon}}\right)$ rounds of no-swap-regret dynamics.*

Proof. We plug the regret guarantee of FTPL from Corollary D.3, $M = \frac{n^2\theta^5T^6}{\varepsilon^2}$, and $d = \frac{1}{\varepsilon}$ into Theorem A.2. The runtime complexity follows from Corollary D.4 and Theorem A.2. □

A.1 Details of Experimental Implementation

We implement the no-swap-regret algorithm of Corollary A.3 and plot the distance to CE (i.e. swap regret) of the joint play in Figure 6. In our implementation, we use the same game parameters as before: $T = 5, V_1 = 10, V_2 = 10, \theta_L = -5, \theta_U = 5$. Given that $M \gg d$ in theory, we set $M = 150$ and $d = 2$. For each setting of κ, we execute 20 runs of no-swap-regret-dynamics, each consisting of $M^d = 22,500$ rounds.

The results suggest that in practice, the algorithm is indeed hindered by the (exponentially) slow convergence rate given by theory. In fact, we find that FTPL obtains *lower* swap regret than this algorithm (it only guarantees low external regret), even with *much* fewer rounds.