THE UNIVERSITY OF CHICAGO

ACTIVE LEARNING IN MARKETPLACES AND ONLINE PLATFORMS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
YIFAN FENG

CHICAGO, ILLINOIS
JUNE 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Professor René Caldentey, for his continuous support, insight, and feedback throughout my academic journey as a Ph.D. student. He has encouraged me and guided me to develop as an independent researcher since the first day I met him at the University of Chicago. In the meanwhile, I can always gain new perspectives from his sharp comments on my research developments, which further inspires me to forge ahead. I cherish the nurturing environment he has created for me, where I can approach research problems from the first principles. Such experience in graduate school is a once-in-a-life opportunity for me. His strong work ethic has deeply influenced my studies and will inspire me in the years to come.

I would also like to thank my other dissertation committee members: Professors John R. Birge, N. Bora Keskin, Christopher T. Ryan, and Linwei Xin. I have learned a lot from them on how to grow into a scholar in this field. I am also grateful for all the generous resources and firm supports they have provided during my studies and my academic job market. Besides my committee members, I am thankful to my other research collaborators, namely, Professor Yuan Zhong, Adam Schultz, Lijun Zhu, Yinghui Xu, and Bing Wang. Although some research projects I have worked on with them are not included in this dissertation, I gained a unique perspective on research from each of them.

Beyond my collaborators, my great gratitude goes to people who have provided valuable feedback for my dissertation during various seminars and conferences. A quite incomplete list includes Professors Varun Gupta, Dan Adelman, Dennis J. Zhang, Will Ma, Renyu Zhang, and Barış Ata for Chapter 2, as well as Professors Kimon Drakopoulos, Varun Gupta and Amy R. Ward for Chapter 3.

My life in graduate school would not have been so fruitful without the cultivation of the University of Chicago Booth School of Business. Firstly, I would love to thank all of the faculties in the Management Science/Operations Management area beside my committee: Professors Dan Adelman, Barış Ata, Ozan Candogan, Levi DeValve, Donald D. Eisenstein,

Varun Gupta, R. Kipp Martin, Burhaneddin Sandikçi, Amy R. Ward and Yuan Zhong. Special thanks go to Professors Barış Ata, Levi DeValve, Varun Gupta, and Amy R. Ward for their valuable advice on graduate studies, academic job markets, among others.

I would also like to thank the administrative team at Booth: Brown Malaina, Amity James and Kimberly Mayer for their professional support. A special thank you goes to the EconScribe program, which totally changed my view about academic writing. Of course, I am also grateful for the friends I make at the university, who have been a constant source for feedback and encouragement, particularly from Hongfan (Kevin) Chen, Nasser Barjesteh, Ali Cem Randa, among others.

I will never forget how my academic journey started. I feel indebted to Professor Zizhuo Wang, who gave me hands-on guidance on conducting research in OM/OR when I was an undergraduate student. Professor Yichuan Ding, Professor Renyu Zhang, and Chenhao Du were the people from whom I first learned about the field of OM/OR. Let me also take this opportunity to thank Professors Zhenyue Zhang and Qifan Yang, from whom I took the introductory real analysis and mathematical modeling courses, respectively, two courses that have changed my life.

Last but not least, I would like to thank my family, which always reminds me of the meaning of life besides research. My parents have constantly been supporting me and giving me endless love. My final and most important thank-you goes to my dear wife and son. My wife is my best friend and great companion, who always gives me confidence and the ultimate emotional support. My son is my greatest gift in life. He has made me stronger and more fulfilled than I could have ever imagined.

ABSTRACT

Decision makers rely on observations to make better decisions. Hence mastering the interplay between data and decision-making is a central topic to the field of OM/OR. In this dissertation, we study active learning problems (defined broadly) in the context of managing marketplaces and online platforms. Here active learning means that the data flow that the decision maker (DM) observes could be *proactively* and *endogenously* determined by the actions of the DM and possibly other agents in the environment as well.

Classic dynamic learning problems typically involve resolving the tension between exploration (i.e., choosing informative actions to reduce model uncertainty) and exploitation (i.e., determining "reward maximizing" actions based on the estimated model). In Chapter 2, we consider a *pure exploration* problem instead, where the DM cares less about the reward flow along the way. Instead, the DM strives to take "information maximizing" actions to make high-confidence statistical inferences based on a minimal amount of data. Such pure exploration problems are very relevant in online platform operations, especially applied to survey/questionnaire design for preference learning, new product introduction, among others. Specifically, Chapter 2 studies how to pick the menu of products shown to each consumer so that the platform only needs a minimal amount of samples to identify the consumer population's favorite product with high confidence.

Chapter 3 deviates from the classic dynamic learning literature in a different dimension: it examines an environment where the data flow could be *strategically manipulated* by powerful agents. Specifically, in a spread betting market, the market maker (i.e., DM) wishes to learn from market transactions to move her spread lines in a way to correct mispricing. In the meanwhile, she may face an informed bettor who can profit from "flooding" the market (in the opposite direction) to exacerbate the market maker's mispricing. The goal of the DM in Chapter 3 is to design a learning algorithm that not only gathers information from the market, but also protects the DM from strategic manipulations.

CHAPTER 1

INTRODUCTION

Two concrete problems are studied in this dissertation.

Chapter 2. This paper studies a class of ranking and selection problems faced by a company that wants to identify the most preferred product out of a finite set of alternatives when consumer preferences are *a priori* unknown. The only information available is that consumer preferences satisfy two key properties: (*i*) they are consistent with some unknown true ranking of the alternatives and (*ii*) they are strict, namely, no two products are equally preferred. To learn the unknown ranking, the company is able to sample consumer preferences by sequentially showing different subsets of products to different consumers and asking them to report their top preference within the displayed set. The objective of the company is to design a display policy that minimizes the expected number of samples needed to identify the top-ranked product with high probability. We prove an instance-specific lower bound on the sample complexity of any policy that identifies the top-ranked version within a given (probabilistic) confidence. We also propose a robust formulation of the company's problem and derive a sampling policy (Myopic Tracking Policy), which is both worst-case asymptotically optimal and intuitive to implement. Roughly speaking, the Myopic Tracking Policy randomly alternates between two extreme type of displaying strategies: (i) *full display* that shows a consumer the entire menu so as to learn something about every version and (ii) *pair display* that shows a consumer only two versions so as to maximize the informativeness of the choice made by the consumer. To assess the performance of our proposed Myopic Tracking Policy, we conduct a comprehensive set of computational studies and compare it to alternative methods in the literature.

Chapter 3. We study the profit maximization problem of a market maker in a spread betting market. In this market, the market maker quotes cutoff lines for the outcome of a certain future event as "prices," and bettors bet on whether the event outcome exceeds the cutoff lines. Anonymous bettors with heterogeneous strategic behavior and information

levels participate in the market. The market maker has limited information on the event outcome distribution, aiming to extract information from the market (i.e., "learning") while guarding against an informed bettor's strategic manipulation (i.e., "bluff-proofing").

We show that Bayesian policies that ignore bluffing are typically vulnerable to the informed bettor's strategic manipulation, resulting in exceedingly large profit losses for the market maker as well as market inefficiency. We develop and analyze a novel family of policies, called *inertial policies*, that balance the tradeoff between learning and bluff-proofing. We construct a simple instance of this family which (i) enables the market maker to achieve a near-optimal profit loss and (ii) eventually yields market efficiency.

1.1 Related Papers

The material presented in this dissertation is based on the following papers.

Chapter 2. Y. Feng, R. Caldentey, and C. T. Ryan. *Robust learning of consumer preferences*. Available at `https://ssrn.com/abstract=3215614`.

Chapter 3. J. R. Birge, Y. Feng, N. B. Keskin and A. Schultz. *Dynamic learning and market making in spread betting markets with informed bettors*. Available at `https://ssrn.com/abstract=3283392`.

CHAPTER 2

ROBUST LEARNING OF CONSUMER PREFERENCES

2.1 Introduction

Problem Overview. A company wants to identify the 'best' version of a product to commercialize in the marketplace from a menu $[K] = \{1, 2, \ldots, K\}$ of alternative versions. The company does not know consumer preferences over these versions and implements a consumer feedback system to collect information. Specifically, the platform is able to display different subset of versions to different consumers, who then give feedback in the form of a *vote* for their most preferred version within the subset they see. In addition, the system must decide when to stop the feedback process and make a recommendation on which version to commercialize.

There are many possible feedback mechanisms that a company can use. Traditional examples include "taste tests", focus groups, or surveys of potential consumers. With the advent of the Internet, methods of feedback have become more sophisticated. One trend in online commerce is the use of *crowdvoting* platforms to collect consumer feedback about possible new products or other business innovations. For example, Chicago-based fashion company Threadless uses a crowdvoting platform to feature T-shirt designs from freelance designers on a weekly basis and solicits consumer opinions online for preferred designs. Threadless uses this consumer feedback to narrow down the number of designs that are sent to production (Brabham, 2010, King and Lakhani, 2013).[1]

The task of efficiently managing the feedback platform –*i.e.*, balancing the quality and speed of the learning process– is complex, particularly when (i) the number of alternative versions is large and (ii) making inferences about consumer preferences from votes is limited.

1. Other examples include: (i) Amazon, which leverages reader nominations to e-publish books (the *Kindle Scount* platform, Pee, 2016); (ii) LEGO, which uses crowdvoting to generate and pick new designs of toy sets (the *LEGO Ideas* platform, LEGO, 2018); and (iii) Betabrand, which uses both crowdfunding and crowdvoting to solicit design ideas and converts selected designs into real products (Betabrand, 2018).

In such cases, the company needs to judiciously and dynamically choose which subset of versions to display to each arriving consumer, with the objective of maximizing the *amount of information* generated by each consumer choice. The optimal choice of display sets is contingent on the history of votes observed over time. In terms of the length of the feedback process, the company would like to minimize the time required to make a final recommendation, but without jumping too quickly to recommendation. Commercializing the '*wrong*' version can be costly (Schneider and Hall, 2011).

In this paper, we propose a methodology to (i) dynamically choose which *display set* to show each arriving consumer, (ii) decide when to stop the feedback process, and (iii) select the version to commercialize. This methodology minimizes the amount of feedback needed to achieve a fixed probabilistic guarantee of choosing the best version. Our methodology builds on a general class of ranking-based choice models where consumer choice probabilities are determined by a fixed (unknown to the company) ranking over the set of versions that represent consumer preferences. We provide a detailed mathematical description of this choice model in Section 2.4.

To get some intuition for the trade-offs involved, and how our methodology balances them, let us discuss two extreme display strategies. On one extreme, the company can use a *full-display* policy where all versions are displayed to every consumer. This allows the company to learn something about each consumer's preference over every version. In this regard, a full-display policy maximizes the *coverage* achieved by a display policy. However, under reasonable assumptions on the underlying choice model, the probability that a consumer votes for the best version within a given display set decreases in the cardinality of the set. For instance, consumers can become overwhelmed if many alternatives are displayed, making it harder for them to identify their true most-preferred version, (e.g., the paradox of choice). Hence, larger display sets may provide less accurate information than smaller ones. To maximize the *accuracy* of the inference made on each consumer choice, the company can use a *pair-display* policy where only two versions are shown to each consumer. Of course,

the choice of which pair of versions to display to each consumer should depend on the history of the feedback process. This choice makes implementing an optimal pair-display strategy substantially more complex than a full-display strategy, which displays the same set of versions to all consumers.

In general, neither the full-display nor the pair-display strategies are optimal. This is as expected, the one optimizes for coverage (at the cost of accuracy) while the other optimizes for accuracy (at the cost of coverage). An optimal strategy must strikes the right balance between coverage and accuracy. The goal in this paper is to shed some light on this trade-off.

Summary of Methodology and Results. The development of rigorous methods to learn consumer preferences has been the focus of much research in computer science, economics, marketing, and operations (see Section 2.2 for a review of related literature). A dominant approach to tackle this problem is to impose a parametric structure on the underlying choice model governing consumer behavior. For instance, Luce-type models –and in particular the Multinomial Choice Model (MNL)– are regularly used (e.g., Sauré and Zeevi, 2013a, Chen et al., 2018). This parametric approach, however, puts us in an uncomfortable predicament. In order to learn unknown consumer preferences, we must assume that we have a fair amount of knowledge about the parametric structure of those preferences.[2]

With the aforementioned predicament in mind, we develop an efficient active learning algorithm to learn these preferences '*on the fly*' while imposing minimal parametric assumptions. We tackle this challenge using a worst-case formulation of the problem that relies on a mild separability condition. This worst-case analysis singles out a specific choice model we call the *Ordinal Attraction* (OA) model. The OA model is, in some sense, the "noisiest" (and hence, the hardest) consumer choice model to learn among those satisfying our separability condition. Fortunately, as in robust optimization, where the worst-case distribution often

2. See Heckel et al. (2019) for a discussion on the limited benefits of parametric models in the context of rankings from pairwise comparisons.

has a tractable structure, the OA model has many attractive analytical features that we exploit to propose a robust and computationally efficient display policy.

Our proposed strategy – the Myopic Tracking Policy in Algorithm 1– judiciously balances the coverage/accuracy trade-off and satisfies two key optimality properties: (i) it chooses the best version *with high probability* and (ii) it *asymptotically* minimizes the amount of consumer feedback needed to make a final recommendation. To be precise, for any small $\delta > 0$, we first derive a lower bound on the number of votes needed by any display strategy to achieve a $1 - \delta$ probability of selecting the top-ranked version (Theorem 1). We then show that as $\delta \downarrow 0$, the Myopic Tracking policy needs an expected number of votes that matches this lower bound (Theorem 6) with up to δ error probability (Theorem 7).

To get a panoramic view of our approach and the challenges that we need to address, we summarize its main characteristics. Loosely speaking, at every consumer arrival, the Myopic Tracking policy goes through the following sequence of steps:

Myopic Tracking policy: Schematic Description

Step 1: Using the available history of consumer votes, identify the 'best' preference ranking over all versions.

Step 2: Using the identified preference in Step 1, and the available data, update a stopping criteria. If the criterion is met, stop the feedback process and select the top-ranked version according to the current ranking. Otherwise, go to Step 3.

Step 3: Randomly (according to a pre-specified distribution) select a display set to show the next consumer. Record the vote outcome and go to Step 1.

In Step 1, the algorithm computes a ranking of the versions that best reflects consumer preferences given the feedback history. There are two issues that need to be handled in this step. First, we need to decide how to identify the 'best ranking' in every iteration. Second,

we need to be able to compute this ranking efficiently. Our proposed methodology relies on a Maximum Likelihood Estimation (MLE) criteria to select the best ranking at every iteration.

The stopping criterion in Step 2 is of a threshold-hitting type *à la* Chernoff (1959). Specifically, the Myopic Tracking policy keeps track of a stochastic process that measures the discrepancy between the preference ranking identified in Step 1 and the available data. The algorithm stops as soon as this stochastic process hits a fixed lower bound whose value is appropriately calibrated to ensure that the algorithm will select the best version with probability at least $1 - \delta$. The evolution of the underlying stochastic process is related to the MLE computations used in Step 1.

Turning to Step 3, our proposed policy randomly selects the display set shown to each arriving consumer. In particular, the probability distribution that is used to randomize over the display sets is constant –invariant to the feedback history– up to the permutation of the versions induced by the preference ranking identified in Step 1. It follows that this randomization distribution can be computed offline before the actual implementation of the feedback system. Furthermore, under the OA model, these randomization probabilities depend on the relative ranking of each version within the given display set. Through a closed-form characterization (Theorem 4), we find that the Myopic Tracking Policy restricts this randomization to a small subset of nested display sets, i.e., those including the top n most preferred versions according to the preference ranking identified in Step 1.[3] This means that there are only $K - 1$ nested sets that are being considered for display, a much smaller number than the $2^K - K - 1$ possible display subsets of $[K]$.[4] This key property of the Myopic Tracking policy dramatically reduces the complexity of its implementation.

3. Specifically, let $\sigma : [K] \to [K]$ be the *preference ranking* (*i.e.*, a permutation of the elements in $[K]$) identified in Step 1. The collection $\{S_n^\sigma\}_{n=2}^K$ of nested display sets associated to ranking σ is such that $S_n^\sigma = \{\sigma^{-1}(1), \sigma^{-1}(2), \ldots, \sigma^{-1}(n)\}$.

4. The possible display sets are all the subsets of $[K]$ excluding the empty set and singletons, which are obviously never optimal to display.

Finally, although we derive the OA model as a worst-case consumer preference, for the purpose of proposing a robust solution to the top-ranked selection problem, the OA model has a number of alternative interpretations. For instance, it can be viewed as a generalization of the pair-wise comparison model commonly used in the tournament literature. In particular, the MLE problem in Step 1 in our Myopic Tracking Policy is equivalent to the classical dispersion minimization criterion in the sense of Young (1988)(Proposition 2). Thus, computationally, the MLE problem and stopping criterion verification problem can be cast as versions of the *weighted feedback arc set problem on tournaments* that admits an effective integer linear programming algorithm (see Section 2.7.2). The OA model is also connected to voting theory and, specifically, the Condorcet criterion (see Section 2.7.3).

2.2 Related Literature

Methodology: In terms of methodology, our paper builds on the following three areas of research:

1. Sequential Hypothesis Testing: At a high level, we interpret our problem as an active, sequential, and composite multi-hypothesis testing problem where each hypothesis corresponds to one version being the top-ranked version. When the experimenter is a passive observer of data, the generalized sequential likelihood ratio test is known to be asymptotically optimal under various settings (e.g. Wald, 1973, Chernoff, 1972, Draglia et al., 1999, Li et al., 2014). This provides some support for Steps 1 and 2 in the Myopic Tracking Policy, which essentially implement a generalized sequential likelihood ratio test. On the other hand, the sampling rule in Step 3 is motivated by classical results in active hypothesis testing, in particular, the Max-Min problem studied in Chernoff (1959) (see also Naghshvar et al., 2013).

   There are, however, several key distinctions in our model that prevent us from directly applying the results in Chernoff (1959). First, Chernoff (1959) considers a different

objective criterion that incorporates a penalty term for selecting the wrong hypothesis (in our case the wrong version), while our problem has an explicit hard constraint that bounds the error probability upon stopping. Second, Chernoff (1959) analysis is restricted to settings in which each hypothesis consists of a finite number of possible states while we allow for an infinite number of states. Moreover, there are certain separability conditions that are imposed among the alternative hypotheses in Chernoff (1959), which our model relaxes. Lastly, our worst-case analysis allows us to completely solve the corresponding Max-Min problem, which is intractable if the experimenter uses the naïve LP formulation of Chernoff's Max-Min problem.

2. Ranking and Selection: Our paper also contributes to the emerging literature on ranking and selection from pairwise to multiwise noisy comparisons (e.g., Braverman and Mossel, 2008, Ailon, 2012, Braverman and Mossel, 2009, Jiang et al., 2011, Wauthier et al., 2013, Shah and Wainwright, 2017, Falahatgar et al., 2017, Heckel et al. (2019), among others).

   Among the very few papers that also consider a multi-wise comparison setting, the closest paper to ours is probably Chen et al. (2018) that studies a top-k selection problem under a Multinomial Logit (MNL) model. The class of choice models we consider is, in general, different than the one used in Chen et al. (2018), although some results are comparable. For instance, our methodology improves the lower bound on sample complexity in Chen et al. (2018) (Theorem 1.3), from a fixed success rate, to a generic error rate of $\delta \in (0, 1)$ (Theorem 1). Our lower bound is also asymptotically tight because we can find an admissible policy to match the lower bound when δ is small.

   We also approach the selection problem from a different angle and so the optimality regimes in Chen et al. (2018) and our paper are different. Chen et al. (2018)'s algorithm (nearly) matches their lower bound when (i) fixing the success probability and (ii)

ignoring poly-logarithmic factors of K and the reciprocal of MNL parameter gaps (Theorems 1.2 and 1.3). By comparison, our proposed algorithm (Theorem 6) (i) is asymptotically optimal with respect to the error rate δ, even with the coefficient term matched and (ii) allows parameters other than δ to grow large, as long as they grow not too fast compared to $1/\delta$.

3. Best Arm Identification: Our solution method is also related to the best arm identification (BAI) literature (e.g., Audibert and Bubeck, 2010, Bubeck et al., 2011, Gabillon et al., 2012). In a generic BAI problem, the experimenter tries to distinguish the best arm (i.e., the one with highest expected reward) using as few samples as possible, where a sample corresponds to pulling an individual arm and observing a realization of that arm's reward distribution. Our problem could be vaguely cast as a BAI problem by treating each version as an arm and each display set as a "super-arm", i.e., a subset of arms. With this interpretation, the company decision is to select which super-arm to pull at every time epoch. Two key distinctions between BAI and our problem are (i) by pulling a super-arm, the company is able to learn something about every arm included in the super-arm and (ii) after pulling a super-arm, the observed response is a particular arm rather than a realization of the arm's reward distribution. There is a growing awareness of the relationship between best arm identification and active sequential hypothesis testing (Russo, 2016, Garivier and Kaufmann, 2016, Kaufmann et al., 2016), especially regarding the importance of the Max-Min problem proposed in Chernoff (1959).

Applications: In terms of applications, our paper is related to two streams of work:

1. Crowdvoting/Wisdom of Crowd: Our paper brings an optimal learning view to crowdvoting or more generally, leveraging the "wisdom of the crowd" to help with product offering decisions (e.g. Raykar et al. (2010), Marinesi and Girotra (2012), Huang et al. (2014), Araman and Caldentey (2016), to name a few). For example, Marinesi and

Girotra (2012) study a two-period model where the company uses an online voting platform as an information acquisition mechanism. Araman and Caldentey (2016) decide on the optimal length of the voting period, so as to balance quality of learning and delay cost. Our paper differs from the previous literature, in the sense that we consider consumer choice behavior among many versions, and focus on how to customize each consumer's choice set to maximize the learning speed.

2. Dynamic Assortment Planning with Learning. Our paper is also related to the growing literature on dynamic assortment planning with demand learning (e.g. Caro and Gallien, 2007, Rusmevichientong et al., 2010, Ulu et al., 2012, Sauré and Zeevi, 2013b, Agrawal et al., 2019, Agrawal et al., 2017, Chen and Wang, 2017, among others). The vast majority of this literature formulates the assortment problem as a revenue maximization (or regret minimization) problem and relies on a "learn and earn" approach to solve it (e.g. Rusmevichientong et al., 2010 and Sauré and Zeevi, 2013b). A popular strategy is to divide the selling season into two periods: (1) a "pure learning" period in which assortments are offered sequentially to maximize the amount of learning without any revenue consideration, and (2) a "pure earning" period in which a myopic static strategy (based on the knowledge obtained in the pure learning period) is implemented to maximize revenues. As a general rule, the assortments used during the pure learning period have maximal cardinality, typically determined by an exogenously-imposed capacity constraint.

   By contrast, our model is solely concerned with maximizing the likelihood of selecting the best version and thus resembles the "pure learning" period previously mentioned. A key insight that emerges from our work is that the exclusive use of maximal cardinality assortments is, in general, suboptimal. The learning process can be accelerated by judiciously balancing the sizes of the display sets over time.

2.3 Roadmap of Analysis and Results

In this section we provide a high-level outline of our analysis and main results with the objective of explaining our methodology in simple and intuitive terms. Formal definitions and precise mathematical statements are presented in the sections that follow.

A company wants to identify the product (or version) that is the most attractive to a given population of individuals out of a set of $K \geq 2$ available alternatives. We assume that the preferences of these individuals (consumers) over the different versions are governed by a probabilistic choice model that satisfies two key properties. First, the choice model belongs to the class of ranking-based preferences, namely, there is an underlying (unknown to the company) ranking of the versions and the likelihood that a consumer would select one specific version out of a given subset is consistent (in a probabilistic sense) with this ranking. Second, we will assume that the preferences satisfy a separability condition under which no two versions are equally preferred. We will denote by $\mathcal{M}_p$ the class of choice models that satisfy these requirements, by $f \in \mathcal{M}_p$ a specific consumer preference, and by $f_* \in \mathcal{M}_p$ the true unknown consensus preference of the consumers.

Because preferences are ranking based, for each $f \in \mathcal{M}_p$ there exists a unique permutation of the K versions that is consistent with f. We let Σ denote the set of permutations (or rankings) of the set of versions $[K] := \{1, 2, \ldots, K\}$ and $\sigma_f \in \Sigma$ denote the ranking associated with a preference $f \in \mathcal{M}_p$. It follows that the set of preferences $\mathcal{M}_p$ can be partitioned into a collection $\{\mathcal{M}_p(k) \colon k \in [K]\}$, where each member $\mathcal{M}_p(k)$ of the partition is the set of preferences f for which $\sigma_f(k) = 1$, that is, k is the top-ranked version under σ_f. Thus, the company's problem can be cast as the multi-hypothesis testing problem of deciding which set $\mathcal{M}_p(k)$ contains the true preference f_*. It is worth highlighting that the company is not directly interested in identifying f_*, simply the top-ranked product under it.

To tackle this problem, the company sets a sequential experimentation (or voting) strategy under which individuals from the target population are exposed –one by one– to a subset

of the versions (a display set) and are asked to select the version they most like. The display set can vary from individual to individual and the goal of the company is to design a strategy that will identify, as quickly as possible, the true hypothesis with a given probabilistic confidence. Specifically, for a given error tolerance $\delta \in (0,1)$, the company wants to design a display policy that is δ-*accurate*; that is, chooses the true hypothesis with probability at least $1-\delta$.

Our first result (Theorem 1) shows every δ-accurate policy needs at least $\log(1/\delta)/I_*(f_*)$ samples in expectation to learn the true hypothesis, where $I_*(f_*)$ is *Chernoff's information measure* that determines the informativeness of preference f_*. Our second result (Theorem 2) shows –under an additional constraint on the cardinality of $\mathcal{M}_p$– that there exists a δ-accurate policy that asymptotically, as $\delta \downarrow 0$, uses no more than $\big(\log(1/\delta) + o(\log(1/\delta))\big)/I_*(f_*)$ samples in expectation to learn the true hypothesis. Combined, these two results formalize the intuitive fact that some preferences are easier (or harder) to learn than others and show that Chernoff's information measure $I_*(f)$ quantifies the learning complexity of a given f.

Motivated by the lower and upper bounds on the sample complexity of any δ-accurate policy identified by Theorems 1 and 2, we adopt a worst-case view of the problem and take on the challenge of identifying the set of preferences that minimize Chernoff's information measure $I_*(f)$ over all the preferences in the set $\mathcal{M}_p$, i.e., the preferences that are the hardest to learn. In Theorem 3, we characterize of a subset of this class of worst-case preferences that we call the *Ordinal Attraction* (OA) choice model. The OA model has a number of distinguishing properties that we discuss in detail in Section 2.6. One property worth mentioning here is that, under the OA model, the probability that a given version k is selected within an arbitrary display set S depends exclusively on the relative ranking of k with respect to the other versions in S. Hence, under the OA choice model, consumer preferences have an ordinal structure. It is this property that motivates the name Ordinal Attraction.

We use the worst-case nature of the OA model to formulate a robust version of the company's problem. In particular, we introduce the subset $\mathcal{M}_p^{\text{OA}} \subseteq \mathcal{M}_p$ of OA preferences and look for a δ-accurate policy with minimum expected sample complexity in $\mathcal{M}_p^{\text{OA}}$. To this end, we propose the *Myopic Tracking Policy* (MTP) and show, in Theorem 5, that it is worst-case asymptotically optimal as $\delta \downarrow 0$ within the class of δ-accurate policies not just in $\mathcal{M}_p^{\text{OA}}$ but in the larger set of preferences $\mathcal{M}_p$. Roughly speaking, this means that for any δ-accurate policy π there exists a preference $f_\pi \in \mathcal{M}_p$ where the expected number of samples needed to identify the top-ranked version under f_π using the MTP policy is less than or equal to the expected number of samples needed by policy π. The proof of Theorem 5 is based on Theorems 6 and 7 that show the MTP policy (i) matches the lower bound on the samples complexity in Theorem 1 asymptotically as $\delta \downarrow 0$ and (ii) is δ-accurate. We also show that the Myopic Tracking Policy relies on a simple randomization strategy over a relatively small subset of all possible display sets. Finally, the stopping criteria of the Myopic Tracking Policy is also simple and takes the form of a first hitting time of an appropriate log-likelihood process, very much in the same spirit as Wald's SPRT method.

In Section 8, we use a set of computational experiments to test the performance, in terms of accuracy and sample complexity, of our proposed Myopic Tracking Policy. Using synthetic data, we find that the Myopic Tracking Policy is particularly well suited for an environment where: (i) the number versions is large; (ii) responses are noisy; and (iii) the tolerance for error probability is low.

2.4 Model and Problem Formulation

Consumer preferences over the set $[K]$ of available versions are represented by a *consumer choice model* that defines the probability $f(X|S)$ that a consumer will select version $X \in [K]$ when presented with display set $S \subseteq [K]$. The set $\mathcal{S} := \{S \subseteq [K], |S| \geq 2\}$ denotes the collection of all display sets with at least two versions. We refer to f as a consumer preference

or simply a *preference*.

We restrict attention to the class of ranking-based preferences that satisfy a specific separability condition. Let Σ denote the set of all permutations of the elements in $[K]$, an element $\sigma \in \Sigma$ is called a *ranking*.

Definition 1. (p-Separable Preferences) *Let $p \in [0,1)$ be a fixed constant. A preference f belongs to the class $\underset{\surd}{\mathcal{M}}$ of ranking-based p-Separable preferences if:*

(A-1) `Non-degeneracy:` *For any $S \in \mathcal{S}$, $f(X|S) > 0$ if $X \in S$ and $f(X|S) = 0$ otherwise;*

(A-2) `Probability Mass Function:` *For any $S \in \mathcal{S}$, $\sum_{X \in S} f(X|S) = 1$;*

(A-3) `Ranking-based Preference:` *There exists a ranking $\sigma_f \in \Sigma$ such that for any $S \in \mathcal{S}$ and any $X, X' \in S$, $f(X'|S) \leq f(X|S)$ if and only if $\sigma_f(X) < \sigma_f(X')$;*

(A-4) `(`p`-Separability)` *For any $S \in \mathcal{S}$ and any $X, X' \in S$ such that $\sigma_f(X) < \sigma_f(X')$, the preference f satisfies $f(X'|S) \leq p\, f(X|S)$.* $\diamond$

Conditions (A-1) and (A-2) are rather intuitive requirements to impose on any probabilistic choice model. Condition (A-3) imposes a minimum level of consistency on the consumers' preferences to ensure that the problem of identifying the the top-ranked problem is well defined. Specifically, under this condition, preferences are independent of the display set. Finally, Condition (A-4) is needed for technical reasons to ensure that we can effectively identify the top-ranked version as the number of votes grows large. The condition, however, is rather mild as it only requires that (i) versions with lower ranking are more likely to be chosen and (ii) consumers are not indifferent between two products in any display set. It is not hard to see that Luce-type models, such as the MNL, are p-Separable (for some value of p) as long as the attraction score of the different versions are all different. The parameter p measures the degree of informativeness of the choice model. For example, in the extreme case of $p = 0$, consumers select the best alternative among any display set with probability one. In this case, the identification problem is trivial and the company needs to display the entire

assortment to a single consumer to identify the best version. If $p = 1$, it is possible that the preference is completely uninformative since the uniform preference $f(X|S) = \mathbb{I}\{X \in S\}/|S|$ belongs to the class of 1-Separable choice models.

Remark 1. For any $p \in [0, 1)$ and any preference $f \in \mathcal{M}_p$, the ranking σ_f in Condition (A-3) is uniquely defined. Indeed, for every version $k \in [K]$, define $q_k = 1 - f(k|[K])$ and let $q_{(k)}$ be the k^{th} order statistic of the sequence $(q_1, q_2, \ldots, q_K)$. Then, $\sigma_f(k) = i$ if and only if $q_k = q_{(i)}$. $\diamond$

Turning to the company's problem, recall that $f_* \in \mathcal{M}_p$ is the *consensus preference*, which is the true (unknown to the company) consumer preferences over the versions in $[K]$. The objective of the company is to design a display strategy to identify $X^* = \sigma_{f_*}^{-1}(1)$, the most preferred version under f_*, as fast as possible. We do not undertake the more ambitious objective of identifying the top k versions (or even the complete ranking σ_{f_*}). Without loss of generality, assume that σ_{f_*} is the identity ranking $\sigma_* := (1, 2, \ldots, K)$.

Consumers arrive sequentially and are indexed by $t = 1, 2, 3, \ldots$ (we use index t to index both time and consumers, since only one consumer arrives per time period.) At time t, the company selects a subset $S_t \in \mathcal{S}$ and displays it to the arriving consumer. Consumer t chooses a version $X_t \in S_t$ and the company records his/her choice. We call X_t the "vote" of consumer t. The history of display sets and votes is captured by the filtration $\mathcal{F}_t$, the smallest sigma-algebra generated by $H_t = (S_1, X_1, \ldots, S_t, X_t)$. We also let $\Delta(\mathcal{S})$ denote the set of probability distributions on $\mathcal{S}$.

The Company's Decision Problem. An admissible policy has three parts:

1. a *display rule*, i.e., a sequence $\{\lambda_t\}_{t=1}^{\infty}$ of probability distributions $\lambda_t \in \Delta(\mathcal{S})$ adapted to the history $H_{t-1} := (S_1, X_1, \ldots, S_{t-1}, X_{t-1})$,

2. a *stopping rule*, i.e., an $\mathcal{F}_t$ stopping time τ for when the feedback process stops, and

3. a *final selection rule*, i.e., $d_\tau \in [K]$ that identifies which version to select in the end.

We let $\pi = (\{\lambda_t\}_{t=1}^{\infty}, \tau, d_\tau)$ denote an admissible policy. A preference $f \in \mathcal{M}_p$ and an admissible policy π induce a probability distribution $\mathbb{P}_f^\pi(\cdot)$ over the history $\{H_t\}$. We also denote by $\mathbb{E}_f^\pi[\cdot]$ the expectation operator under $\mathbb{P}_f^\pi(\cdot)$. With a slight abuse of notation, we may also suppress the superscript, and use the notations $\mathbb{P}_f(\cdot)$ and $\mathbb{E}_f[\cdot]$, when the context is clear.

The company's objective is to implement a policy π that identifies the top-ranked version as quickly as possible. The challenge in determining an optimal policy is the classical sequential learning trade-off between *confidence* (*i.e.*, how sure is the company about selecting the top-ranked version) and *speed* (*i.e.*, how fast can the company reach a final decision). To mathematically formalize this trade-off, we follow the best-arm identification literature (see, e.g., Gabillon et al., 2012) and use a *fixed confidence* approach in which the company's objective is to minimize the (expected) number of votes subject to a hard constraint that upper bounds the probability of selecting the wrong version. To this end, let us define the notion of a δ-accurate policy:

Definition 2. (δ-accurate policy) *Let $\mathcal{M} \subseteq \mathcal{M}_p$ be a subset of preferences. An admissible policy π is* $\delta(\mathcal{M})$-accurate *if*

$$\mathbb{P}_f^\pi(\tau < \infty) = 1 \qquad \textit{and} \qquad \mathbb{P}_f^\pi(d_\tau \neq \sigma_f^{-1}(1)) \leq \delta \qquad \textit{for any } f \in \mathcal{M};$$

that is, if the voting process terminates almost surely and the probability of selecting the wrong version (according to any given preference $f \in \mathcal{M}$) is less than or equal to δ.
An admissible policy π is δ-accurate *if it is* $\delta(\mathcal{M}_p)$-accurate. $\diamond$

In what follows, we tackle the problem of finding a policy π that minimizes the expected number of votes $\mathbb{E}_{f_*}^\pi[\tau]$ within the class of δ-accurate policies.

2.5 Lower and Upper Bounds on the Required Number of Votes

In this section, we derive upper and lower bounds for the sample complexity (i.e., required number of votes) of any δ-accurate policy. Our lower bound result is rather general, as we provide an instance-specific and non-asymptotic lower bound on $\mathbb{E}_f^\pi[\tau]$ for any $f \in \mathcal{M}_p$ and

any δ-accurate policy π. On the other hand, our upper bound is derived in an asymptotic regime under more restrictive conditions on the set of feasible preferences.

Let us start by introducing some notation. Given any display set $S \in \mathcal{S}$ and probability distribution $\lambda \in \Delta(\mathcal{S})$, the Kullback-Leibler (KL) divergence between two preferences f_1 and f_2 with respect to S and λ are given by

$$D_S\left(f_1||f_2\right) := \sum_{k\in S} f_1\left(k|S\right) \log \frac{f_1\left(k|S\right)}{f_2\left(k|S\right)} \qquad \text{and} \qquad D_\lambda\left(f_1||f_2\right) := \sum_{S\in\mathcal{S}} \lambda(S)\, D_S\left(f_1||f_2\right), \tag{2.1}$$

respectively. We define $\mathcal{M}_p(f) := \{f' \in \mathcal{M}_p \colon \sigma_f(1) = \sigma_{f'}(1)\}$ to be the set of preferences that have the same top-ranked version as f and its complement $\overline{\mathcal{M}}_p(f) := \mathcal{M}_p \setminus \mathcal{M}_{\surd}(\{)$, which is the set of preferences that have a top-ranked version different from f. We also introduce *Chernoff's information measure* $I_*(f)$ to be the value of the following Max-Min problem (parameterized by preference f):

$$I_*(f) := \sup_{\lambda\in\Delta(\mathcal{S})} \inf_{\bar{f}\in\overline{\mathcal{M}}_p(f)} D_\lambda\left(f||\bar{f}\right). \tag{Max-Min}$$

Here, $I_*(f)$ quantifies the inherent difficulty of the learning problem when the underlying preference is f. As we will see below, $I_*(f)$ is a measure of separability among alternative hypotheses. In particular, the larger the value of $I_*(f)$, the easier the top-ranked identification problem is under f. Chernoff's information measure $I_*(f)$ is closely related to the expected number of votes needed by any δ-accurate policy. We use this relationship to calibrate the design of an optimal policy (see Section 2.7).

In passing, we note that the max-min nature of $I_*(f)$ allows for a game-theoretic interpretation. Under this interpretation, the decision maker selects a randomized display strategy $\lambda \in \Delta(\mathcal{S})$ to maximize the KL divergence between a preference f and an alternative preference $\bar{f}$, which is being selected in an adversarial fashion from the set of preferences $\overline{\mathcal{M}}_p(f)$ that differ with f on the top-ranked product.

Given any $\delta_1, \delta_2 \in (0,1)$, let $kl(\delta_1,\delta_2) := \delta_1 \log \frac{\delta_1}{\delta_2} + (1-\delta_1)\log\frac{1-\delta_1}{1-\delta_2}$ denote the Kullback-

Leibler divergence between two Bernoulli distributions with means δ_1 and δ_2. Theorem 1 below identifies a lower bound on the number of votes needed by any δ-accurate policy.

Theorem 1. (Lower Bound on $\mathbb{E}^{\pi}_{\sigma}[\tau]$) *Let $\delta \in (0,1)$. For any δ-accurate policy π and $f \in \mathcal{M}_p$,*

$$\mathbb{E}^{\pi}_{f}[\tau] \geq \frac{kl(\delta, 1-\delta)}{I_*(f)} \qquad \textit{and} \qquad \liminf_{\delta \downarrow 0} \frac{\mathbb{E}^{\pi}_{f}[\tau]}{\log\left(\frac{1}{\delta}\right)} \geq \frac{1}{I_*(f)}.$$

Our proof of Theorem 1 in Appendix A.1 is based on Kaufmann et al. (2016) and Garivier and Kaufmann (2016). Specifically, we adapt the change of measure Lemmas 18 and 19 in Kaufmann et al. (2016) that they develop for the best arm identification setting to a more general hypothesis testing framework that we describe in Appendix A.1.1, which captures our setting for identifying the top-ranked alternative as a special case. In the process we also provide a slight generalization of the 'transportation' Lemma 1 in Kaufmann et al. (2016). This extra level of generality in the proof of Theorem 1 reveals that the lower bound above can be applied to a broader class of problems such as identifying the top-k versions or the complete ranking σ_{f_*}, to name a few. We illustrate this point in Section A.1.5 in the appendix, where we compare our lower bound to the one proposed by Jamieson et al. (2015) (see also Heckel et al., 2019) in the context of dueling bandits. In Proposition 8 we show that our bound is actually tighter than the one proposed by Jamieson et al. (2015) (Theorem 3 in their paper).

Let us now turn to the derivation of the upper bound. As we mentioned above, to derive this upper bound we need to impose some additional conditions. Specifically, we need to limit the cardinality of the set of feasible preferences.

Theorem 2. (Upper Bound on $\mathbb{E}^{\pi}_{\sigma}[\tau]$) *Let $\mathcal{M}^{\text{F}}_p$ be an arbitrary finite subset of $\mathcal{M}_p$ and $\delta \in (0,1)$. There exists a $\delta(\mathcal{M}^{\text{F}}_p)$-accurate policy $\hat{\pi}$ such that $\mathbb{E}_f[\tau] < \infty$ for every $\delta \in (0,1)$ and*

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}^{\hat{\pi}}_{f}[\tau]}{\log\left(\frac{1}{\delta}\right)} \leq \frac{1}{I_*(f)}.$$

The proof of Theorem 2 is in Appendix A.2. Asymptotically speaking, Theorems 1 and 2 imply that $\mathbb{E}_f[\tau] \approx \frac{\log \frac{1}{\delta}}{I_*(f)}$ under any $\delta(\mathcal{M}^{\text{F}}_p)$-accurate policy and for any $f \in \mathcal{M}^{\text{F}}_p$.

Since the set of preferences $\mathcal{M}_p^{\text{F}}$ is arbitrary (except for the finiteness requirement), this relationship suggests that the dependence between consumers' preference f and the speed of learning of any δ-accurate policy can be quantified by Chernoff's information measure $I_*(f)$. In particular, we expect that the larger the value of $I_*(f)$ the faster one can learn the underlying preference f. Since our goal is to find a δ-accurate policy that minimizes the amount of time needed to learn the top-ranked version uniformly over the class of p-Separable choice models, the objective now turns to identifying the policy that performs well against the p-Separable choice model with the smallest value of $I_*(f)$. In the next section, we characterize this hardest-to-learn consumer choice model (which we refer to it as *Ordinal Attraction* choice model or OA) by solving a robust version of the Max-Min problem above. Then, in Section 2.7, we propose a δ-accurate policy that (asymptotically) achieves the lower bound in Theorem 1 and satisfies $\mathbb{E}_f[\tau] \approx \frac{\log \frac{1}{\delta}}{I_*(f)}$ under the OA model.

2.6 Worst-Case Analysis: Ordinal Attraction Model

Motivated by the lower bound in Theorem 1, in this section we find a subset of preferences in $\mathcal{M}_p$ that all have the smallest value of $I_*(f)$. To this end, let

$$I_*^{\text{OA}} := \inf_{f \in \mathcal{M}_p} I_*(f). \tag{2.2}$$

The reason for the superscript "OA" defining I_*^{OA} will become apparent in Theorem 3 below. If f is any preference such that $I_*(f) = I_*^{\text{OA}}$ then any permutation of f also minimizes $I_*(f)$ since the labeling of the K versions is completely arbitrary.[5]

A subset of $\arg\min_{f \in \mathcal{M}_p} I_*(f)$ is described in Theorem 3 below. These minimizing preferences make use of the following definition:

$$\sigma(X|S) := \sum_{k \in S} \mathbb{I}\{\sigma(k) \leq \sigma(X)\} \quad \text{for any } X \in S.$$

5. Indeed, given any $f \in \mathcal{M}_p$ and any permutation $\sigma \in \Sigma$, let $f_\sigma \in \mathcal{M}_p$ be defined by $f_\sigma(X|S) = f(\sigma(X)|\sigma(S))$ for all $X \in [K]$ and $S \in \mathcal{S}$. Then, it is not hard to see that $I_*(f) = I_*(f_\sigma)$.

That is, $\sigma(\cdot|S) : S \to [|S|]$ is the restriction of σ to S so that $\sigma(k_1|S) < \sigma(k_2|S)$ if and only if $\sigma(k_1) < \sigma(k_2)$, for any $k_1, k_2 \in S$.

Theorem 3. *Let $p \in [0,1)$. For any $\sigma \in \Sigma$ define the preference*

$$f_\sigma^{\text{OA}}(X|S) := \frac{1-p}{1-p^{|S|}}\, p^{\sigma(X|S)-1} \qquad S \in \mathcal{S} \text{ and } X \in S. \tag{2.3}$$

and let

$$\mathcal{M}_p^{\text{OA}} := \{f_\sigma^{\text{OA}} \colon \sigma \in \Sigma\}. \tag{2.4}$$

Then $\mathcal{M}_p^{\text{OA}} \subseteq \arg\min_{f \in \mathcal{M}_p} I_*(f)$.

The proof of Theorem 3 is in Appendix A.3.

Let $\overline{\mathcal{M}}_p^{\text{OA}}(f) := \{\bar{f} \in \mathcal{M}_p^{\text{OA}} \colon \sigma_{\bar{f}}(1) \neq \sigma_f(1)\}$ be the class of OA preferences that disagree with f on the top-ranked version. The following result follows from the proof of Theorem 3.

Corollary 1. *For any $f \in \mathcal{M}_p^{\text{OA}}$ we have that*

$$I_*^{\text{OA}} = I_*(f) = \max_{\lambda \in \Delta(\mathcal{S})} \min_{\bar{f} \in \overline{\mathcal{M}}_p^{\text{OA}}(f)} D_\lambda\left(f||\bar{f}\right).$$

In other words, when computing the value of I_*^{OA}, we can replace the set of alternative preferences $\overline{\mathcal{M}}_p(f)$ in the Max-Min problem by the considerably smaller set of alternatives OA preferences $\overline{\mathcal{M}}_p^{\text{OA}}(f)$. As we will see, this reduces significantly the complexity of characterizing the optimal randomization strategy $\lambda \in \Delta(\mathcal{S})$. We will also exploit Corollary 1 to formulate a robust version of the company's problem.

Discussion of the set of OA preferences $\mathcal{M}_p^{\text{OA}}$. To get some intuition about the structure of a preference $f_\sigma^{\text{OA}} \in \mathcal{M}_p^{\text{OA}}$, note that Theorem 3 implies that the likelihood that a consumer selects version X out of the display set S is proportional to the relative ranking of product X within S. In other words, the "attractiveness" of product X has an ordinal dependence on the set S. It is because of this property that we refer to f_σ^{OA} as an *Ordinal Attraction* (OA) choice model. Mathematically, it is not hard to see that f_σ^{OA} satisfies the p-Separability requirement (A-4) in Definition 1 with equalities.

Although we derived the OA preferences for the purpose of characterizing the smallest value of Chernoff's information measure I_* within the class of p-Separable choice model, the OA

model has a number of additional properties. For instance, the OA model is an extension of the popular class of noisy pairwise comparison models that have been widely used in the literature (e.g. Braverman and Mossel, 2008, Braverman and Mossel, 2009, Caragiannis et al., 2013, Wauthier et al., 2013, to name a few). As its name suggests, in a pairwise noisy comparison model, only pairs of products are displayed and the better version is chosen with a fixed probability independent of the actual pair being displayed. It is easy to see that when only pairs are displayed, the Ordinal Attraction model reduces to the pairwise comparison model.

From a practical standpoint, another appealing feature of the OA model is that it provides a parsimonious framework to study consumer choice behavior using limited information about product version characteristics. Indeed, by its ordinal nature, the only relevant attribute of a version for the purpose of affecting consumers' voting choices is its relative ranking within the display set.

Finally, as we will see below, the OA model is also tractable and allows us to derive a complete closed-form characterization of our proposed Myopic Tracking algorithm. We leverage this solution to derive valuable insights into the structure of an optimal policy and how to effectively balance the coverage-accuracy trade-off discussed in the Introduction.

Computing the Information Measure I_*^{OA}. Using the result in Theorem 3 and Corollary 1, let us now turn to the question of determining the lower bound for Chernoff's information measure I_* that solves the Max-Min problem above. That is, for any $f_\sigma^{\text{OA}} \in \mathcal{M}_p^{\text{OA}}$, we are interested in computing

$$I_*^{\text{OA}} = \max_{\lambda \in \Delta(\mathcal{S})} \min_{\bar{f} \in \overline{\mathcal{M}}_p(f_\sigma^{\text{OA}})} D_\lambda \left(f_\sigma^{\text{OA}} || \bar{f} \right).$$

We note that the specific ranking σ that we use in this definition is immaterial since the preference f_σ^{OA} is only defined up to permutations of σ (see footnote 5). So, to simplify the notation, we will assume that σ is equal to the consensus ranking $\sigma_*(k) = k$ for all $k \in [K]$ and denote $f_*^{\text{OA}} = f_{\sigma_*}^{\text{OA}}$.

Our next result characterizes optimal solutions for the randomized display strategy problem

$$\max_{\lambda\in\Delta(\mathcal{S})}\min_{\bar{f}\in\overline{\mathcal{M}}_p(f_*^{\text{OA}})} D_\lambda\left(f_*^{\text{OA}}||\bar{f}\right). \tag{2.5}$$

Theorem 4. *Define the sequences* $\{\mathfrak{a}_n\}$, $\{\mathfrak{b}_n\}$ *and* $\{\lambda_n^*\}$ *of positive real numbers as follows:*

$$\begin{aligned}\mathfrak{a}_n &:= \log\left(\tfrac{1}{p}\right)\left[1-np^{n-1}+(n-1)p^n\right], \quad \mathfrak{b}_n = 1-p^n,\\ \textit{and}\quad \lambda_n^* &= \begin{cases}\mathfrak{b}_n\left(\frac{1}{\mathfrak{a}_n}-\frac{1}{\mathfrak{a}_{n+1}}\right), & \textit{if } n=2,\ldots,K-1;\\ \frac{\mathfrak{b}_n}{\mathfrak{a}_n}, & \textit{if } n=K.\end{cases}\end{aligned} \tag{2.6}$$

Then, the unique optimal solution (of the outer maximization) of (2.5) *is given by*

$$\lambda_*^{\text{OA}}(S) = \begin{cases}\frac{\lambda_n^*}{\lambda_2^*+\cdots+\lambda_K^*} & \textit{if } S=[n] \textit{ for some } n\in\{2,\ldots,K\}\\ 0 & \textit{otherwise.}\end{cases} \tag{2.7}$$

The result in Theorem 4 is significant for a number of reasons. First, it shows that the randomized display rule in (2.7) is static, independent of the voting history, and can be computed offline. Second, it reveals a nested structure of the display sets that have a positive probability of being offered. Indeed, note that the support of λ_*^{OA} is the collection of display sets: $\big\{[2],[3],\ldots,[K]\big\}$, where $[k]=\{1,2,\ldots,k-1,k\}$. Finally, this collection is rather sparse (there are only $K-1$ members out of the 2^K-1 possible display sets) which is a fact that simplifies significantly its computation and implementation.

Equipped with Theorem 4, we can now compute the smallest value of $I_*(f)$ within the class of p-Separable preferences.

Proposition 1. *The value of* I_*^{OA} *is*

$$I_*^{\text{OA}} = (1-p)\,\log\left(\tfrac{1}{p}\right)\left(1+\sum_{n=2}^{K}\frac{p^{n-1}}{1+2p+\cdots+(n-1)p^{n-2}}\right)^{-1}.$$

It follows that $(1-p)\,\log\left(\frac{1}{p}\right)K\,\left(K+2p(K-1)\right)^{-1} \le I_*^{\text{OA}} \le (1-p)\,\log\left(\frac{1}{p}\right)\,(1+p)^{-1}$.

From Proposition 1, one can see that I_*^{OA} decreases in both K and p. Also, numerical computations show that I_*^{OA} is not particularly sensitive to the value of K (the total number

of versions). On the other hand, the impact of the noise parameter p is roughly given by $I_*^{\mathrm{OA}} \approx (1-p)^2$ as $p \uparrow 1$ and $I_*^{\mathrm{OA}} \approx \log \frac{1}{p}$ as $p \downarrow 0$.

2.7 Robust Learning and Myopic Tracking Policy

In this section we propose a particular δ-accurate display policy, which we refer to it as *Myopic Tracking Policy* (MTP), and show that it is worst-case asymptotically optimal in a sense that we make precise in Theorem 5.

Inspired by the discussion in the previous section, the implementation of the Myopic Tracking Policy leverages the structure of the OA model. More specifically, it has three main steps: (1) an MLE estimation of the true ranking (*i.e.*, most likely hypothesis) restricted to the OA model, (2) a stopping criteria to decide whether to stop the voting process or to continue it, given the available information, and (3) a random selection of the display set to show to the next consumer in case the voting process is continued. The specific details of the Myopic Tracking policy are provided in Algorithm 1 below, whose statement makes use of the following definitions. First, given any history $H_t = (S_1, X_1, \ldots, S_t, X_t)$ and any pair of preferences $f, \bar{f} \in \mathcal{M}_p$, we define the log-likelihood ratio process

$$L_t^{f,\bar{f}} := \sum_{\ell=1}^{t} \log\left(\frac{f(X_\ell|S_\ell)}{\bar{f}(X_\ell|S_\ell)}\right). \tag{2.8}$$

Second, for any preference $f \in \mathcal{M}_p$, we define a nested sequence of display sets $\{S_f(k)\colon k = 2, \ldots, K\}$ such that $S_f(k) := \{\sigma_f^{-1}(\ell)\colon \ell = 1, \ldots, k\}$ is the set that includes the top-k versions under f.

Algorithm 1 is parameterized by a single exogenous parameter β (possibly depending on the error tolerance δ) that is used in the stopping criterion in Step 2. The choice of the parameter β is critical to ensure that the Myopic Tracking policy is both fast and δ-accurate. We discuss these two issues next.

Algorithm 1: Myopic Tracking Policy (MTP)

INPUT: A scalar $\beta = \beta(\delta) > 0$.

STEP 0: (Initialization). Set $t = 1$, select an arbitrary display set $S_1 \in \mathcal{S}$ to show to the first consumer and record the vote X_1.

STEP 1: At time epoch t, given the history of votes $(S_1, X_1, \ldots, S_t, X_t)$, compute a most likely consensus preference by solving the MLE problem

$$f_t^{\text{OA}} \in \arg\max_{f \in \mathcal{M}_p^{\text{OA}}} \sum_{\ell=1}^{t} \log f(X_\ell | S_\ell). \tag{MLE}$$

We break ties uniformly at random if the arg max in (MLE) is not a singleton.

STEP 2: Update the value of the generalized log-likelihood ratio process

$$\mathcal{L}_t = \min_{\bar{f} \in \overline{\mathcal{M}}_p^{\text{OA}}(f_t^{\text{OA}})} L_t^{f_t^{\text{OA}}, \bar{f}}. \tag{L}$$

If $\mathcal{L}_t \geq \beta$, then stop set $\tau = t$ and select the top-ranked version according to f_t^{OA}; that is, $\sigma^{-1}_{f_t^{\text{OA}}}(1)$. Otherwise, go to Step 3.

STEP 3: For $k = 2, \ldots, K$, let $\hat{S}_t(k) = S_{f_t^{\text{OA}}}(k)$ and $\hat{\lambda}_t(k) = \lambda_*^{\text{OA}}([k])$ (see Theorem 4).

Using the vector of probabilities $\hat{\lambda}_t(k)$, randomly select from the set $\{\hat{S}_t(2), \hat{S}_t(3), \ldots, \hat{S}_t(K)\}$ the next display set S_{t+1} to be displayed to the next consumer and record her choice X_{t+1}. Go to Step 1 and iterate. □

2.7.1 Worst-case Asymptotic Optimality of the Myopic Tracking Policy.

We now show that the Myopic Tracking Policy is worst-case asymptotically optimal for an appropriate value of parameter β. This result is summarized in the theorem below.

Theorem 5. (Worst-case asymptotic optimality of MTP) *There exists a threshold $\beta = \beta(\delta)$ such that*

$$\text{MTP} \in \arg\min_{\pi \text{ is } \delta\text{-accurate}} \sup_{f \in \mathcal{M}_p} \limsup_{\delta \downarrow 0} \frac{\mathbb{E}_f^{\pi}[\tau]}{\log(1/\delta)}. \tag{2.9}$$

The proof of Theorem 5 consists of two parts. First, in Theorem 6, we show how to select β so that MTP uses an expected number of votes that asymptotically matches the

lower bound in Theorem 1. Second, in Theorem 7, we show how to select β to ensure that MTP belongs to the family of δ-accurate policies.

Both the stopping time τ and threshold β depend on the error tolerance δ. More specifically, because of the hitting time property (see Figure 2.1), τ is an increasing function of β, and let us call this dependence $\tau(\beta)$. Meanwhile, we also expect that the lower the tolerance δ is, the higher β needs to be. Let us denote this dependence as $\beta(\delta)$. Intuitively, β is an decreasing function of δ. Combining $\tau(\beta)$ and $\beta(\delta)$, we expect $\tau = \tau(\beta(\delta))$ to decrease in δ.

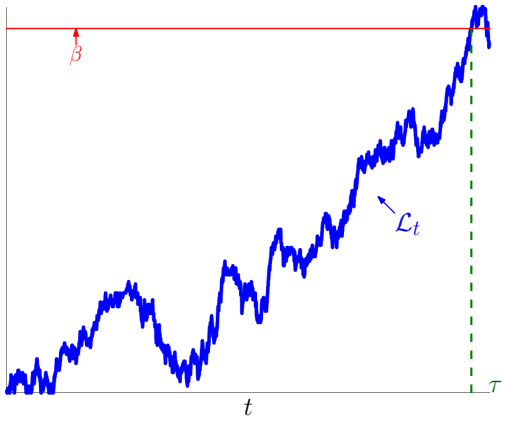


Figure 2.1: **Illustration of the Myopic Tracking Policy.** The stopping time τ is a hitting time.

Theorem 6 below gives a sufficient condition on how "small" $\beta = \beta(\delta)$ needs to be for the Myopic Tracking Policy to be fast in the sense of Theorem 1.

Theorem 6. (Sample Complexity of MTP) *For any constant C_0 (independent of δ), threshold $\beta = \beta(\delta)$ such that $\beta \leq C_0 + \log \frac{1}{\delta}$ and preference $f \in \mathcal{M}_p$, we have that $\mathbb{E}_f[\tau] < \infty$ for every $\delta > 0$ under the Myopic Tracking Policy. Moreover,*

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_f[\tau]}{\log\left(\frac{1}{\delta}\right)} \leq \frac{1}{I_*^{\mathrm{OA}}}. \tag{2.10}$$

The inequality above can be replaced by an equality when $f \in \mathcal{M}_p^{\mathrm{OA}}$.

Theorem 6 implies that the MTP uses a number of votes that asymptotically matches the lower bound of any δ-accurate policy. The key idea behind the proof of Theorem 6 is that

the MTP matches the randomization strategy λ_*^{OA} in Theorem 4. As a result, the stochastic process $\mathcal{L}_t$ achieves the fastest rate of growth (*i.e.*, maximum drift). We will expand more on this reasoning in Section 2.7.3.

Our next result gives a sufficient condition on how "large" $\beta = \beta(\delta)$ needs to be for the Myopic Tracking Policy to be δ-accurate.

Theorem 7. (Accuracy of MTP) *There exists a constant C_1, that depends only on K, such that as long as $\beta \geq C_1 + \log\left(\frac{1}{\delta}\right)$, the Myopic Tracking policy is δ-accurate for every $\delta \in (0,1)$.*

The hitting threshold β controls the error probability of the MTP algorithm. Intuitively, the higher the threshold the less likely it is that we will end up selecting the wrong version in the end. The proof of Theorem 7 is based on a change-of-measure argument, and does not rely on the specific structure of the MTP (except the requirement that $\tau < \infty$ $\mathbb{P}_f-$a.s. for any $f \in \mathcal{M}_p^{\text{OA}}$). The construction of threshold β is based on an estimate of the error probability $\mathbb{P}_f(d_\tau \neq \sigma_f^{-1}(1))$ for every $f \in \mathcal{M}_p$. The main idea behind the estimation is two-fold: first, by tracking the generalized log-likelihood process restricted to the OA model, the MTP achieves δ-accuracy within the OA model (based on a change-of-measure argument); second, since OA model is the "hardest to learn", achieving δ-accuracy within the OA model also implies achieving δ-accuracy within $\mathcal{M}_p$ (based on a dominance argument).[6]

Finally, we can combine the results in Theorems 1, 6 and 7 to establish the worst-case asymptotic optimality of the MTP.

Proof of Theorem 5. Select an arbitrary $\delta-$accurate policy π. Invoking Theorem 1 and

6. Chernoff (1959) proposed a similar type of hitting threshold for an alternative performance criterion. His analysis, however, assumed that the number of states (i.e., preferences in our setting) is finite and does not extend to our case in which the cardinality of $\mathcal{M}_p$ is infinite. More recently, in the context of a best arm identification problem, Garivier and Kaufmann (2016) have also proposed a sampling policy that relies on a similar hitting threshold like the one in Theorem 7. However, because of the structure of their problem, to ensure δ-accuracy their threshold must grow at a rate of $\log(t)$ over time while in our case the threshold is constant independent of t.

(2.2),

$$\sup_{f\in\mathcal{M}_p} \limsup_{\delta\downarrow 0} \frac{\mathbb{E}_f^{\pi}[\tau]}{\log(1/\delta)} \;\geq\; \sup_{f\in\mathcal{M}_p} \frac{1}{I_*(f)} \;=\; \frac{1}{I_*^{\mathrm{OA}}}.$$

By Theorem 7, MTP is $\delta-$accurate if we pick $\beta = C_1 + \log\left(\frac{1}{\delta}\right)$. Invoking Theorem 6 with that β yields

$$\sup_{f\in\mathcal{M}_p} \limsup_{\delta\downarrow 0} \frac{\mathbb{E}_f^{MTP}[\tau]}{\log(1/\delta)} \;\leq\; \sup_{f\in\mathcal{M}_p} \frac{1}{I_*^{\mathrm{OA}}} \;=\; \frac{1}{I_*^{\mathrm{OA}}}.$$

As a result, (2.9) holds. ■

Remark 2. The conditions in Theorem 6 and 7 may be further weakened. For example, f_t^{OA} defined in (MLE) does not have to be the maximum likelihood estimator, but any statistics so that:

$$\mathbb{E}_{f_*^{\mathrm{OA}}}[\hat{\tau}^2] < \infty, \text{ where } \hat{\tau} := \max\{t : f_t^{\mathrm{OA}} \neq f_*^{\mathrm{OA}}\} \tag{2.11}$$

Here $\hat{\tau}$ is a $\mathbb{Z}_+ \cup \{+\infty\}$-valued, $\mathcal{F}_\infty$-measurable[7] random variable that denotes the last time period in which the estimated preference $f_t^{\mathrm{OA}} \neq f_*^{\mathrm{OA}}$. ◇

2.7.2 On the Complexity of the Myopic Tracking Policy

In this section we discuss the computational complexity of the Myopic Tracking Policy. Each iteration of Algorithm 1 involves three steps. Because of the result in Theorem 4, the third step is rather simple as it involves randomly selecting a display set out of $K-1$ alternatives. Since the randomization probabilities are fixed, independent of the history of the learning process, this step can be executed offline before the feedback process begins. Steps 1 and 2, on the other hand, involve solving a possibly large combinatorial optimization problem at each iteration. Fortunately, the structure of the underlying OA preference model allows us to solve these combinatorial problems without much computational burden. The rest of this section is devoted to supporting this claim.

To this end, let us represent the voting history $H_t = (S_1, X_1, \ldots, S_t, X_t)$ using a complete directed graph $\mathcal{G}_K$ with K nodes, each representing a version. For each arc $(i,j) \in [K]\times[K]$,

7. $\mathcal{F}_\infty$ is defined as the smallest sigma algebra containing $\cup_{t=1}^{\infty}\mathcal{F}_t$.

we define its weight $w_{i,j}^t$ by

$$w_{ij}^t := \sum_{\ell=1}^{t} \mathbb{I}\big\{\{i,j\} \subseteq S_\ell \text{ and } X_\ell = i\big\}. \tag{2.12}$$

That is to say, w_{ij}^t is the total number of instances, up to time t, where both version i and j are jointly displayed and version i is voted for. Intuitively, if $w_{ij}^t - w_{ji}^t$ is large then there is a strong indication that version i has a higher ranking than version j.

We use the graph $\mathcal{G}_K$ to quantify the discrepancy between any given preference f and the voting history using the total weight of the *feedback arc set* $\{(i,j) : \sigma_f(j) < \sigma_f(i)\}$. We introduce the discrepancy cost

$$c(f, \vec{w}^t) := \sum_{(i,j):i\neq j} \mathbb{I}\{\sigma_f(j) < \sigma_f(i)\}\, w_{ij}^t. \tag{2.13}$$

The argument inside the summation is the total number of instances where a pair of versions (i,j) are jointly displayed and the less preferred version i under f (that is, $\sigma_f(i) > \sigma_f(j)$) is chosen.

In the result below, we demonstrate that the log likelihood of any preference f given voting history H_t is proportional to its discrepancy cost $c(f, \vec{w}^t)$. As a result, (MLE) (resp. (L)) corresponds to an unconstrained (resp. constrained) discrepancy cost minimization problem.

Proposition 2. *Given the voting history $H_t = (S_1, X_1, \ldots, S_t, X_t)$, the following facts hold:*

1. *There exists a constant ϕ such that for any $f \in \mathcal{M}_p^{\text{OA}}$, $\sum_{\ell=1}^t \log f(X_\ell; S_\ell) = \log(p) \cdot c(f, \vec{w}^t) + \phi$. Hence* (MLE) *is equivalent to finding an*

$$f_t^{\text{OA}} \in \operatorname*{arg\,min}_{f \in \mathcal{M}_p^{\text{OA}}} c(f, \vec{w}^t).$$

2. *Given any $f, \bar{f} \in \mathcal{M}_p^{\text{OA}}$, $L_t^{f,\bar{f}} = \log\left(\frac{1}{p}\right) \cdot d(f, \bar{f})$, where $d(\cdot,\cdot)$ is defined as*

$$d(f,\bar{f}) := c(\bar{f}, \vec{w}^t) - c(f, \vec{w}^t) = \sum_{(i,j):i\neq j} (w_{i,j}^t - w_{j,i}^t)\, \mathbb{I}\{\sigma_{\bar{f}}(j) < \sigma_{\bar{f}}(i) \text{ but } \sigma_f(j) > \sigma_f(i)\}. \tag{2.14}$$

Hence (L) *is equivalent to solving*

$$\mathcal{L}_t = \min_{\bar{f} \in \overline{\mathcal{M}}_p^{\text{OA}}(f_t^{\text{OA}})} d(f_t^{\text{OA}}, \bar{f}).$$

Proposition 2 reveals that the (MLE) problem in Algorithm 1 is computationally equivalent to the *weighted feedback arc set problem on tournaments* on graph $\mathcal{G}_K$ with weights $\vec{w}^t$ (Ailon et al., 2005). This type of problems has been extensively studied in the computer science (Davenport and Kalagnanam, 2004, Ailon et al., 2005, Alon, 2006, Conitzer et al., 2006, Charbit et al., 2007, Kenyon-Mathieu and Schudy, 2007, Schalekamp and Zuylen, 2009, Fomin et al., 2010, etc.) and operations research (Grötschel et al., 1984, Mitchell and Borchers, 1996, Charon and Hudry, 2010) literature. For example, acceleration algorithms are available for the following integer programming formulation of this problem (Grötschel et al., 1984): $\sigma_{\hat{f}}(i) = \sum_{j \in [K] \setminus \{i\}} \hat{x}_{ji} + 1$, where:

$$\begin{aligned} \hat{x} \in \arg\min_{\vec{x}} & \sum_{(i,j): i \neq j} x_{ji} w_{ij}^t \\ s.t. \quad & x_{ij} + x_{jk} + x_{ki} \geq 1, \quad \forall \text{ distinct } i,j,k \in [K] \\ & x_{ij} + x_{ji} = 1, \qquad \forall \text{ distinct } i,j \in [K] \\ & x_{ij} \in \{0,1\}. \qquad \forall \text{ distinct } i,j \in [K] \end{aligned} \tag{2.15}$$

It has also being reported that the integer programming formulation is effective both on randomly generated data sets (Conitzer et al., 2006) and on real data sets (Ali and Meila, 2012). Moreover, Kenyon-Mathieu and Schudy (2007) propose a polynomial-time approximation scheme (PTAS) to approximate the problem. That is, when the number of versions is large and computational tractability is a concern, for any fixed ϵ, a polynomial-time algorithm with $1-\epsilon$ optimal sample complexity is attainable. Note also that (L) can be reduced to (MLE) in linear time: for any $k \in [K] \setminus \{\sigma_{\hat{f}}^{-1}(1)\}$, we can solve for the most likely ranking conditional on k being top ranked. To do that, we just need to solve a weighted feedback arc set problem with sub-graph $\mathcal{G}_K \setminus \{k\}$, and insert k back to the ranking. We can solve

(L) by comparing the optimal costs for all the sub-problems.

2.7.3 Discussion of the Myopic Tracking Policy

We conclude this section with a brief discussion of some key features of our proposed Myopic Tracking Policy. Let us start by providing some intuition behind the (asymptotic) optimality of MTP both in terms of the stopping rule and the choice of a display policy.

On the optimality of the stopping rule. To get some intuition behind the stopping rule used in the Myopic Tracking policy, we note that if the display policy $\{S_t\}$ is fixed, our problem reduces to a classical sequential (composite and multi-hypothesis) testing problem.[8] In this case, the idea of tracking the generalized log-likelihood ratio process $\mathcal{L}_t$, and stopping when $\mathcal{L}_t$ hits a pre-specified threshold, is known to be asymptotically optimal under various settings (e.g., Wald, 1973, Chernoff, 1972, Draglia et al., 1999, Li et al., 2014). Steps 1 and 2 in the Myopic Tracking policy extend this idea to our more general setting.

On the optimality of the display set policy. Given the (asymptotic) optimality of the sequential likelihood ratio test discussed above, the goal of an optimal display rule is to speed up the learning process by minimizing the time it takes the likelihood ratio process $\mathcal{L}_t$ to hit the threshold β. In other words, to solve the problem

$$\inf_{\{S_t\}} \mathbb{E}^{\pi}_{f_*} \inf_t \big\{t : \mathcal{L}_t \geq \beta\big\}. \tag{2.16}$$

To build some intuition on the growth of $\mathcal{L}_t$, let us pick f_* from the OA model, for otherwise $\mathcal{L}_t$ will grow more quickly, reflecting the fact that the company's learning problem is easier. Without loss of generality, let us pick $f_* = f^{\text{OA}}_{\sigma_*}$ within the OA model specifically. The Myopic Tracking Policy is able to recover the ranking represented by f_*, or in other words, f^{OA}_t is absorbed into the preference f^{OA}_* quickly (in the sense of Equation (2.11)). Hence $\mathcal{L}_t = \min_{\bar{f} \in \overline{\mathcal{M}}^{\text{OA}}_p(f^{\text{OA}}_t)} L_t^{f^{\text{OA}}_t, \bar{f}}$ is well approximated by the process $\min_{\bar{f} \in \overline{\mathcal{M}}^{\text{OA}}_p(\sigma_*)} L_t^{f^{\text{OA}}_*, \bar{f}}$

8. In our problem, each hypothesis corresponds to a version being the top-ranked one, and each hypothesis contains the family of rankings that rank the same version at the top.

as t grows. Further, we may understand the latter process to consist of two components: a deterministic part (which is a deterministic process that grows linearly) and a noise part (which is a random process that diverges sub-linearly). The drift part captures the growth rate of $\mathcal{L}_t$ and could be written in the following manner:

$$\tilde{\mathcal{L}}_t := \min_{\bar{f}\in\bar{f}^{\text{OA}}} \sum_{\ell=1}^{t}\sum_{k\in S_\ell} f_*(k|S_\ell)\,\log\left(\frac{f_*(k|S_\ell)}{\bar{f}(k|S_\ell)}\right) = \min_{\bar{f}\in\overline{\mathcal{M}}_p^{\text{OA}}(f_*)} \sum_{\ell=1}^{t} D_{S_\ell}\left(f_*||\bar{f}\right)$$

$$= t\cdot \underbrace{\min_{\bar{f}\in\overline{\mathcal{M}}_p^{\text{OA}}(f_*)} D_{\bar{\lambda}}\left(f_*||\bar{f}\right)}_{\text{growth rate}}, \quad \text{where } \bar{\lambda}(S) = \underbrace{\frac{\sum_{l=1}^{t}\mathbb{I}\{S_l = S\}}{t}}_{\text{display frequency of set } S}.$$

As a result, by replacing $\mathcal{L}_t$ with $\tilde{\mathcal{L}}_t$, we can replace the optimal hitting problem in (2.16) by the problem of maximizing the average growth rate of $\tilde{\mathcal{L}}_t$. Furthermore, to maximize this average growth rate, it suffices to balance the long-run display frequency of each set to achieve the fastest growth rate of $\tilde{\mathcal{L}}_t$ (see Figure 2.2). This can be done by selecting

$$\bar{\lambda} \in \arg\max_{\lambda\in\Delta(\mathcal{S})} \min_{\bar{f}\in\overline{\mathcal{M}}_p^{\text{OA}}(f_*)} D_{\lambda}\left(f_*||\bar{f}\right).$$

That is, selecting $\bar{\lambda}$ that solves (Max-Min) with preference f_*, so that

$$\tau \approx \beta/\max_{\lambda\in\Delta(\mathcal{S})} \min_{\bar{f}\in\overline{\mathcal{M}}_p^{\text{OA}}(f_*)} D_{\lambda}\left(f_*||\bar{f}\right) = \beta/I_*(f_*).$$

Accuracy versus Coverage. Recall that in the Introduction we have argued that an effective display policy should balance the underlying accuracy/coverage trade-off embedded in the problem. On one hand, the company can show every consumer the entire set $[K]$ to use every vote to learn about every product. On the other hand, the quality of the information collected on each vote is typically higher when the cardinality of the display set is small. As we show next, the Myopic Tracking Policy resolves this tension in a rather parsimonious fashion.

Indeed, one of the key features of the MTP is the simplicity of its display policy (i.e, Step 3 in Algorithm 1). As mentioned above, the policy randomizes over a nested collection of

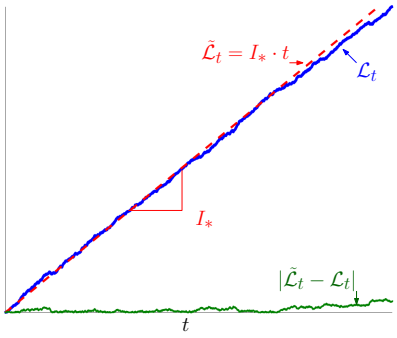


Figure 2.2: **Intuition behind the Myopic Tracking Policy.** Over a long time, $\mathcal{L}_t$ is well approximated by $\tilde{\mathcal{L}}_t$, a linear function. The display policy is chosen to maximize the slope of $\tilde{\mathcal{L}}_t$.

$K-1$ display sets. Specifically, if f_t^{OA} is the MLE estimate of f_* at period t, then the MTP policy randomizes over the sets $\{S_{f_t^{\text{OA}}}(k)\colon k = 2, \ldots, K\}$. (Recall that $S_{f_t^{\text{OA}}}(k)$ is the set that includes the top-k versions under f_t^{OA}.) Furthermore, from Theorem 4, the probability of displaying set $S_{f_t^{\text{OA}}}(k)$ is equal to

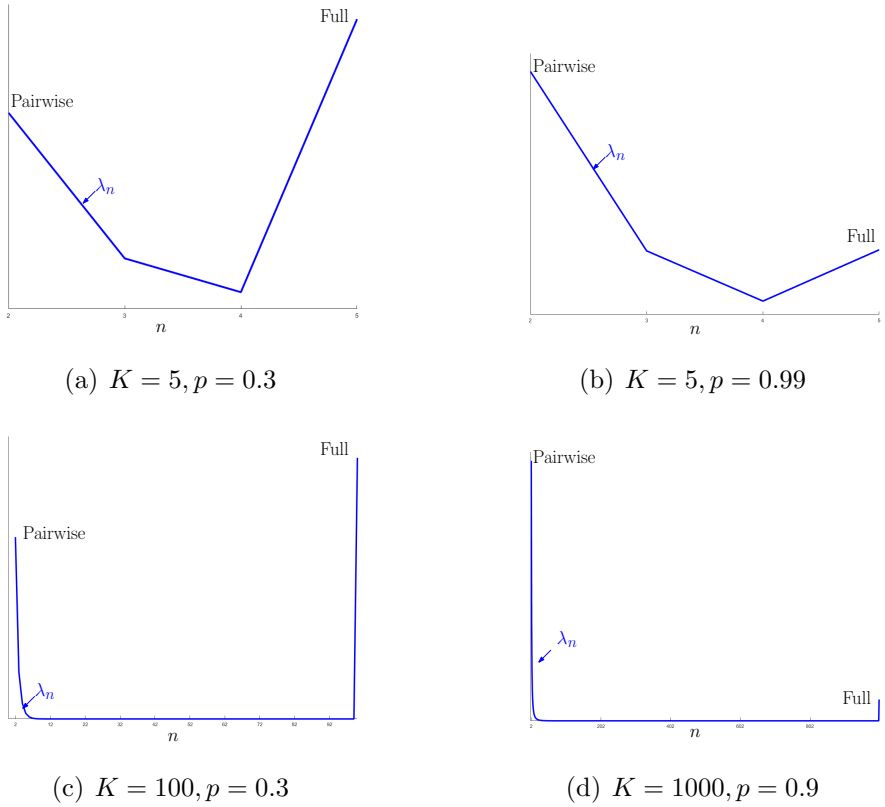
$$\lambda_*^{\text{OA}}(S_{f_t^{\text{OA}}}(k)) = \lambda_k^*/(\lambda_2^* + \cdots + \lambda_K^*) \qquad \text{for } k = 2, \ldots, K, \tag{2.17}$$

where the values of the $\{\lambda_k^*\}$ are given in the theorem. One can show that these randomization probabilities satisfy the following property.

Corollary 2. (U-shaped λ_*^{OA}) *For any $K \geq 4$, we have $\lambda_2^* > \lambda_3^* > \cdots > \lambda_{K-1}^*$ and $\lambda_{K-1}^* < \lambda_K^*$.*

In other words, the vector of randomization probabilities is "U" shaped as a function of the cardinality of the display sets; it decreases with the cardinality and then increases for the full display set $[K]$. The proof of Corollary 2 is in Appendix A.5. Figure 2.3 illustrates the values of $\{\lambda_k^*\}$ for different values of K and p. Roughly speaking, the proposed display strategy in the MTP policy exhibits the following pattern:

- The noisier the environment is (*i.e.,* p is larger), the more probability is allocated to smaller display sets. The less noisy the environment, the more probability is allocated to the full set (*i.e.,* full-display).



(a) $K = 5, p = 0.3$ (b) $K = 5, p = 0.99$

(c) $K = 100, p = 0.3$ (d) $K = 1000, p = 0.9$

Figure 2.3: **Display probabilities λ_*^{OA} of Myopic Tracking Policy for different values of K and p.** In each panel, n is the cardinality of the nested display set $\hat{S}(n)$.

- As the number of versions grows (*i.e.,* K is larger), MTP tends to allocate larger probabilities to either very small display sets (and mostly pairwise) or full-display.

- For sufficiently large values of K, MTP appears to randomize only between pairwise and full-display.

To further underscore these patterns, Table 2.1 reports the values of

$$\underline{\Lambda}_i := \sum_{k=2}^{i} \lambda_k^* \qquad \text{and} \qquad \overline{\Lambda}_i = \sum_{k=K-i+1}^{K} \lambda_k^*,$$

which are the probabilities of selecting a display set with cardinality less than or equal to i or greater than or equal to $N - i + 1$, respectively

As we can see from the table, for small values of p, $\underline{\Lambda}_2 + \overline{\Lambda}_2$ is almost one and so the

(K,p)	$\underline{\Lambda}_2$	$\overline{\Lambda}_2$	$\underline{\Lambda}_5$	$\overline{\Lambda}_5$	$\underline{\Lambda}_{10}$	$\overline{\Lambda}_{10}$
(20,0.1)	0.165	0.811	0.189	0.811	0.189	0.811
(20,0.5)	0.440	0.293	0.671	0.293	0.705	0.294
(20,0.9)	0.468	0.053	0.790	0.065	0.897	0.094
(100,0.1)	0.165	0.811	0.189	0.811	0.189	0.811
(100,0.5)	0.440	0.293	0.671	0.293	0.705	0.293
(100,0.9)	0.465	0.038	0.781	0.038	0.891	0.038
(200,0.1)	0.165	0.811	0.189	0.811	0.189	0.811
(200,0.5)	0.440	0.293	0.671	0.293	0.705	0.293
(200,0.9)	0.465	0.038	0.781	0.038	0.891	0.038

Table 2.1: Λ_i: Probability that MTP selects a display set with cardinality less than or equal to i or greater than or equal to $N - i + 1$ for different values of K and p.

MTP essentially uses two display sets: (i) a pair with the top two versions and (ii) the full set with all K versions, with more than 80% allocated to the full display set. The table also shows that as p grows the display policy of MTP relies on additional display sets of small cardinality. For instance, even when $p = 0.9$ and consumer preferences are rather noisy, $\underline{\Lambda}_{10} \approx 90\%$ independent of K. In other words, if the company has $K = 200$ versions and use the MTP policy, about 90% of the display sets would include less than ten versions.

Pairwise versus Multiwise Comparisons. Proposition 2 reveals an interesting feature about how information accumulates under an OA choice model. It shows that it is *ex-post* equivalent to replace a vote from a multi-wise comparison with a collection of pairwise comparisons. For example, if we replace a vote on Version 1 from the display set $S = \{1, 2, 3\}$ with two consecutive votes on Version 1 from two pairs $\{1, 2\}$ and $\{1, 3\}$ respectively, we do not change the discrepancy cost or likelihood of any preference f. This property explains why we can use a simple graph representation to summarize the voting data. The idea of breaking a multi-wise comparison into several pairwise comparisons relates to Jiang et al. (2011).

Connection to Voting Theory. The discrepancy metric in (2.13) used to solve (MLE) and (L) resembles the one used in the Kemeny-Young method in voting theory (Young, 1988,

Levin and Nalebuff, 1995). As a result, our choice model inherits a number of desirable properties of the Kemeny-Young model. For example, if there exists a version which gets more votes than any other version when they are jointly displayed (known as the *Condorcet winner*), that version is the optimal choice.

2.8 Numerical Experiments

In this section, we numerically investigate the performance of the Myopic Tracking Policy. First, in Section 2.8.1, we study the running time of Algorithm 1. In Section 2.8.2, we compare the sample complexity of MTP and three benchmark policies that use alternative display strategies.

2.8.1 Running Time of the Myopic Tracking policy

Most of the computational time required to implement the Myopic Tracking policy in Algorithm 1 is allocated to the optimization problems in Steps 1 and 2. As mentioned above, Step 3 is computationally inexpensive since it is static and can be computed offline (see Theorem 4). Furthermore, one can show that the optimization in Step 2 can be reduced to a formulation like the one in Step 1 in linear time. Thus, to assess the computational performance of the Myopic Tracking policy we will focus on analyzing the running time needed to solve the MLE problem in Step 1.

Recall that Step 1 is equivalent to a *weighted feedback arc set problem on tournaments* (see Proposition 2), which admits an integer programming (IP) formulation. In our numerical experiments, we solve this IP using an out-of-box solver as well as a heuristic based on its linear programming (LP) relaxation. We evaluate the running times of both methods and also record the relative optimality gaps of our heuristic in comparison to the LP relaxation of (2.15). Specifically, we generate a sequence of instances of the $\vec{w}$ data in (2.12) that arises as MTP iterates under the OA model and compute average running times and optimality gaps (relative to the LP relaxation) of these representative instances of (2.15). The out-of-the box

solver we use is Gurobi 9.0.0 (win64, Python).[9] Both the Gurobi solver and our heuristic are run on an Intel Core i7-6700 CPU with a frequency of 3.40 GHz. Table 2.2 presents a summary of our numerical study and we refer the reader to Section A.9 for more details on our heuristic and its implementation.

(K, p)	T_{IP}	T_H	Δ	(K, p)	T_{IP}	T_H	Δ
$(25, 0.5)$	0.052	0.015	0%	$(25, 0.9)$	0.059	0.014	0%
$(50, 0.5)$	0.991	0.143	0%	$(50, 0.9)$	0.837	0.143	0%
$(75, 0.5)$	3.805	0.548	0%	$(75, 0.9)$	4.576	0.695	0.001%
$(100, 0.5)$	11.019	1.560	0%	$(100, 0.9)$	13.692	1.878	0.028%
$(125, 0.5)$	20.707	3.279	0%	$(125, 0.9)$	30.576	3.808	0%
$(150, 0.5)$	40.838	6.288	0%	$(150, 0.9)$	60.025	6.383	0%

Table 2.2: Running time of the integer programming formulation T_{IP}, running time (in seconds) of our heuristic T_H, and relative optimality gap Δ of the heuristic for Step 1 of MTP.

The following are some of our main findings:

(i) The exact integer programming formulation can be solved within a reasonable amount of time, even when the number of items is relatively large. For example, even when $K = 100$ and $p = 0.9$, the average running time is within 15 seconds.

(ii) Our proposed heuristic is about an order of magnitude faster than the out-of-box solver. For example, even when $K = 100$ and $p = 0.9$, the average running time is less than 2 seconds.

(iii) In most cases, our heuristic achieves zero optimality gap with respect to the LP relaxation, and when the gap is nonzero, it is nonetheless small. In fact, our heuristic achieves zero optimality gap relative to the LP relaxation in 99.6% of the cases and the maximum gap we observe is 3.3%. This finding is consistent with those in related studies in earlier literature (see, e.g., Conitzer et al., 2006, Schalekamp and Zuylen, 2009).

9. We note that we did not attempt to accelerate the implementation using techniques such as constraint generation, or problem-specific cutting plane methods as in Grötschel et al. (1984). Our main goal in this study is to test the performance of the integer programming formulation while not requiring special purpose software.

Our interpretation of the effectiveness of the heuristic and the speed of our solver is that, while the *weighted feedback arc set problem on tournaments* is NP-hard in general, the actual choice data encountered by MTP concentrate on an "easier-to-solve" subclass of instances (at least under the OA model).

2.8.2 Sample Complexity of MTP

In this subsection, we evaluate the sample complexity (i.e., average number of samples) of our proposed Myopic Tracking Policy and compare it to a number of alternative policies using two separate sets of numerical experiments. First, we consider three variants of the MTP policy that employ different display sampling rules but similar stopping and recommendation rules. In our second set of experiments, we compare the MTP to three policies that have been proposed in the ranking and selection literature.

MTP-based Benchmark Policies: In our first set of numerical experiments we compare the sample complexity of the Myopic Tracking Policy against the following three variations:

1. The Full Display Policy, under which $S_t = [K]$ for all t.

2. The Pairwise Display Policy, under which S_t is randomized over the space of all pairs of items, i.e., $|S_t| = 2$. The randomization probabilities are given in equation (2.18) in Proposition 3 below. Using a similar analysis to one use for the Myopic Tracking Policy, one can show that this Pairwise Display Policy is worst-case asymptotically optimal in $\mathcal{M}_p^{\text{OA}}$ if we restrict ourselves to policies that only use pairwise comparisons.

3. The Pair & Full Display Policy, under which S_t is randomized over the Top 2 (i.e., $\{\sigma^{-1}(1), \sigma^{-1}(2)\}$) and the full set $[K]$. The randomization probabilities are given in equation (2.19) below. Again, for this choice of randomization probabilities, one can show that this Pair & Full Display Policy is worst-case asymptotically optimal in $\mathcal{M}_p^{\text{OA}}$ if we restrict ourselves to policies to only display the Top 2 versions or the full display set.

In what follows, we will use M, F, P, and PF to identify the Myopic Tracking policy, Full Display policy, Pairwise Display policy, and Pair & Full Display policy, respectively. Also, for a given policy $\pi \in \{\text{M,F,P, PF}\}$, we let $\mathcal{S}^\pi \subseteq \mathcal{S}$ denote the set of display sets that are admissible under π, that is, $\mathcal{S}^M = \mathcal{S}$, $\mathcal{S}^P = \{S \subseteq [K] : |S| = 2\}$, $\mathcal{S}^F = \{[K]\}$, and $\mathcal{S}^{PF} = \{[2], [K]\}$.

Our next result justifies the specific choice of the randomized display strategies that we use in the implementation of the P and PF policies.

Proposition 3. *Let the randomization distributions $\lambda_*^{\text{P,OA}}(S) \in \Delta(\mathcal{S}^P)$, $\lambda_*^{\text{PF,OA}}(S) \in \Delta(\mathcal{S}^{PF})$ be given as:*

$$\lambda_*^{\text{P,OA}}(S) := \begin{cases} \frac{1}{K-1} & \text{if } S = \{i, i+1\} \text{ for some } i \in \{1, \ldots, K-1\} \\ 0 & \text{otherwise.} \end{cases} \tag{2.18}$$

and

$$\lambda_*^{\text{PF,OA}}(S) := \begin{cases} \frac{(\mathfrak{a}_3 - \mathfrak{a}_2)/\mathfrak{b}_K}{\mathfrak{a}_2/\mathfrak{b}_2 + (\mathfrak{a}_3 - \mathfrak{a}_2)/\mathfrak{b}_K} & \text{if } S = \{1, 2\} \\ \frac{\mathfrak{a}_2/\mathfrak{b}_2}{\mathfrak{a}_2/\mathfrak{b}_2 + (\mathfrak{a}_3 - \mathfrak{a}_2)/\mathfrak{b}_K} & \text{if } S = [K]. \end{cases} \tag{2.19}$$

where $\mathfrak{a}_n$ and $\mathfrak{b}_n$ are defined in (2.6)*. Then*

$$\lambda_*^{\text{P,OA}} \in \operatorname*{arg\,max}_{\lambda \in \Delta(\mathcal{S}^P)} \min_{\bar{f} \in \overline{\mathcal{M}}_p(f_*^{\text{OA}})} D_\lambda\left(f_*^{\text{OA}} || \bar{f}\right) \quad \text{and} \quad \lambda_*^{\text{PF,OA}} \in \operatorname*{arg\,max}_{\lambda \in \Delta(\mathcal{S}^{PF})} \min_{\bar{f} \in \overline{\mathcal{M}}_p(f_*^{\text{OA}})} D_\lambda\left(f_*^{\text{OA}} || \bar{f}\right).$$

The following proposition gives performance guarantees and theoretical justifications of our policies F, P and PF. The statement of Proposition 4 uses the shorthand notation $I^\pi = \max_{\lambda \in \Delta(\mathcal{S}^\pi)} \min_{\bar{f} \in \overline{\mathcal{M}}_p(f_*^{\text{OA}})} D_\lambda\left(f_*^{\text{OA}} || \bar{f}\right)$.

Proposition 4. *With the stopping threshold $\beta = C_1 + \log(1/\delta)$ as in Theorem 7, all of the policies $\{\text{F}, \text{P}, \text{PF}\}$ are $\delta-$accurate. Their sample complexities are such that for all $f \in \mathcal{M}_p$,*

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_f^F[\tau]}{\log\left(\frac{1}{\delta}\right)} \leq \frac{1}{I^F}, \quad \limsup_{\delta \downarrow 0} \frac{\mathbb{E}_f^P[\tau]}{\log\left(\frac{1}{\delta}\right)} \leq \frac{1}{I^P}, \quad \text{and} \quad \limsup_{\delta \downarrow 0} \frac{\mathbb{E}_f^{PF}[\tau]}{\log\left(\frac{1}{\delta}\right)} \leq \frac{1}{I^{PF}}, \tag{2.20}$$

where

$$I^F = (1-p)\log\left(\tfrac{1}{p}\right)\frac{1}{(1-p^K)/(1-p)}, \qquad I^P = (1-p)\log\left(\tfrac{1}{p}\right)\frac{1}{(K-1)(1+p)} \quad \textit{and}$$
$$I^{PF} = (1-p)\log\left(\tfrac{1}{p}\right)\frac{1+2p}{(1-p^K)/(1-p)+2p(1+p)}.$$

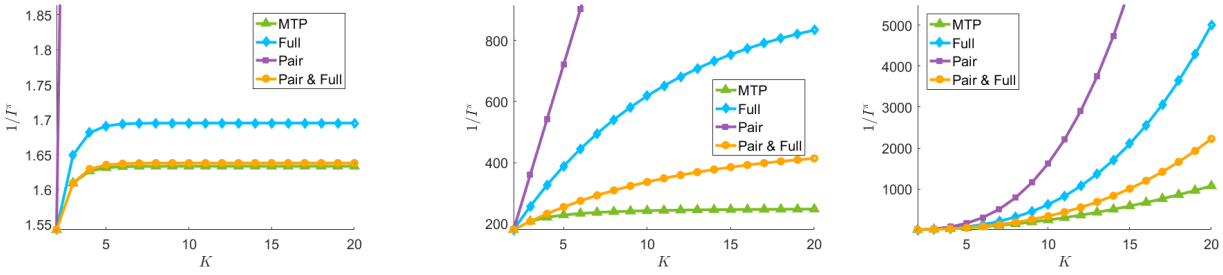
Moreover, if $f \in \mathcal{M}_p^{\text{OA}}$ *then*

$$\lim_{\delta\downarrow 0}\frac{\mathbb{E}_f^F[\tau]}{\log\left(\frac{1}{\delta}\right)} = \frac{1}{I^F} \quad \textit{and} \quad \lim_{\delta\downarrow 0}\frac{\mathbb{E}_f^P[\tau]}{\log\left(\frac{1}{\delta}\right)} = \frac{1}{I^P}. \tag{2.21}$$

As a result, F *and* P *are worst-case asymptotically optimal within the classes of full and pairwise display* δ*-accurate policies, respectively. That is, if we let* $\mathcal{A}^F$ *(resp.* $\mathcal{A}^P$*) be the space of* δ*-accurate policies such that* $S_t \in \mathcal{S}^F$ *(resp.* $S_t \in \mathcal{S}^P$*), we have*

$$\text{F} \in \arg\min_{\pi\in\mathcal{A}^F}\sup_{f\in\mathcal{M}_p}\limsup_{\delta\downarrow 0}\frac{\mathbb{E}_f^\pi[\tau]}{\log(1/\delta)} \quad \textit{and} \quad \text{P} \in \arg\min_{\pi\in\mathcal{A}^P}\sup_{f\in\mathcal{M}_p}\limsup_{\delta\downarrow 0}\frac{\mathbb{E}_f^\pi[\tau]}{\log(1/\delta)}. \tag{2.22}$$

We use Proposition 4 to conduct a sensitivity analysis of the policies $\pi \in \{\text{M,F,P, PF}\}$ on p and K when δ is sufficiently small. Figure 2.4 depicts the values of the $1/I^\pi$ for for different values of p and K.



(a) $p = 0.3$ (b) $p = 0.9$ (c) $p = 1 - 1/K$

Figure 2.4: I^π **as function of** K. Values of p are taken for $p = 0.3$, $p = 0.9$, and $p = 1-1/K$ respectively.

We can see from the figure that the Myopic Tracking Policy offers a significant advantage over the other policies, specially over F and P, when the number of versions K and/or the noise parameter p are large. In particular, when $K \uparrow \infty$ and $p \uparrow 1$ at the same time (far-most right panel), the values of $1/I^\pi$ under Full, Pairwise, and Pair & Full display policies perform

arbitrarily bad compared to that of MTP. The following corollary formalizes some of these observations.

Corollary 3. *As a direct consequence of Theorem 6, Proposition 1, and Proposition 4, we have that for* $f \in \mathcal{M}_p^{\text{OA}}$

$$\lim_{K\uparrow\infty,p\uparrow 1} \lim_{\delta\downarrow 0} \frac{\mathbb{E}_f^F[\tau]}{\mathbb{E}_f^M[\tau]} \geq \lim_{K\uparrow\infty,p\uparrow 1} \frac{K}{K+2p(K-1)}(1+p+\cdots+p^{K-1}) = \infty;$$
$$\lim_{K\uparrow\infty,p\uparrow 1} \lim_{\delta\downarrow 0} \frac{\mathbb{E}_f^P[\tau]}{\mathbb{E}_f^M[\tau]} \geq \lim_{K\uparrow\infty,p\uparrow 1} \frac{K}{K+2p(K-1)}(K-1)(1+p) = \infty.$$

That is, the policies P *and* F *could perform "arbitrarily bad" compared to MTP when* $p \uparrow 1$*,* $K \uparrow \infty$*, and* $\delta \downarrow 0$*.*

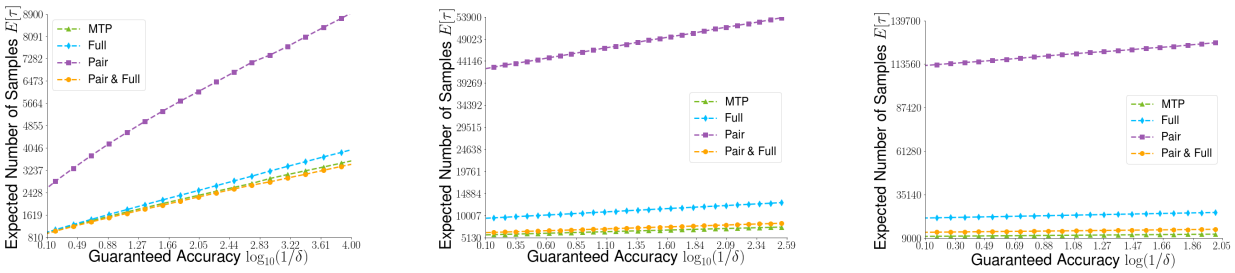
Let us now conduct a set of numerical simulations to compare the value of $\mathbb{E}_{f_*}^{\pi}[\tau]$ for $\pi \in \{\text{M, F, P, PF}\}$ and to investigate the sensitivity of the MTP sample complexity to δ. In our simulation study, we fix $p = 0.9$, and set $K \in \{4, 10, 15\}$ with appropriate values of β. For each problem instance (K, p), we evaluate two metrics simultaneously:

- The sample complexity as a function of δ for the learning algorithm to be $\delta(\mathcal{M}_p)$ accurate. This metric is the closest to what we theoretically studied earlier. While many values of β are appropriate (i.e., guarantee low error probability and asymptotically optimal sample complexity), we select $\beta = C_1 + \log\left(\frac{1}{\delta}\right)$, where $C_1 = \log\big((K-1)(K-1)!\big)$. Note that the current value of β is only an upper bound of what is needed to guarantee δ-accuracy. With that said, it is the smallest value we know in our proof. Moreover, the resulting loss of performance by picking a conservative β is negligible asymptotically as $\delta \downarrow 0$;

- The sample complexity as a function of the empirical error probability $\hat{\delta}$, as we vary different levels of β. Here the empirical error probability is defined as the fraction of instances in which the algorithm terminates with an item different from $\sigma_{f_*}^{-1}(1)$ under the preference f_*. This metric characterizes where a learning algorithm stands in the

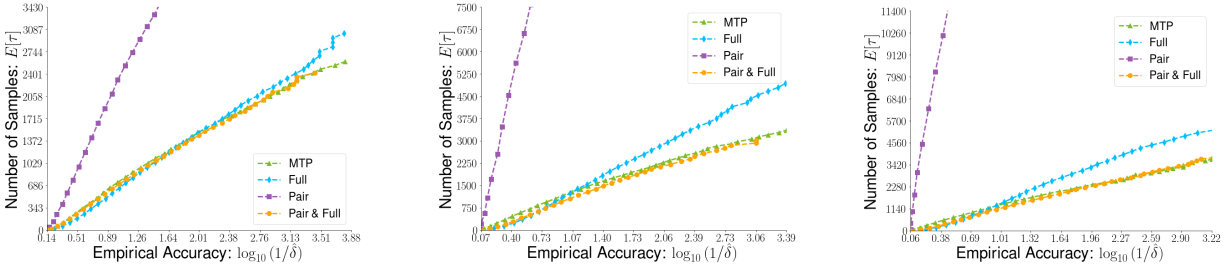
trade-off between speed (i.e., being able to stop early) and accuracy (i.e., achieving low error probability) under the specific preference model f_*.

The performance of the four policies {M, F, P, PF} under the OA model is illustrated in Figures 2.5 and 2.6. Our computational experiments suggest that policy P is significantly



(a) $K = 4$ (b) $K = 10$ (c) $K = 15$

Figure 2.5: **Sample complexity vs. theoretically guaranteed error probability (log scale).**



(a) $K = 4$ (b) $K = 10$ (c) $K = 15$

Figure 2.6: **Sample complexity vs. empirical error probability (log scale).**

outperformed by policy F which, in turn, is significantly outperformed by policies PF and M. Furthermore, the performance gap among these policies increases with K and $1/\delta$, which is consistent with our theoretical analysis. It is interesting to note that the empirical errors of PF and M are comparable, suggesting that the PF policy might be the right policy to use in practice as it offers a good compromise between performance and implementation simplicity. This echoes our previous discussion on the need for an optimal policy to balance the trade-off between accuracy and coverage. Our numerical results suggest that only displaying pairs and the full set is close to an optimal policy, if one selects the randomization probabilities

carefully (that is; according to equation (2.19)).

Benchmarks from the Ranking and Selection Literature: We conclude our numerical experiments by comparing the performance of the MTP policy to three alternative policies that have been proposed in the literature on ranking and selection under noisy pairwise comparisons. Specifically, we consider policies AR and AR2 proposed by Heckel et al. (2019) and policy PLPAC proposed by Szörényi et al. (2015). Since these policies were developed in the context of pairwise comparisons, we also include policy P in our numerical experiments. Also, the parameters of the different algorithms were selected so that all policies are δ-accurate[10] with $\delta = 0.01$ for the different problem instances (K, p) that we consider.

(K,p)	M ($\times 10^4$)	P ($\times 10^4$)	AR ($\times 10^4$)	AR2 ($\times 10^4$)	PLPAC ($\times 10^4$)
$(4, 0.5)$	0.0048	0.014	0.235	0.162	0.261
$(10, 0.5)$	0.0134	0.125	2.49	1.77	9.95
$(15, 0.5)$	0.0207	0.302	6.29	4.57	1.68

(K,p)	M ($\times 10^4$)	P ($\times 10^4$)	AR ($\times 10^4$)	AR2 ($\times 10^4$)	PLPAC ($\times 10^4$)
$(4, 0.9)$	0.231	0.613	9.67	6.73	14.4
$(10, 0.9)$	0.712	5.13	101	73.2	53.2
$(15, 0.9)$	1.13	12.5	259	188	88.9

Table 2.3: Sample complexity comparisons for $\delta = 0.01$.

Table 2.3 presents a summary of our numerical study. We find that policy M uses on average an order of magnitude less samples than policy P which, in turn, uses an order of magnitude less samples than the AR, AR2, and PLPAC policies. These results show promise for the MTP methodology, but we also need to take them 'with a grain of salt'. The reason is that policies AR, AR2 and PLPAC were developed without imposing any separability requirement on the underlying choice model and so they do not explicitly take advantage –as the M and P policies do– of the fact that preferences are ranking based or that they belong to the set $\mathcal{M}_p$. Nevertheless, our results shed some light into the question of how much

10. PLPAC need the additional assumption that the properties of the Bradley-Terry-Luce model holds.

one can gain from incorporating additional structure (e.g., parametric and/or separability assumptions) into the model. While the work of Heckel et al. (2019) reveals that imposing specific type of parametric assumptions (e.g., consumers' choice preferences belong the popular class of Bradley-Terry-Luce models) offers limited benefits for stochastic comparisons, our numerical results suggest that imposing some mild structure (as in Definition 1) goes a long way in reducing sample complexity. Furthermore, from the result in Proposition 4, the sample complexity of policy P satisfies

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_f^P[\tau]}{\log(1/\delta)} = \frac{(1+p)(K-1)}{(1-p)\log(1/p)} \quad \text{for any } f \in \mathcal{M}_p^{\text{OA}},$$

while the sample complexity of the AR and AR2 policies satisfy (see Theorem 1b in Heckel et al., 2019 and Theorem 1 in Jamieson et al., 2015)

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_f^\pi[\tau]}{\log(1/\delta)} \geq \frac{(K-1)^2(1+p)^2}{(1-p)^2} \sum_{i=2}^{K} \frac{1}{(i-1)^2}, \quad \forall f \in \mathcal{M}_p^{\text{OA}} \text{ and } \pi \in \{\text{AR,AR2}\}.$$

So for a given p, the P policy has a asymptotic sample complexity $O(K)$ while the AR and AR2 have asymptotic sample complexity $\Omega(K^2)$.[11]

2.9 Concluding Remarks and Further Directions

In this paper, we have studied a class of ranking and selection problems faced by a company that wants to identify the most preferred product out of a finite set of alternatives when consumer preferences are *a priori* unknown. Specifically, we have assumed that the only information available is that consumer preferences belong to a class $\mathcal{M}_p$ that satisfies two key properties: (*i*) preferences are ranking-based (i.e., choice probabilities are consistent with some unknown true ranking of the alternatives) and (*ii*) choice probabilities satisfy a mild

11. Note that this observation does not contradict the results in Heckel et al. (2019) and Jamieson et al. (2015), because they need to identify a δ-accurate algorithm over preference models where the consistency assumption (A-3) and the separability assumption (A-4) may be violated. Our finding does not violate Theorem 2a in Heckel et al. (2019) either because under their parametric structure, the lower bound for the OA model is vacuous: Φ is a step function, and therefore, $\phi_{\min} = 0$, $\phi_{\max} = \infty$, and $c_{\text{par}} = 0$ in their notations.

separability condition under which no two products are equally preferred (i.e., consumers have strict preferences over the different alternatives). To learn the unknown ranking over the products, we have assumed that the company is able to sample consumer preferences by sequentially showing different subsets of products to different consumers and asking them to report their top preference within the display set they were offered. In this setting, we have formulated the problem as one of designing a display policy that minimizes the expected number of samples needed to identify the top-ranked product for a given error probability $\delta \in (0, 1)$.

Because of the minimal assumptions imposed on consumer preferences, we proposed a *robust learning* methodology to derive a display policy, our *Myopic Tracking Policy* (MTP), that is worst-case asymptotically optimal as $\delta \downarrow 0$. This means that for any other policy π there exists an error probability δ and a consumer's preference in $\mathcal{M}_p$ under which the expected number of samples needed to identify the top-ranked version using MTP is less than or equal to the expected number of samples needed by policy π. Besides this theoretical performance guarantee, the Myopic Tracking policy also shows good non-asymptotic numerical performance. Through a set of computational experiments, we showed that the Myopic Tracking policy has consistently better sample performance (i.e., learns faster for a given level of accuracy) when compared to various alternative policies. Our numerical experiments also show that the Full & Pair Display policy –a variation of the MTP policy in which only pairs of the full set are displayed– offers a good compromise between performance and simplicity, which makes it particularly appealing from a practical standpoint.

In terms of future work, we envision a few directions in which our results can be extended. First, in the context of the top-ranked identification problem that we have considered in this paper, one could explore the possibility of relaxing some of the requirements that we have imposed on the set $\mathcal{M}_p$ to consider a larger set $\widetilde{\mathcal{M}}_p$ of admissible preferences. In particular, rather than requiring that consumers have strict preferences over the entire menu of products, we could only require strict preference for the top-ranked product, allowing for indifference

among the rest of the products. This can be accomplished by replacing conditions (A-3) and (A-4) in Definition 1 by the following weaker condition:

(A-5) For every $f \in \widetilde{\mathcal{M}}_p$ there exists $X_f \in [K]$ such that $p\, f(X_f|S) \geq f(X'|S)$ for all $S \ni X_f$.

Extending our robust framework to this larger set $\widetilde{\mathcal{M}}_p$ of consumer preference requires a number changes in our methodology. We anticipate that the most challenging adjustment would be to find the set of hardest-to-learn preferences and adapting our analysis accordingly. We believe, however, that our proposed Myopic Tracking policy will still perform well in this more general setting. Although a formal support of this claim is beyond the scope of this paper, we have conducted a set of numerical experiments using a real data set for which the strict ranking assumption does not hold and yet the Myopic Tracking policy still performs well when compared to other benchmark policies. The data comes from a survey conducted at the AGH University of Science and Technology in which students were asked to provide a rank ordering over a set of courses with no missing elements (further details of the dataset and the numerical experiments are provided in Appendix A.10).

A second direction that we believe is worth exploring is to extend our results on instance-specific sample optimality. The analysis in this paper gives us a good understanding on how to achieve instance-specific optimality if we restrict the set of admissible preferences to any arbitrarily finite subset of $\mathcal{M}_p$ (Theorem 2), as well as how to achieve worst-case optimality for the whole of $\mathcal{M}_p$ (Theorem 5). We conjecture that a Max-Min-type problem is still central in developing an asymptotically optimal algorithm.

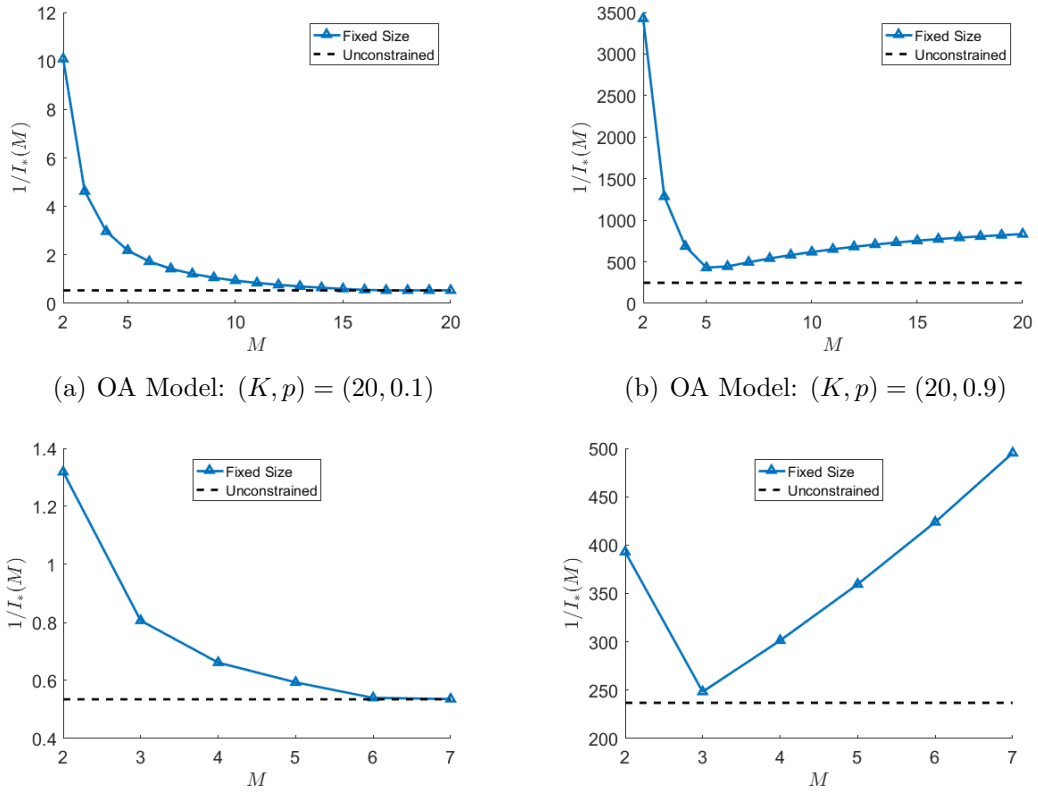
Another direction in which our paper can be extended relates to how to push our analysis for moderately-sized δ and to derive a policy that is "higher-order" optimal than the Myopic Tracking Policy. In fact, Theorem 6 implies that under the MTP policy, $\mathbb{E}_f^{\tilde{\pi}}[\tau] \leq \frac{\log(1/\delta)}{I_*^{\mathrm{OA}}} + o(\log \frac{1}{\delta})$ for all $f \in \mathcal{M}_p$. We conjecture that there exists an algorithm $\tilde{\pi}$ such that $\mathbb{E}_f^{\tilde{\pi}}[\tau] \leq \frac{\log(1/\delta)}{I_*^{\mathrm{OA}}} + O(1)$ for all $f \in \mathcal{M}_p$, where the improvement is in the residual term from $o(\log \frac{1}{\delta})$

to $O(1)$ as a function of δ.

Additionally, using the current framework of analysis, there is a potential that the Myopic Tracking Policy can be generalized to a broader class of problem formulations. For example, we could consider (i) a constrained set $\mathcal{S} \subseteq \mathcal{P}([K])$ of possible display sets; (ii) other problem objectives such as top-k selection, or the full ranking identification problem; (iii) a general class of probabilistic choice model $f(\cdot|S)$ that goes beyond p-separability; or even (iv) a more general feedback mechanism such as asking consumers to provide full rankings rather than single choices. Of course, with these extensions, the strategies (i.e. the randomization distribution) will change as the Max-Min problem changes. So will the MLE solution change if we change the probabilistic model $f(\cdot|S)$. We are interested understanding the structural properties of MTP under these different formulations. For instance, we conjecture that the randomization distribution, as a result of the new Max-Min problem, is sparse in general.

As mentioned above, our computational experiments in Section 2.8 revealed that the simple Full & Pair Display policy can achieve good perform. This raises the question of whether there are other simple policies that could also have good performance. This is a particularly important issue in many practical settings in which companies are restricted, or must commit, to display sets of a given fixed cardinality (see Vinayak and Hassibi, 2016 for a discussion comparing display sets of cardinality two and three). It would be interesting to extend our methodology to identify an optimal policy within the class of strategy that use display sets of a given size. To illustrate this point, suppose the company is restricted to offer display sets of fixed size M and let $I_*(M)$ denote Chernoff's information measure subject to this additional cardinality constraint. Figure 2.7 depicts the value of $1/I_*(M)$ (solid line) and the unconstrained value of $1/I_*$ (dashed line) as a function of M for different values of K and p and two preference models: OA model (top panels) and Mallows model (bottom panels).

These preliminary results suggests that, for low values of p, a display policy that maximizes the cardinality of the display set could be optimal among those policies with fixed



(a) OA Model: $(K, p) = (20, 0.1)$

(b) OA Model: $(K, p) = (20, 0.9)$

(c) Mallows Model: $(K, p) = (7, 0.1)$

(d) Mallows Model: $(K, p) = (7, 0.9)$

Figure 2.7: **Chernoff's inverse information measure $1/I_*(M)$ as a function of the display set cardinality M.** Different values of K and p are taken and two preference models are chosen: OA model (top panels) and Mallows model (bottom panel).

cardinality. Furthermore, Chernoff's inverse information measure for the full display policy is almost identical to the unconstrained value. On the hand, when p is large is not true that a policy that maximizes the cardinality is better. In the example above, for the OA model with $(K, p) = (20, 0.9)$ the optimal fixed size display set is $M^* = 5$ and for the Mallows model with $(K = 7, p = 0.9)$ the optimal fixed size display set is $M^* = 3$. In general, it would be interesting to understand how M^* changes as function of K and p as well as what is the optimality of gap (measured by the difference $1/I_*(M^*) - 1/I_*$) resulting from using a fixed size display strategy.

Finally, we also see room for extending the optimization and computational methods used in the implementation of the MTP policy. For instance, one can think of developing

more specialized methods that take full advantage of the structure of the problem to speed up the solution time of the MLE problem in Steps 1 and 2 of Algorithm 1. Similarly, one could consider variations of the MTP policy in which the MLE problem is not solved in every iteration but at a less frequent rate (e.g., at a monotonically increasing sequence of (possibly random) time epochs $1 \leq \tau_1 < \tau_2 < \cdots$). In theory, by carefully selecting the values of $\{\tau_i\}$, one could increase the speed of the algorithm without a significant decay in performance.

CHAPTER 3

DYNAMIC LEARNING AND MARKET MAKING IN SPREAD BETTING WITH INFORMED BETTORS

3.1 Introduction

3.1.1 Background and Overview

Spread betting markets are a prevalent form of prediction market, where market makers quote cutoff lines (a.k.a. "spread lines") for the outcome of an uncertain future event, and participants take sides on whether the outcome will exceed the spread line. As a salient example, in a point-spread market for sports betting, a bookie (the market maker) sets a sequence of point spreads (the spread lines), which can be interpreted as the number of points taken from the favorite side. The players (bettors) then observe and decide whether to wager on either the favorite side winning with a margin larger than the point spread or the favorite not winning with such a margin.[1]

Mostly popular in sports betting, spread betting constitutes a multibillion-dollar industry in the U.S.; see NGISC (1999), Statista (2018) and Schwartz (2018). For example, Schwartz (2018) reports that the amount of wagers placed in 2017 in the Nevada regulated sports betting market was worth around $4.9 billion. Due to the Supreme Court's recent decision to clear the way for states to legalize sports betting (NYTimes, 2019, CNN, 2018), the size of the U.S. sports betting market is expected to grow considerably in the near future (OE, 2017).

It is of great value to understand the market making problem in this context, i.e., how to set the spread lines as "prices" on the part of market makers. From a market maker's perspective, mispriced spread lines are costly, since the market makers take the opposite

1. Other popular examples of spread betting involve the total number of points in a sports game (Moskowitz, 2015) as well as the percentage difference of votes in a political election (Wolfers and Zitzewitz, 2004).

side of every bet offered.[2] The consequence of mispricing is exacerbated by the fact that professional bettors may systematically exploit the mispricing events; see Haralabos "Bob" Voulgaris as a vivid example in the National Basketball Association (NBA) betting market.[3] From a market designer's perspective, an effective spread betting market (as a prediction market) serves a role in information aggregation. More specifically, the spread line should reveal intrinsic characteristics of the event outcome distribution (at least in the long run).

Despite its value, understanding of the aforementioned market making problem is limited. Levitt (2004) provides some guidelines for a *clairvoyant* market maker, i.e., one who has perfect information about both the event outcome distribution and the systematic bias of the public.[4] However, Gandar et al. (1998) suggest that opening line biases exist in general in the NBA betting market, but the lines change relatively frequently in a way to eliminate the opening line biases over time. Such empirical evidence indicates that market makers (sportsbooks) may not necessarily know the "correct" spread lines in the beginning. Rather, Gandar et al. (1998) imply that the spread lines are influenced by the interplay between sportsbooks and the bettors: informed bettors—those who can identify the teams undervalued by the sportsbooks—are both present and influential in this market; while sportsbooks may adjust the spread lines significantly to correct their prediction errors over time.

Motivated by this empirical evidence, this paper aims to deepen the understanding of the above market making problem in the following directions:

1. In the presence of sophisticated bettors, how should a (non-omnipotent) market maker move the spread lines (dynamically) to maximize overall profits?

2. What is the overall cost of the market maker's lack of information?

2. See Levitt (2004) for several notorious examples of bookmakers suffering large losses in history.

3. Among the top gamblers in this market (or at least, among the ones who have revealed their identities), Voulgaris reportedly "routinely wagered a million dollars in a single day," with his average winning percentage being around 70% (Eden, 2013).

4. In particular, in an unbiased betting market where there is no systematic bias of the public, a clairvoyant market maker should consistently set the spread line at the median of the event outcome distribution, which equalizes the probabilities of bets on both sides.

3. Can spread lines yield market efficiency (i.e., converge to an unbiased predictor of the event outcome) in the long run?

We formulate a dynamic learning problem for the market maker (hereafter referred to as "she"). In particular, we consider an unbiased market where the market maker has a binary prior belief on the correct spread line. The market maker strives to dynamically extract information from the market. For example, too many bets on one side of the spread line may be treated as a signal of mispricing and the market maker can respond to it by moving the spread lines in the opposite direction. We study policies that respond to such market signals in a profit-maximizing way and characterize the corresponding spread line dynamics in the market.

Our model incorporates bettors with heterogeneous strategic behaviors and information levels. Specifically, we consider two types of bettors: a population of myopic bettors and an informed bettor (each hereafter referred to as "he"). The informed bettor has superior knowledge about the event outcome distribution, and can bet repeatedly and strategically to maximize his expected profit. On the other hand, myopic bettors do not exhibit the same level of strategic sophistication as the informed bettor. They form idiosyncratic estimates about the event outcome and bet according to their individual estimates in a myopic way. All bets are anonymous, i.e., the market maker can only base her spread lines on the aggregate statistics of bets rather than on each individual bettor's betting history.

To maximize profit, the market maker faces a trade-off between two goals: *learning* and *bluff-proofing*. On one hand, she needs to extract information from the market and incorporate it into her spread lines. We refer to this goal as *learning*. On the other hand, if the market maker adjusts spread lines in a particular way, the informed bettor may strictly prefer to "bluff," i.e., bet counter to his private information to exacerbate the market maker's mispricing. We refer to the market maker's goal to protect herself from bluffing as *bluff-proofing*. A good pricing policy should balance the trade-off between learning and bluff-

proofing.[5] For this purpose, we develop a policy that collects information at a judiciously selected rate. We show that our policy (i) achieves near-optimal profit performance for the market maker, and (ii) eventually yields market efficiency by pushing the spread line to the median of the event outcome distribution.

3.1.2 Summary of Results and Main Contributions

In our analysis, we first study Bayesian policies (BPs)—a popular class of pricing policies in the literature on dynamic pricing with demand learning (see, e.g., Harrison et al. (2012), Chen and Wang (2016) and references therein). Under a BP, the market maker (i) computes her posterior belief about the event outcome distribution using Bayes' rule but ignoring the informed bettor; and (ii) sets the spread line as a time-invariant function of her posterior belief. The BP family contains various well-studied policies, such as the myopic Bayesian policy that uses myopic profit maximization to set the spread line (see, e.g., Harrison et al., 2012, Chen and Wang, 2016).

We find that BPs are weak against the informed bettor. To be more precise, the informed bettor could earn a profit that is linearly growing in the number of bets (Proposition 5). In general, the market maker's regret typically grows linearly when the commission rate is small (Theorem 8), where regret is defined as the profit loss compared to the clairvoyant in Levitt (2004). We also show that the poor performance of BPs in our setting is not due to incomplete learning. When the informed bettor is absent in our setting, many simple BPs (including the myopic Bayesian policy) eventually learn the event outcome distribution and achieve good performance. Specifically, the spread line converges to the optimal one at an exponential rate, leading to a constant regret independent of the number of bets (Theorem 9). We also show the robustness of our results by considering two extensions where the informed bettor has restricted ability to place bets (Theorems 12 and 13).

5. See Huddart et al. (2001) and Back and Baruch (2004) for similar ideas of bluffing in the financial market. Chen and Wang (2016) also consider a trade-off between learning and bluff-proofing in dynamic pricing for a single strategic customer with unknown valuation.

We develop a policy framework that protects the market maker from strategic manipulation. Our solution, called the *inertial policy* (IP), is similar to BP except that IP moves the spread line at a slower rate, but at the same time makes it more costly for the informed bettor to bluff. IP is based on a different (yet parsimonious) state variable: the difference between bets on both sides of the spread line. This state variable resembles the market maker's log-likelihood ratio process, but it aggregates historical data differently, effectively discounting the statistical power of each single data point. We construct a simple instance of IP that achieves three goals simultaneously. First, the informed bettor never bluffs and bets according to a threshold strategy (Theorem 10). Second, the spread line converges to the optimal one almost surely, although at a sub-exponential rate (Theorem 11). Third, IP achieves a regret that grows logarithmically in the number of bets (Theorem 11). To gain deeper insights on our design choices for IP, we also provide a generalized analysis (Propositions 14 and 15). Our analysis implies that even if the informed bettor is absent, it is impossible to improve from logarithmic regret to bounded regret by choosing a different instance of IP under mild regularity conditions (Theorem 14). Our proof techniques for deriving these results are based on the exact analysis of a certain Markov chain we construct; this approach differs from the commonly used arguments in the antecedent dynamic learning literature.

3.1.3 Literature Review

Our work is related to three streams of literature, which are discussed in detail below.

Dynamic pricing and learning in the presence of strategic customers. The literature on dynamic pricing and demand learning is vast (see, for instance, Araman and Caldentey, 2009, Harrison et al., 2012, Keskin and Zeevi, 2014, 2016, 2018, Chen et al., 2015, Ciocan and Farias, 2014, Ferreira et al., 2017, den Boer and Keskin, 2017, 2019, Shin and Zeevi, 2017, Ban and Keskin, 2018, Keskin and Birge, 2019). Within this literature, a handful of studies focus on strategic customers; see Levina et al. (2009), Kanoria and

Nazerzadeh (2019), Devanur et al. (2014), Chen and Wang (2016), and Huang et al. (2018). Our methodologies and solution concepts are inherently connected to this literature, the closest work being that of Chen and Wang (2016), which builds on the work of Harrison et al. (2012). Like these authors, we consider a binary prior belief on the event outcome distribution in our setting. Our work further develops that of Chen and Wang (2016) in two ways. First, we consider bettors with heterogeneous strategic behaviors and information levels, whereas Chen and Wang (2016) consider a single strategic customer with unknown valuation. Second, we propose a different, yet simple policy family to defeat the informed bettor. Our policy is deterministic, and therefore it is verifiable whether the market maker deviates from the policy *ex-post*. This property is helpful to reinforce the market maker's policy.

We make two contributions to the literature discussed above. First, our paper expands the boundary of applications in this literature from online advertising and retail to prediction markets. Second, we hope that our ideas behind the construction of IP, as well as the proof techniques in characterizing its performance, will in turn motivate analogous pricing policies in other contexts.

Insider trading in financial markets. Our spread betting market model is akin to the insider trading literature in securities markets (see, e.g., Kyle, 1985, Glosten and Milgrom, 1985, Lin and Howe, 1990, Back, 1992, Back and Baruch, 2004, Caldentey and Stacchetti, 2010, and Ostrovsky, 2012 for related work on financial markets) but considers the special organization of the spread betting market.[6]

Mulitple modeling differences distinguish our paper from other studies in this literature. First, a common characterization of the market maker in this literature is a zero-profit condition, i.e., the market maker's expected profit conditional on the current filtration is zero;

6. A differentiating feature of a betting market is that there is no inherent value in the outcome for the participants. Hence participants trade primarily based on their predictions of the event outcome. Moreover, the typical contract structure in a spread betting market is different from a financial exchange market: in a spread betting market, a typical contract is specified by a spread line and a (usually fixed) commission rate, instead of the bid/ask prices.

see Kyle (1985) and Caldentey and Stacchetti (2010) for more details. Our market maker solves a dynamic learning and profit maximization problem. Therefore, a simple optimality equation may not be available.[7] Second, we impose a "wisdom of crowd" condition for myopic bettors, who bet according to idiosyncratic but unbiased signals of the event outcome. That is different from the purely noisy trading condition as in Kyle (1985). Finally, our market model is closely related to those in Glosten and Milgrom (1985) and Back and Baruch (2004). The main difference is that our model is fully sequential, i.e., we do not assume any probabilistic structure on the bet arrival process (such as Poisson arrivals). Thus the market maker cannot rely on detection of abnormalities in the arrival rate to distinguish the informed bettor from myopic ones.

Spread betting markets. Our paper further develops Levitt's model (Levitt, 2004) by considering the uninformed market maker's dynamic profit maximization problem. Ultimately, our paper contributes to a better understanding of market efficiency in spread betting markets (see also Sauer, 2005, Hausch and Ziemba, 2008). A spread betting market is statistically efficient if the spread line is an unbiased estimator of the event outcome (Lacey, 1990, Golec and Tamarkin, 1991). It is economically efficient if there are no profit-earning betting strategies (Zuber et al., 1985 and Gray and Gray, 1997). In general, the efficiency of this market may depend on multiple factors. Among these are spread line dynamics; in particular, market inefficiency tends to vanish over time (Gray and Gray, 1997, Gandar et al., 1998). Another factor is the strategic behavior of bettors; e.g., Golec and Tamarkin (1991) point out that in American football betting, the college market is more efficient than the National Football League (NFL) market, which is consistent with the fact that there are more professional bettors in the college market than in the NFL market. Market efficiency is also

7. Regarding our modeling choice of a profit maximization problem, a common justification for the zero-profit condition is competition among market makers. In the context of the spread betting markets, the presence of competition would probably drive the commission rate down to the market maker's marginal cost of a single bet. In that case, the market maker takes commission rates as given and the only way for a market maker to avoid a systematic loss is to price the spread line at the median, akin to the profit maximization problem in our paper.

influenced by bettors' misconceptions about random events like the "hot hand" (Camerer, 1989) and even corruption, as in point-shaving scandals (Wolfers, 2006). Regarding spread betting market efficiency, our paper conveys the insight that in an unbiased market, even if the market maker is initially uninformed of the event outcome and there is potential strategic manipulation by informed bettors, the market maker is still able to drive the spread line to the efficient one using an inertial policy.

By showing the effectiveness of the market maker's policy, our paper also sheds light on the question of why the spread betting market usually has a market maker instead of being organized as a pure exchange market. While Bossaerts et al. (2002) discuss this question from a market thickness perspective, our paper suggests that the manipulation of informed bettors could also contribute to this phenomenon.[8] Thus, a market maker with commitment power is needed to mitigate informed bettors' strategic manipulations.

Prediction markets. Our paper also fits into the broad theme of prediction markets and belief aggregation rules. For example, in a formulation where participants bet on the specific outcomes of an event, different wagering mechanisms are considered such as the scoring rules (see, e.g., Hanson, 2003, Hanson, 2007, Chen and Pennock, 2007, Chen and Vaughan, 2010, Ban, 2018) among others (see, e.g., Agrawal et al., 2011, Freeman et al., 2017, Freeman and Pennock, 2018). Besides the noticeable difference between the market organization in our paper and those in the common settings in this literature,[9] a unique feature of our paper is that we formulate an online learning problem for the market maker in an environment where a strategic expert can bet *multiple* times.

8. For example, consider a pure exchange market where a mixture of strategic (but uninformed) traders with a common prior and myopic traders participate. The symmetric equilibrium should be that the strategic traders submit the same bid/ask prices (based on their posterior beliefs) and adjust the prices by learning from the myopic traders. In that case, the strategic traders behave like the market maker in our paper, and the informed trader may harm the strategic traders in the same way that the informed bettor harms the market maker in our paper. In this case, the strategic bettors rationally would not enter and the market would not achieve efficiency.

9. For example, in scoring rules, participants submit their entire belief distributions over outcomes, while in spread betting, participants only give binary responses to the market maker's spread lines.

3.2 Problem Formulation

In this section, we build a sequential model for the spread betting market with a mixture of bettors with heterogeneous strategic behaviors and information levels. In this market, we formulate a learning-and-pricing problem for the market maker.

3.2.1 Universal Notations

Throughout the sequel, we use $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{Z}_+$, $\mathbb{Z}_-$, $\mathbb{N}$ and $\mathbb{Q}$ to denote the sets of real numbers, integers, strictly positive integers, strictly negative integers, natural numbers (including zero), and rational numbers, respectively. For all $x, y \in \mathbb{R}$, we use the following notation: $x \wedge y := \min\{x, y\}$, and $x \vee y := \max\{x, y\}$. In particular, $x^+ := x \vee 0$ and $x^- := -(x \wedge 0)$. For a function $f : \mathbb{R} \to \mathbb{R}$, we use f' and f'' to denote the derivative and second derivative functions of f, respectively.

3.2.2 Spread Betting Market

Organization of the market. We consider a betting market for a specific event happening in the future. We model the event outcome as a continuous random variable $X \in \mathbb{R}$ with cumulative distribution function (c.d.f.) $F(\cdot)$. Anonymous bets are placed sequentially, and we index them by $t \in \mathbb{Z}_+$. (We indistinguishably use the index t to refer to period t or bet t.) For bet t, the market maker first quotes a spread line s_t chosen from a compact interval $\mathcal{S} := [s_L, s_H]$, where $-\infty < s_L < s_H < \infty$. The bettor who bets in period t can then bet on either the event $\{X > s_t\}$ (in which case, we denote the bet as $d_t = +1$, or a "positive bet") or the event $\{X < s_t\}$ (in which case, we denote the bet as $d_t = -1$, or a "negative bet"). The payment to the bettor is made after the event outcome is realized. Letting $c \in (0, 1)$ denote the commission rate charged by the market maker, the normalized net payment to the bettor is $1 - c$ if the bettor wins (i.e., $(X - s_t)\, d_t > 0$), and -1 if the bettor loses (i.e., $(X - s_t)\, d_t < 0$).

Uncertain event outcome. The event outcome distribution is of either "high type" or "low type." To be precise, in our model, $X = m + \epsilon$, where: $m \in \{m_0, m_1\}$ is the (unknown) median of $F(\cdot)$, ϵ is the noise term with c.d.f. $F_\epsilon(\cdot)$, and $s_L < m_0 < m_1 < s_H$. We propose regularization conditions for $F_\epsilon(\cdot)$ in Assumption 1 below. We introduce *hypothesis* i, H_i, to be the hypothesis that $m = m_i$. We also denote by $F_i(\cdot)$ the event outcome distribution under H_i. We illustrate the event outcome distributions in Figure 3.1.

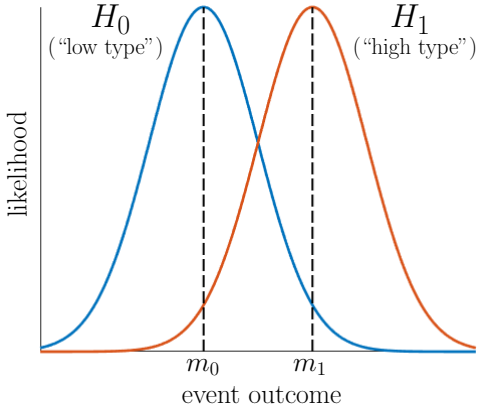


Figure 3.1: **Illustration of the event outcome distributions.** The bell-shaped curve on the left displays the probability density function of $F_0(\cdot)$ (i.e., the event outcome distribution is of "low type"). The bell-shaped curve on the right displays the probability density function of $F_1(\cdot)$ (i.e., the event outcome distribution is of "high type"). For this graph, $m_0 = 0$, $m_1 = 1$, and $\epsilon \sim \text{Normal}(0, 1)$.

Myopic bettors. There is a population of myopic bettors who bet according to idiosyncratic signals of the event. As a whole, they represent the market's (unbiased) knowledge about the event outcome X. In our model, the myopic bettor who bets in period t receives a signal x_t, which is independently drawn from the true event outcome distribution $F(\cdot)$. He bets on the side $X > s_t$ if and only if $x_t > s_t$. That is, his bet $\vartheta_t := \mathbb{I}\{x_t > s_t\} - \mathbb{I}\{x_t \leq s_t\}$ follows a binary distribution that equals $+1$ with probability $\bar{F}_i(s_t)$ and -1 with probability $F_i(s_t)$.

Informed bettor. There is an informed bettor who knows the correct median m and bets in a strategic way. For brevity, if the median $m = m_i$, then the informed bettor is referred to as the *type-i informed bettor*. An admissible strategy ξ_i of the type-i informed bettor specifies a (possibly randomized) action $a_t \in \mathcal{A} := \{-1, 0, +1\}$, which can depend

on both the market maker's policy π and the underlying hypothesis H_i. Here, actions $a_t = +1$ and $a_t = -1$ correspond to placing the bets $d_t = +1$ and $d_t = -1$, respectively, and $a_t = 0$ corresponds to a "waiting" action. That is, if $a_t = 0$, then bet t is placed by a myopic bettor. The action a_t is adapted to (i) the (public) transaction history $h_{t-1} := (s_1, d_1, \ldots, s_{t-1}, d_{t-1})$, (ii) the informed bettor's action history $A_{t-1} := (a_1, \ldots, a_{t-1})$, and (iii) the most recent spread line s_t. Accordingly, bet t can be expressed as $d_t = \mathbb{I}\{a_t \neq 0\}a_t + \mathbb{I}\{a_t = 0\}\vartheta_t$.

Market maker's policy. An admissible policy for the market maker is any function π that maps the public transaction history h_{t-1} to the spread line $s_t \in \mathcal{S}$. To represent the market maker's knowledge, the mapping π takes neither $\{a_t\}$ nor $\{\vartheta_t\}$ as arguments. This means that the market maker neither knows nor observes whether a bet is placed by the informed bettor or a myopic one. Moreover, her pricing function can depend only on the problem input parameter $\Xi := (c, m_0, m_1, F_\epsilon)$, but neither on the correct median m nor the total number of bets T. In our model, the market maker picks the policy π in the beginning and commits to her policy afterward. To make the commitment credible, we also require that her pricing function can be verified ex-post by the market. For example, this condition holds when the market maker's spread line decision is a deterministic function of the bet history.

Informed bettor's decision problem as a best response strategy. The informed bettor seeks to maximize his total expected profit, given his private information about m, as well as the public knowledge of the market maker's pricing policy π.[10] We introduce some notation to describe the informed bettor's decision making problem. Given the event outcome distribution $F_i(\cdot)$ and the market maker's spread line s, we denote the type-i informed

10. This assumption implies the market maker is able to credibly commit to a policy. In our case, bettors can verify that the market maker is following the policy, making the commitment assumption reasonable.

bettor's expected profit from a single positive and negative bet as

$$\begin{cases} j_i^+(s) := (1-c)\mathbb{P}_i(X > s) + (-1)\mathbb{P}_i(X < s) = (c-2)F_i(s) + 1 - c, \\ j_i^-(s) := (1-c)\mathbb{P}_i(X < s) + (-1)\mathbb{P}_i(X > s) = (2-c)F_i(s) - 1, \end{cases} \tag{3.1}$$

respectively, where the probabilities are taken over the event outcome X. Throughout the sequel, we refer to the quantities in (3.1) above as the informed bettor's one-stage profit function. Furthermore, given the market maker's policy π and informed bettor's response strategy ξ, the informed bettor's T-period expected profit function under H_i is

$$V_i^{\pi,\xi}(T) := \mathbb{E}_i^{\pi,\xi}\left[\sum_{t=1}^{T}\left[j_i^+(s_t)\,\mathbb{I}\{a_t = +1\} + j_i^-(s_t)\,\mathbb{I}\{a_t = -1\}\right]\right], \tag{3.2}$$

where the expectation is taken over the spread lines $\{s_t\}$ and the informed bettor's (possibly randomized) actions $\{a_t\}$, which are specified by the market maker's policy π, informed bettor's response strategy ξ, and the underlying hypothesis H_i. Given π and H_i, the informed bettor aims to find a policy ξ that maximizes his total expected profit, and in case the informed bettor's profit becomes unbounded, the informed bettor chooses a policy that maximizes his long-run average profit per bet. Formally speaking, we say that policy ξ_i^* is a best response strategy if it satisfies the following condition:

$$\xi_i^* \in \begin{cases} \arg\max_\xi \liminf_{T\to\infty}\left\{V_i^{\pi,\xi}(T)\right\} & \text{if } \sup_\xi \liminf_{T\to\infty}\left\{V_i^{\pi,\xi}(T)\right\} < \infty, \\ \arg\max_\xi \liminf_{T\to\infty}\left\{\frac{1}{T}V_i^{\pi,\xi}(T)\right\} & \text{if } \sup_\xi \liminf_{T\to\infty}\left\{V_i^{\pi,\xi}(T)\right\} = \infty. \end{cases} \tag{3.3}$$

We consider an undiscounted formulation because common spread betting markets, including sports betting, typically have frequent bets within short deadlines (Moskowitz, 2015).

3.2.3 Market Maker's Decision Problem

The market maker's goal is to choose a policy π to maximize her T-period profit, given by

$$\sum_{t=1}^{T} \mathbb{E}_i^{\pi,\xi}\big[\underbrace{\mathbb{I}\{(X-s_t)\,d_t<0\}}_{\text{bettor loses}} - (1-c)\underbrace{\mathbb{I}\{(X-s_t)\,d_t>0\}}_{\text{bettor wins}}\big],$$

where the expectation is taken over the history h_T generated by strategy profile (π,ξ). To evaluate a given policy π, we use as a performance metric the market maker's *regret*, which is the profit loss of π relative to a clairvoyant market maker who knows the underlying event outcome distribution $F(\cdot)$. The clairvoyant's optimal policy is to consistently set the spread line at the median of $F(\cdot)$. To see this, let us first consider a myopic bettor. Let $r_i(s)$ be the market maker's expected profit from a myopic bettor under hypothesis $i \in \{0,1\}$ when the spread line is s; i.e.,

$$r_i(s) := \underbrace{\mathbb{P}_i(X>s)}_{\text{prob. of positive bet}} \underbrace{[\mathbb{P}_i(X<s)+(c-1)\mathbb{P}_i(X>s)]}_{\text{market maker's profit}} + \underbrace{\mathbb{P}_i(X<s)}_{\text{prob. of negative bet}} \underbrace{[\mathbb{P}_i(X>s)+(c-1)\mathbb{P}_i(X<s)]}_{\text{market maker's profit}} \tag{3.4}$$

$$= (2c-4)\left(F_i(s)-\tfrac{1}{2}\right)^2 + \tfrac{c}{2}. \tag{3.5}$$

According to Equation (3.4), the clairvoyant earns the optimal expected profit $\frac{c}{2}$ if and only if $F_i(s)=\frac{1}{2}$. Meanwhile, the same pricing policy drives the informed bettor out of the market because of the commission cost c. That is, if $F_i(s)=\frac{1}{2}$, the informed bettor's profit from betting, $j_i^+(s)=j_i^-(s)=-\frac{c}{2}$, is strictly negative; hence, the informed bettor is incentivized to refrain from betting. For every $i\in\{0,1\}$ and $T\in\mathbb{Z}_+$, the market maker's regret is her T-period profit loss relative to the clairvoyant under hypothesis H_i; that is,

$$\Delta_i^{\pi,\xi}(T) := \tfrac{cT}{2} - \sum_{t=1}^{T} \mathbb{E}_i^{\pi,\xi}\big[\mathbb{I}\{(X-s_t)\,d_t<0\} - (1-c)\mathbb{I}\{(X-s_t)\,d_t>0\}\big]. \tag{3.6}$$

We also let $\Delta^{\pi}(T) := \max\left\{\Delta_0^{\pi,\xi_0^*}(T), \Delta_1^{\pi,\xi_1^*}(T)\right\}$ be the worst-case regret of policy π. Throughout the sequel, we study how $\Delta^{\pi}(T)$ increases in T under certain classes of policies. Specifically, we use the following Big O notation in our asymptotic performance evaluations.

Definition 3. *For every pair of functions $f(\cdot), g(\cdot) : \mathbb{Z}_+ \to \mathbb{R}$, we say that:*

- *$f(T) = O(g(T))$ if there exists $M < \infty$ and $T_0 \in \mathbb{Z}_+$ such that $|f(T)| \leq Mg(T)$ for all $T \geq T_0$;*
- *$f(T) = \Omega(g(T))$ if there exists $\delta > 0$ and $T_0 \in \mathbb{Z}_+$ such that $f(T) \geq \delta|g(T)|$ for all $T \geq T_0$;*
- *$f(T) = \Theta(g(T))$ if both $f(T) = O(g(T))$ and $f(T) = \Omega(g(T))$.*

3.2.4 Assumptions and Discussions

Assumption 1. *The noise distribution $F_\epsilon(\cdot)$ satisfies the following properties:*

(A1:1) $F_\epsilon(\cdot)$ has a continuously differentiable probability density function (p.d.f.) $f_\epsilon(\cdot)$.

(A1:2) $f_\epsilon(\cdot)$ is symmetric around zero; i.e., $f_\epsilon(x) = f_\epsilon(-x)$ for all $x \in \mathbb{R}$.

(A1:3) $f_\epsilon(x) > 0$ for all x satisfying $|x| \leq \max\{s_H - m_0, m_1 - s_L\}$.

Statement (A1:1) implies that, for each $i \in \{0, 1\}$, $F_i(\cdot)$ has a smooth p.d.f., which we denote as $f_i(\cdot)$. This continuous distribution assumption is a reasonable approximation when the event outcome is continuous (e.g., majority vote percentages) or when the number of possible event outcomes is large (e.g., basketball games). This modeling choice grants us greater analytical tractability to provide insights into the market maker's problem. Statement (A1:2) implies that the noise term ϵ is symmetrically distributed around zero. In particular, ϵ has zero mean. This symmetric distribution assumption is supported by empirical tests on sports betting for American football games (Stern, 1991) as well as basketball and baseball games

(Stern, 1994). The last assumption, Statement (A1:3), has two implications: (i) both $f_0(\cdot)$ and $f_1(\cdot)$ are strictly positive on the feasible region $\mathcal{S} = [s_L, s_H]$, and (ii) $F_1(s_L) > 0$ and $F_0(s_H) < 1$. The first implication ensures that $F_0(\cdot)$ and $F_1(\cdot)$ are separable, while the second implication rules out the degenerate case of instant learning. For a more detailed discussion on Assumption 1, we refer readers to Appendix B.1.

Informed bettor. Besides superior information about $F(\cdot)$, there are several implicit assumptions about the informed bettor in our model. First, we treat each bet as anonymous. The reason is that the informed bettor may find agencies or proxies to place the bet for him. As a consequence, the market maker cannot differentiate the informed bettor from myopic ones based on their identities. This anonymity accommodates market manipulation in the spirit of Allen and Gale (1992). Second, the informed bettor can repeatedly bet without budget constraints and can also bet an arbitrarily large amount of money before any myopic bettor (while maintaining anonymity). Since bettors make profits at the cost of the market maker, our assumption of a powerful, informed bettor imposes a "stress test" for the market maker's pricing policy. Specifically, the bet arrival model of our paper can be viewed as a combination of an adversarial model (when the informed bettor bets) and a stochastic model (when the myopic bettors bet), where the informed bettor has the power to choose between these two models to maximize his profit. In practice, we believe that such a "stress test" is relevant because the sizes of game-specific sports betting markets are relatively small. In Section 3.4, we propose a pricing policy, IP, that passes this stress test. That is, it defeats the informed bettor by allowing him to extract at most a constant profit (Lemmas 1 and 2). In Section 3.5, we also consider more restricted versions of the informed bettor to understand better the robustness of our findings.

In connection to the broader literature on stochastic bandit problems with adversarial opponents, it is worth noting that we show the existence of a pricing policy whose performance does not deteriorate infinitely in the opponent's budget.[11] In part, this is due to the

11. Apart from imposing exogenous budget constraints (Lykouris et al., 2018, Jun et al., 2018), there are

informed bettor's outside option of withdrawing the bets. Recall that the profit benchmark for the market maker is $\frac{cT}{2}$. If there is a bluffing strategy under which the market maker's profit is $\frac{cT}{4}$ and the informed bettor's loss is $\frac{cT}{4}$, the informed bettor would prefer to quit (with a profit of 0) even though he could have caused a regret of $\Omega(T)$ for the market maker.

Market maker's decision problem. There is some debate over whether market makers maximize expected profit, or minimize risk by setting the spread line to balance the wagers on both sides (Paul and Weinbach, 2012). In our model, since the market is unbiased, these two spread lines are equal, and there is no ambiguity regarding what the "ideal" spread line is for the market maker. In a biased market, it is still possible to study the market making problem under a profit maximization framework. For example, if the systematic bias is known to the market maker (or can be empirically estimated), we can generalize our current formulation by adding a bias term to the market maker's objective function.

3.3 Failure (and Success) of Bayesian Policies

This section studies Bayesian policies, a fairly general class of simple and intuitive policies for the market maker. A Bayesian policy (BP) consists of two components: a belief state and a pricing function. Under such a policy, the market maker updates her belief about the (unknown) event outcome distribution, but assuming that there is no informed bettor. To be more precise, we denote b_t as the market maker's posterior probability that $m = m_1$ in period t. The market maker's spread line depends exclusively on her belief state through a time-invariant pricing function $s^{\pi_B}(\cdot)$; see Algorithm 4 in Appendix B.2 for details. We find it convenient to equivalently express the market maker's belief state b_t using the log-likelihood ratio between F_1 and F_0; i.e., $b_{t+1} = \frac{b_1}{b_1+(1-b_1)\exp(-L_t)}$, where

$$L_t = \sum_{\ell=1}^{t} \left[\mathbb{I}\{d_\ell = 1\} \log\left(\frac{\bar{F}_1(s_\ell)}{\bar{F}_0(s_\ell)}\right) + \mathbb{I}\{d_\ell = -1\} \log\left(\frac{F_1(s_\ell)}{F_0(s_\ell)}\right) \right]. \tag{3.7}$$

other ways to restrict the opponent that we do not require in our analysis, such as focusing on oblivious strategies (Slivkins, 2019) and restricting the information structure (Jun et al., 2018).

Note that L_t is a linear aggregation of the betting sequence $\{d_t\}$ with weights $\left\{\log \frac{\bar{F}_1(s_t)}{\bar{F}_0(s_t)}\right\}$ and $\log\left\{\frac{F_1(s_t)}{F_0(s_t)}\right\}$. To avoid pathological cases, we restrict our analysis to the case where both $s^{\pi_B}(0+) := \lim_{b\downarrow 0} s^{\pi_B}(b)$ and $s^{\pi_B}(1-) := \lim_{b\uparrow 1} s^{\pi_B}(b)$ exist; that is, the spread line s_t converges to a certain level as the market maker's belief state b_t converges to 0 or 1. This is a fairly mild condition and useful in our asymptotic analysis below.

3.3.1 Performance of Bayesian Policies

In this subsection, we evaluate the performance of Bayesian policies. We find that when the commission rate c is low, the informed bettor is able to earn a constant amount from the market maker per bet on average, and the price does not converge to the median of the event outcome distribution. For ease of notation, we let $\mathfrak{d}_t := |s_t - m_i|$ denote the distance between the spread line s_t and the correct median m_i. In particular, s_t converges to m_i if and only if $\mathfrak{d}_t$ vanishes. Our main finding in this subsection is in Theorem 8 below. The proof of this theorem is in Appendix B.3.

Theorem 8. *Suppose that the market maker uses a Bayesian policy π_B with pricing function $s^{\pi_B}(\cdot)$. Then, for every initial belief $b_1 \in (0,1)$ and sufficiently small $c > 0$, we have the following:*

(T8:1) (non-convergence) For some hypothesis $i \in \{0,1\}$, with strictly positive $\mathbb{P}_i^{\pi_B,\xi_i^*}$-probability, $\mathfrak{d}_t$ does not converge to zero.

(T8:2) (linearly growing regret) $\Delta^{\pi_B}(T) = \Omega(T)$.

Theorem 8 states that, under a BP, the spread line does not converge to the correct median, and the market maker's regret grows linearly in T.

The key step in deriving Theorem 8 is identifying profitable strategies for the informed bettor if the market maker uses a BP. Among all possible cases for $s^{\pi_B}(0+)$ and $s^{\pi_B}(1-)$, the most relevant one is perhaps when the market maker is *asymptotically myopic*, i.e.,

$s^{\pi_B}(0+) = m_0$ and $s^{\pi_B}(1-) = m_1$. In this case, the market maker learns from the bet history and asymptotically moves the spread line to one of the two possible medians as the market maker becomes almost certain about the event outcome distribution under a BP. (We also study all other cases of BPs in Appendix B.3.) We find in this case, the informed bettor may earn a constant amount of profit per bet on average, by betting on both sides proportionally. We formalize this finding in Proposition 5 below. We briefly discuss the intuition behind Proposition 5 in Section 3.3.2 and present its proof in Appendix B.3.4.

Proposition 5. *Suppose that the market maker uses a Bayesian policy π_B with a pricing function $s^{\pi_B}(\cdot)$ such that $s^{\pi_B}(0+) = m_0$ and $s^{\pi_B}(1-) = m_1$. Then, for every initial belief $b_1 \in (0,1)$, hypothesis $i \in \{0,1\}$, and sufficiently small commission rate c, the type-i informed bettor has a "bluffing" policy ξ_b satisfying the following:*

(P5:1) (belief and spread line dynamics) The posterior belief b_t converges to $(1-i)$ and the spread line s_t converges to m_{1-i} almost surely under $\mathbb{P}_i^{\pi_B,\xi_b}$.

(P5:2) (linearly growing profit of the informed bettor) $V_i^{\pi_B,\xi_b}(T) = \Omega(T)$.

Proposition 5 is the key result in characterizing the performance of BPs. It demonstrates that when the market maker learns from market signals, the informed bettor earns a systematic profit by driving the spread line away from the correct median.

3.3.2 On the Informed Bettor's Profitable Manipulation Strategy (Proposition 5)

In this subsection, we provide some intuition behind Proposition 5, i.e., how the informed bettor can obtain a linearly growing profit from the market maker when the market maker is learning. In brief, the informed bettor gains from exploiting the constant learning rate of BPs (formally defined as the strictly positive drift of the log-likelihood process). For ease of illustration, let us focus on hypothesis H_1. (The intuition for the analysis under the other hypothesis is the same.)

The informed bettor faces a trade-off between two effects. On one hand, bluffing (i.e., placing a negative bet) means betting on the losing side, which is costly. On the other hand, an honest (i.e. positive) bet pushes the market maker's belief b_t towards 1, which corrects her spread lines in the future. Our question is the following: is there a balance between bluffing and honest betting for the informed bettor in order to confuse the market maker while maintaining overall profitability? The informed bettor can make a profit by betting honestly more often than he bluffs (in the limit). To see why, suppose that $c = 0$, and observe that $j_1^+(s) = -j_1^-(s) > 0$ for every $s < m_1$. That is, the one-stage cost of bluffing is offset by the profit of honest betting. Thus in the limit where $c \to 0$ and s_t converges to some $s_\infty < m_1$, the informed bettor gains a linearly growing profit if the average fraction of honest betting is strictly larger than one half.

Using honest bets more often than bluffing, the informed bettor may still push b_t down to 0 (in the limit). To see that, first suppose that $s_t = m_0$ for all t. Then the probability of a positive bet is $\bar{F}_0(m_0) = \frac{1}{2}$ under H_0 and $\bar{F}_1(m_0) > \frac{1}{2}$ under H_1. With the fraction of positive bets exactly equal to one half, L_t diverges (linearly) in favor of H_0; that is $L_t \to -\infty$.[12] Since L_t is a linear aggregation of the bet sequence, the drift of L_t remains negative if the informed bettor perturbs the fraction of honest (i.e., positive) bets by ε. That is, there exists $\varepsilon > 0$ such that $L_t \to -\infty$ (i.e., $b_t \to 0$), even if the average fraction of honest bets is $\frac{1}{2} + \varepsilon$.

We have thus identified an opportunity for the informed bettor to obtain a linearly growing profit. The informed bettor may first bet negatively consecutively to drive the market maker's posterior belief close to zero. Even though this is costly for the informed bettor, once the market maker's posterior belief is sufficiently close to zero, he gains a strictly positive net profit per average bet by (i) betting honestly with an average ratio of $\frac{1}{2} + \varepsilon$, and (ii) keeping ε sufficiently small, so as to drive the market maker's belief further closer

12. In fact, the average increment of L_t per bet is the negative of the Kullback-Leibler divergence between two Bernoulli random variables with success rates $\bar{F}_0(m_0)$ and $\bar{F}_1(m_0)$, which is strictly negative.

to zero.

3.3.3 Informed Bettor's Manipulation versus Incomplete Learning

This subsection shows that the impact of the informed bettor on the market maker's profit is distinct from the impact of incomplete learning. We do so by evaluating the performance of BPs in an environment with no informed bettors. Under a mild regularity condition, we find that a BP performs well in our setting when the informed bettor is absent. This finding follows from a separability condition regarding hypotheses H_0 and H_1.

For notational simplicity, we introduce a vacuous policy $\xi_\emptyset$ for the informed bettor, under which his action $a_t = 0$ for all t.[13] We say that a pricing function $s^{\pi_B}(\cdot)$ is *regular*, if $\max\left\{\limsup_{b\downarrow 0} \frac{|s^{\pi_B}(b)-m_0|}{b},\ \limsup_{b\uparrow 1} \frac{|s^{\pi_B}(b)-m_1|}{1-b}\right\} < \infty$. This regularity condition is stronger than than saying that $s^{\pi_B}(\cdot)$ is asymptotically myopic, i.e., $s^{\pi_B}(0+) = m_0$ and $s^{\pi_B}(1-) = m_1$, because it also requires a certain speed of convergence. However, this regularity condition is still mild. In fact, it subsumes many intuitive policies as special cases, for example, the myopic Bayesian policy, as well as the linear interpolation pricing function, $s^{\pi_B}(b) = bm_0 + (1-b)m_1$ (see Appendix B.4.3 for a more detailed discussion on the myopic Bayesian policy).

Theorem 9 below characterizes the performance of a Bayesian policy when the informed bettor is absent. We defer the proof of this theorem to Appendix B.4.

Theorem 9. *Suppose that the market maker uses a Bayesian policy π_B with a regular pricing function $s^{\pi_B}(\cdot)$. Then for every initial belief $b_1 \in (0,1)$ and hypothesis $i \in \{0,1\}$, we have the following:*

(T9:1) (convergence of spread lines) $\mathfrak{d}_t$ converges to zero almost surely under $\mathbb{P}_i^{\pi_B,\xi_\emptyset}$.

(T9:2) (exponential convergence) $\mathbb{E}_i^{\pi_B,\xi_\emptyset}[\mathfrak{d}_t] = O\big(e^{-\lambda t}\big)$ for some constant $\lambda > 0$.

13. Equivalently, we could interpret $\xi_\emptyset$ as the informed bettor's best response strategy if the commission is sufficiently high. See Appendix B.4.2 for a discussion.

(T9:3) (bounded regret) $\Delta_i^{\pi_B,\xi_\emptyset}(T) = O(1)$.

Theorem 9 implies that (statistical) incomplete learning does not happen in our context and many simple Bayesian policies exhibit remarkably good profit performance when there is no informed bettor (for the antecedent work on incomplete learning, see, e.g., McLennan, 1984, Harrison et al., 2012, Keskin and Zeevi, 2018). Thus, the informed bettor's strategic manipulation, instead of incomplete learning, is the market maker's major challenge in the context of our problem formulation. The main intuition behind Theorem 9 is that in our setting, a BP satisfies a separability condition similar to being a δ-discriminative policy as in Harrison et al. (2012) (see Lemma 18 in Appendix B.1 for more details).

3.4 Defeating the Informed Bettor with an Inertial Policy

In this section, we construct a simple policy, called the inertial policy (IP), that allows the market maker to combat the informed bettor. We find that no matter how small the commission rate is, IP is immune to the strategic manipulation of the informed bettor while guaranteeing that the market maker's regret grows at most logarithmically in the number of bets.

3.4.1 Preliminaries

Definition of the inertial policy. IP employs a univariate state variable that represents the evidence in support of $m = m_1$, as well as a pricing function. The aforementioned state variable is the difference between the number of positive and negative bets before period t, given by

$$Z_t := \sum_{\ell=1}^{t-1} d_\ell. \tag{3.8}$$

Intuitively, we may interpret the new state variable Z_t as an "unweighted" version of the log-likelihood ratio process L_t in (3.7). On one hand, both Z_t and L_t are linear aggregations

of the betting sequence $\{d_t\}$. As a result, an intuitive property that Z_t inherits from L_t is that a high value of Z_t corresponds to strong evidence in support of H_1. On the other hand, however, Z_t differs from L_t in how bet outcomes are aggregated. Given a spread line s_t, L_t gives the weight of $\log \frac{\bar{F}_1(s_t)}{\bar{F}_0(s_t)}$ to a positive bet observation and the weight of $\log \frac{F_1(s_t)}{F_0(s_t)}$ to a negative one. In comparison, Z_t weighs these observations with weights $+1$ and -1. Such difference in weights effectively adjusts the statistical power of each data point observed. As shown below, we construct Z_t so that it accounts for the informed bettor's incentives, while maintaining tractability in both policy implementation and performance evaluation.

Similar to BPs, IP specifies the market maker's spread line through a time-invariant pricing function $\tilde{s}(\cdot)$ of state variable Z_t. That is, given $Z_t = z \in \mathbb{Z}$, the market maker's spread line is $s_t = \tilde{s}(z)$.

Representing IP via residual probabilities. For notational simplicity and interpretability, we use a sequence of numbers called *residual probabilities* to provide an alternative representation of the pricing function $\tilde{s}(\cdot)$. The idea behind the residual probabilities is to represent a spread line s by the quantity $|F_i(s) - \frac{1}{2}|$, which captures how far s is from the median m_i. Letting $\alpha := F_1(m_0)$, we explain in Proposition 6 below how we can represent $\tilde{s}(\cdot)$ by residual probabilities. The proof of this result is in Appendix B.5.

Proposition 6. *(residual probability) For all $\rho(\cdot) : \mathbb{Z}_+ \to \left(0, \frac{1}{2} - \alpha\right)$, there uniquely exist a pricing function $\tilde{s}(\cdot) : \mathbb{Z} \to [m_0, m_1]$ and an extension of $\rho(\cdot)$ from $\mathbb{Z}_+$ to $\mathbb{Z}$ such that for all $z \in \mathbb{Z}$,*

$$F_0\big(\tilde{s}(-z)\big) = \tfrac{1}{2} + \rho(z) \text{ and } F_1\big(\tilde{s}(z)\big) = \tfrac{1}{2} - \rho(z). \tag{3.9}$$

The closed-form expression for $\tilde{s}(\cdot)$ is given by:

$$\tilde{s}(z) = \begin{cases} F_1^{-1}\left(\frac{1}{2} - \rho(z)\right) & \text{if } z \in \mathbb{Z}_+, \\ \frac{m_0+m_1}{2} & \text{if } z = 0, \\ F_0^{-1}\left(\frac{1}{2} + \rho(-z)\right) & \text{if } z \in \mathbb{Z}_-. \end{cases} \tag{3.10}$$

The closed-form expression for the extension of $\rho(\cdot)$ is given by (B.5.1) *in Appendix B.5.*

The function $\rho(\cdot)$ quantifies how much the policy incorporates historical information into the next spread line.[14] For example, if $Z_t \in \mathbb{Z}_+$, a small $\rho(Z_t)$ means that $\tilde{s}(Z_t)$ is close to m_1; while if $Z_t \in \mathbb{Z}_-$, a small $\rho(Z_t)$ means that $\tilde{s}(Z_t)$ is close to m_0. We illustrate the correspondence between $\tilde{s}(\cdot)$ and $\rho(\cdot)$ in Figure 3.2; see also Algorithm 5 in Appendix B.2 for details.

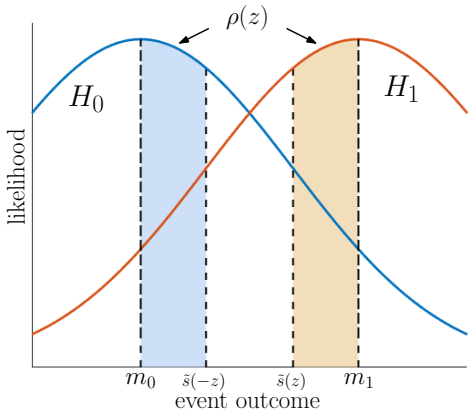


Figure 3.2: **Illustration of the residual probabilities.** For every $z \in \mathbb{Z}_+$, $\tilde{s}(z)$ and $\tilde{s}(-z)$ are chosen such that each of the two shaded regions has an area equal to $\rho(z)$. For this graph, $m_0 = 0$, $m_1 = 1$, $\epsilon \sim \text{Normal}(0, 0.7)$, $\tilde{s}(-z) = 0.3$, and $\tilde{s}(z) = 0.7$.

Constructing a candidate for the residual probability sequence. Noting that IP is broadly defined for a generic function $\rho(\cdot)$, we propose a simple candidate for $\rho(\cdot)$. Let $\rho(z) = \frac{1}{r_0 + rz}$ for $z \in \mathbb{Z}_+$. Here, we choose $r_0 := \left[\frac{1}{2} - F_1\left(\frac{m_0+m_1}{2}\right)\right]^{-1}$ so that $\rho(0) = \frac{1}{r_0}$, where $\rho(0)$ is defined in the sense of Proposition 6. In this construction, the only tuning parameter is $r > 0$, which controls the rate of convergence of $\rho(z)$ as $z \uparrow +\infty$. A larger value of r means that $\rho(z)$ converges to 0 faster.

Discussion on the inertial policy. Our construction of IP possesses several desirable properties. First, it is simple to implement. To be specific, the update of the state variable Z_t, the evaluation of $\rho(\cdot)$, and the calculation of the spread line s_t all require direct function evaluations only.

Second, the market dynamics under IP are tractable from an analytical point of view. Observe that Z_t is a stationary Markov chain that represents the whole market. To see why,

14. Because $\rho(\cdot)$ is a function of integers, we also refer to $\rho = \{\rho(z), z \in \mathbb{Z}_+\}$ as a *residual probability sequence*.

note that IP is a stationary Markov policy that exclusively depends on the state variable Z_t. Therefore, it is sufficient to consider stationary Markov policies for the informed bettor as well. In fact, if we fix the market maker's inertial policy π_I and the informed bettor's policy ξ, then Z_t becomes a birth and death chain under $\mathbb{P}_i^{\pi_I,\xi}$.

Third, IP makes makes manipulation more difficult. Recall from Section 3.3.2 that we described a simple manipulation strategy where the informed bettor mixes bluffing and honest betting to gain a linearly growing profit from BP. It is straightforward to check that IP guards the market maker from the same manipulation strategy. For example, suppose that H_1 is correct while the spread line is near m_0. The informed bettor may push L_t to $-\infty$ by betting a $\left(\frac{1}{2}+\varepsilon\right)$ fraction of honest (i.e. positive) bets. However, the same strategy only pushes Z_t to ∞, which means that the spread line will be corrected eventually.[15] In Theorem 10, we show that IP guards the market maker against all bluffing behaviors in general.

3.4.2 Performance of the Inertial Policy

In this subsection, we quantify the performance of IP. We find that under IP, (i) the informed bettor never bluffs and bets under a threshold strategy (Theorem 10), (ii) the uninformed market maker's regret grows at most logarithmically in the number of bets T (Theorem 11), and (iii) the spread line converges to the median of the event outcome distribution with probability one (Theorem 11).

We first characterize the informed bettor's best response policy ξ_i^* as well as his total profit. Recall that the market state is encoded by Z_t defined in (3.8). With a slight abuse of notation, we specify the type-i informed bettor's optimal strategy ξ_i^* by a function of the state $z \in \mathbb{Z}$. We also introduce the value function $J^i(\cdot) : \mathbb{Z} \to \mathbb{R} \cup \{+\infty\}$ such that $J^i(z)$ is the type-i informed bettor's maximum total expected continuation profit given that the

15. Thus, no matter how small the commission rate $c > 0$ is, the informed bettor does not have an incentive to bluff, at least when the spread line is close to either of the medians m_i.

market maker uses IP and the current market state is z. In particular, $J^i(0)$ is the type-i informed bettor's best total profit, because the market starts with state $Z_1 = 0$. Note that conceptually, $J^i(z)$ is possibly infinite if the market maker's policy is not carefully designed. But IP rules out this possibility as shown in Theorem 10 below. This theorem characterizes the informed bettor's profit and best response to the market maker's inertial policy π_I. We relegate the proof of of Theorem 10 to Appendix B.7, and discuss our key proof approach in Section 3.4.3.

Theorem 10. *There exists $\bar{r} > 0$ such that for every policy parameter $r \in (0, \bar{r})$, we have the following:*

(T10:1) (informed bettor's bounded profit) For every hypothesis $i \in \{0, 1\}$ and $z \in \mathbb{Z}$, $J^i(z) < +\infty$.

(T10:2) (informed bettor's best response strategy) For every hypothesis $i \in \{0, 1\}$, the informed bettor's optimal strategy $\xi_i^*(\cdot)$ is a threshold strategy; i.e., there exists $\bar{z} \in \mathbb{Z} \cup \{-\infty\}$ such that

$$\xi_1^*(z) = \mathbb{I}\{z < \bar{z}\} \quad \text{and} \quad \xi_0^*(z) = -\mathbb{I}\{z > -\bar{z}\} \quad \text{for every } z \in \mathbb{Z}. \tag{3.11}$$

The expressions of $\bar{r}$ and $\bar{z}$ depend only on the problem input parameter Ξ and are given in Appendix B.7.2.

To interpret Theorem 10, let us say that the market state Z_t changes in the "right" direction if it increases under H_1 and decreases under H_0, and in the "wrong" direction otherwise. Theorem 10 means that under IP (with a sufficiently small r), the informed bettor will bet honestly if and only if the market state evolves sufficiently far in the wrong direction. Because the market maker's spread line s_t is a function of the market state Z_t through the pricing function $\tilde{s}(\cdot)$ defined in (3.10), it is equivalent to say that the informed bettor bets if and only if the market maker's spread line is sufficiently close to the wrong median. Otherwise, the informed bettor chooses not to bet on either side because of the

transaction cost (in the form of the market maker's commission rate). In either case, the informed bettor does not have an incentive to bluff. To illustrate our inertial policy as well as the informed bettor's best response, we show in Figure 3.3 a sample path of $\{Z_t\}$ in a numerical example. Under IP, the total profit that the informed bettor gains from the market maker is finite.

Finally, Theorem 11 below characterizes the performance of our inertial policy π_I with the residual probability sequence $\{\rho(z) = \frac{1}{r_0+rz}, z \in \mathbb{Z}_+\}$. We relegate the proof of this result to Section 3.4.4.

Theorem 11. *For every commission rate $c \in (0,1)$, hypothesis $i \in \{0,1\}$, and policy parameter $r \in (0, \bar{r})$, we have the following:*

(T11:1) (convergence of spread lines) $Z_t \to \infty$ (resp. $Z_t \to -\infty$) almost surely under $\mathbb{P}_1^{\pi_I,\xi_1^*}$ (resp. $\mathbb{P}_0^{\pi_I,\xi_0^*}$). As a result, $\mathfrak{d}_t$ converges to zero almost surely under $\mathbb{P}_i^{\pi_I,\xi_i^*}$.

(T11:2) (sub-exponential convergence) $\sum_t \mathbb{E}_i^{\pi_I,\xi_i^*}[\mathfrak{d}_t]$ diverges at a rate satisfying $\sum_{t=1}^{T} \mathbb{E}_i^{\pi_I,\xi_i^*}[\mathfrak{d}_t] = O(\sqrt{T \log T})$.

(T11:3) (logarithmic regret) $\Delta^{\pi_I}(T) = O(\log T)$.

Theorem 11 implies that IP asymptotically sets the spread line s_t at the correct median. Since the informed bettor's best response policy is of threshold type (as shown in Theorem 10), Theorem 11 also implies that IP eventually drives the informed bettor out of the market with probability one. Consequently, under IP, the market maker's T-period regret is at most in the order of $\log T$. Together with Theorem 10, Theorem 11 gives a comprehensive characterization of IP.

Noticeably, if the commission rate is sufficiently high, the informed bettor's best response strategy is to never bet at all; see Appendix B.4.2 for more details. As a result, in connection with Theorem 9, our analysis in Theorems 10 and 11 subsumes the environment with no

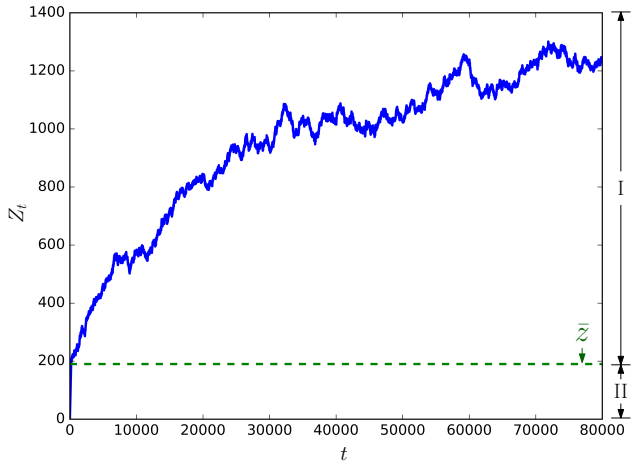


Figure 3.3: **Sample path illustration for the inertial policy under hypothesis H_1.** The solid curve displays a sample path of $\{Z_t\}$ whereas the dashed line displays the informed bettor's betting threshold $\bar{z}$. When the market state Z_t is in Region I (i.e., above the dashed line), the informed bettor is inactive and only myopic bettors bet. In comparison, when Z_t is in Region II (i.e., below the dashed line), the informed bettor actively exploits his inside information by betting honestly. In this graph, $m_0 = 0$, $m_1 = 1$, $\epsilon \sim \text{Normal}(0, 1)$, $c = 0.1$, and $r = 0.99\bar{r}$, where $\bar{r} = 0.1667$ is calculated as in Appendix B.7.2.

informed bettors as a special case. Corollary 4 below formally states this result, which characterizes the performance of IP with the residual probability sequence $\{\rho(z) = \frac{1}{r_0+rz}, z \in \mathbb{Z}_+\}$ when the informed bettor is absent.

Corollary 4. *For every hypothesis $i \in \{0, 1\}$ and policy parameter $r \in (0, \bar{r})$, we have the following:*

(C4:1) (convergence of spread lines) $Z_t \to \infty$ *(resp.* $Z_t \to -\infty$*) almost surely under* $\mathbb{P}_1^{\pi_I, \xi_\emptyset}$ *(resp.* $\mathbb{P}_0^{\pi_I, \xi_\emptyset}$*). As a result,* $\mathfrak{d}_t$ *converges to zero almost surely under* $\mathbb{P}_i^{\pi_I, \xi_\emptyset}$.

(C4:2) (sub-exponential convergence) $\sum_t \mathbb{E}_i^{\pi_I, \xi_\emptyset}[\mathfrak{d}_t]$ *diverges at a rate satisfying* $\sum_{t=1}^{T} \mathbb{E}_i^{\pi_I, \xi_\emptyset}[\mathfrak{d}_t] = O(\sqrt{T \log T})$.

(C4:3) (logarithmic regret) $\Delta_i^{\pi_I, \xi_\emptyset}(T) = O(\log T)$.

From a managerial standpoint, IP stands in stark contrast to the Bayesian policies in Section 3.3. On one hand, IP effectively protects the market maker against the informed

bettor's strategic manipulation. On the other hand, this protection is at a cost: IP learns (from myopic bettors) more slowly than BPs do. For example, suppose there are no informed bettors. Under hypothesis H_1, the drift (i.e., expected one-stage increment) of Z_t equals $2\rho(z)$, which converges to zero as z grows to infinity, while the drift of L_t is bounded away from zero. Roughly speaking, this means Z_t has a sublinear growth rate, which leads to a sub-exponential convergence rate of s_t and logarithmic regret under IP. In contrast, L_t has a linear growth rate, which leads to an exponential convergence rate of s_t and constant regret under BP. The related results are formally stated in Theorems 9 and 11 as well as Corollary 4. Moreover, we also discuss how to intuitively understand the dynamics of $\{Z_t\}$ in Section 3.4.4.

3.4.3 On the Informed Bettor's Optimal Strategy and Profit (Theorem 10)

This subsection summarizes our proof approach for Theorem 10, explaining why the informed bettor never bluffs and gains a finite total profit under IP. In a nutshell, IP pushes the market state Z_t in the right direction. Meanwhile, IP judiciously controls the growth rate of Z_t: the drift of Z_t vanishes as Z_t grows (so that the informed bettor finds it costly to bluff), but slowly (so that the informed bettor cannot gain an infinite profit from simply waiting for mispricing events).

Market state as a birth and death Markov chain. As alluded to earlier, we represent the market by Z_t, which is a birth and death Markov chain. That is, Z_t increases or decreases by one after each bet, and its transition rule is determined by the market participants' stationary Markovian policies. Based on the informed bettor's best response strategy $\xi_i^*(\cdot)$ described in Theorem 10, the transition rule of Z_t can be described by two cases. In the first case, Z_t is sufficiently far in the right direction (i.e., $Z_t \geq \bar{z}$ under hypothesis H_1 and $Z_t \leq -\bar{z}$ under hypothesis H_0), the informed bettor is inactive and only myopic bettors participate. As a result, Z_t increases (resp. decreases) with probability $\frac{1}{2}+\rho(Z_t)$ $\big($resp. $\frac{1}{2}+\rho(-Z_t)\big)$ under hypothesis H_1 (resp. hypothesis H_0). In the second case,

our informed bettor actively exploits his inside information by betting honestly. Hence Z_t moves in the right direction with probability one. As a result, the birth and death Markov chain Z_t has a reflecting boundary point $\bar{z}-1$ (resp. $-\bar{z}+1$) under hypothesis H_1 (resp. hypothesis H_0) after a finitely many of steps.

To formally describe the dynamics of Z_t, it is convenient to use the following notation:

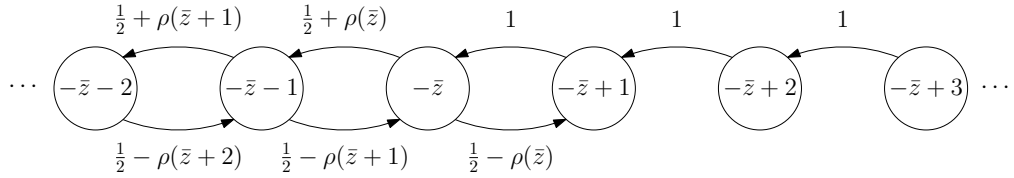
$$\mathbb{P}_i^z(\cdot) := \mathbb{P}_i^{\pi_I,\xi_i^*}(\cdot \mid Z_1 = z) \text{ and } \mathbb{E}_i^z[\cdot] := \mathbb{E}_i^{\pi_I,\xi_i^*}[\cdot \mid Z_1 = z] \text{ for all } i \in \{0,1\} \text{ and } z \in \mathbb{Z}. \tag{3.12}$$

In short, $\mathbb{P}_i^z(\cdot)$ (resp. $\mathbb{E}_i^z[\cdot]$) is a translated version of $\mathbb{P}_i^{\pi_I,\xi_i^*}$ (resp. $\mathbb{E}_i^{\pi_I,\xi_i^*}$), under which Z_t starts with z almost surely. Moreover, in the context of Theorems 10 and 11, it is clear that the market maker implements the inertial policy π_I, and the type-i informed bettor implements the threshold strategy ξ_i^*. Hence we drop the superscript (π_I, ξ_i^*) for notational brevity. In particular, since the market starts with state $Z_1 = 0$, we have $\mathbb{P}_i^0 = \mathbb{P}_i^{\pi_I,\xi_i^*}$. We explicitly write out the transition matrix $\mathcal{P}^i_{z,\breve{z}} := \mathbb{P}_i^z(Z_2 = \breve{z})$ in Equation (3.13) and illustrate the dynamics of Z_t under $\mathbb{P}_i^z$ in Figure 3.4.

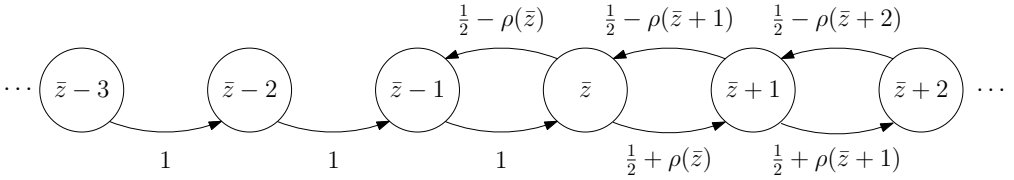
$$\begin{cases} \mathcal{P}^0_{z,\breve{z}} = 1 & \text{if } z \geq -\bar{z}+1 \text{ and } \breve{z} = z-1, \\ \mathcal{P}^0_{z,\breve{z}} = \frac{1}{2} + \rho(-z) & \text{if } z \leq -\bar{z} \text{ and } \breve{z} = z-1, \\ \mathcal{P}^0_{z,\breve{z}} = \frac{1}{2} - \rho(-z) & \text{if } z \leq -\bar{z} \text{ and } \breve{z} = z+1, \end{cases} \tag{3.13}$$

$$\text{and} \begin{cases} \mathcal{P}^1_{z,\breve{z}} = 1 & \text{if } z \leq \bar{z}-1 \text{ and } \breve{z} = z+1, \\ \mathcal{P}^1_{z,\breve{z}} = \frac{1}{2} + \rho(z) & \text{if } z \geq \bar{z} \text{ and } \breve{z} = z+1, \\ \mathcal{P}^1_{z,\breve{z}} = \frac{1}{2} - \rho(z) & \text{if } z \geq \bar{z} \text{ and } \breve{z} = z-1. \end{cases}$$

Informed bettor's one-stage profit function. We introduce a compact representation of the informed bettor's one-stage profit functions as follows. Invoking (3.1) and (3.9), we use use $j_i^+(z)$ and $j_i^-(z)$ in place of $j_i^+\big(\tilde{s}(z)\big)$ and $j_i^-\big(\tilde{s}(z)\big)$, respectively, for shorthand



(a) dynamics of Z_t under $\mathbb{P}_0^z$



(b) dynamics of Z_t under $\mathbb{P}_1^z$

Figure 3.4: **Illustration of the Markov chain $\{Z_t\}$.** The nodes represent the set of integers, $\mathbb{Z}$, as the state space. The numbers associated with the arrows display the transition probabilities.

notation throughout this section; that is,

$$\underbrace{j_1^+(z)=j_0^-(-z)}_{\text{honest betting}}=(2-c)\rho(z)-\tfrac{c}{2} \quad \text{and} \quad \underbrace{j_1^-(z)=j_0^+(-z)}_{\text{bluffing}}=(c-2)\rho(z)-\tfrac{c}{2}. \tag{3.14}$$

Proof sketch for Theorem 10. In the proof of Theorem 10, we employ a verification argument. We evaluate the informed bettor's profit under the threshold strategy ξ_i^* defined in (3.11), showing that: (i) ξ_i^* is a best response strategy, and (ii) ξ_i^* generates a finite profit for the informed bettor.

Constructing a candidate value function $\bar{J}^i(\cdot)$. Let us first define a particular function $\bar{J}^i(\cdot)$ as follows:

$$\bar{J}^1(z)=\bar{J}^0(-z)=$$

$$\begin{cases} 0 \ \text{ for all } z\in\mathbb{Z} & \text{if } \bar{z}=-\infty, \\ j_1^+(\bar{z}-1)\sum_{n=(z-\bar{z})^+}^{\infty}\Lambda_n+\sum_{i=1}^{(\bar{z}-z)^+}j_1^+(\bar{z}-i) \ \text{ for all } z\in\mathbb{Z} & \text{if } \bar{z}>-\infty. \end{cases} \tag{3.15}$$

Here, the constants $\{\Lambda_n\}$ depend only on $\bar{z}$ and $\rho(\cdot)$, and are given by

$$\Lambda_n:=\prod_{k=0}^{n}\frac{\frac{1}{2}-\rho(\bar{z}+k)}{\frac{1}{2}+\rho(\bar{z}+k)}>0. \tag{3.16}$$

The intuition behind this construction is as follows. Intuitively, $\bar{J}^1(\cdot)$ is the informed bettor's continuation profit function under the threshold strategy ξ_1^*. Drawing on the informed

bettor's one-stage profit function and the transition rule of Z_t, we expect $\bar{J}^1(\cdot)$ to satisfy the following recursive relation for $z \in \mathbb{Z}$:

$$\bar{J}^1(z) = \begin{cases} j_1^+(z) + \bar{J}^1(z+1) & \text{if } z < \bar{z}, \\ \bar{F}_1(\tilde{s}(z))\,\bar{J}^1(z+1) + F_1(\tilde{s}(z))\,\bar{J}^1(z-1) & \text{if } z \geq \bar{z}, \end{cases}$$

subject to the boundary condition $\lim_{z\to\infty} \bar{J}^1(z) = 0$. In the meanwhile, by symmetry, $\bar{J}^0(\cdot)$ should be a "reflected" version of $\bar{J}^1(\cdot)$; that is, $\bar{J}^0(z) = \bar{J}^1(-z)$ for all $z \in \mathbb{Z}$. Thus, the construction in (3.15) can be viewed as a solution to the aforementioned recursive relation.

Key properties of $\bar{J}^i(\cdot)$. The construction of $\bar{J}^i(\cdot)$ raises three questions: (i) Is $\bar{J}^i(\cdot)$ well-defined (i.e., finitely valued)? (ii) Is $\bar{J}^i(\cdot)$ indeed the informed bettor's continuation profit function under the threshold strategy ξ_i^*? (iii) Does the informed bettor have an incentive to deviate from ξ_i^*? In Lemmas 1-3 below, we give definite answers to all of the three questions. The proofs of these lemmas are in Appendix B.7.5.

Lemma 1. *(range of the value function) For all $z \in \mathbb{Z}$, we have $0 \leq \bar{J}^1(z) = \bar{J}^0(-z) < \infty$.*

Lemma 2. *(continuation profit) Let $r \in (0, \bar{r})$. Given hypothesis $i \in \{0,1\}$, the market maker's inertial policy π_I, and the informed bettor's response strategy ξ_i^*, $\bar{J}^i(\cdot)$ is the expected continuation profit function for the type-i informed bettor. That is, for all $z \in \mathbb{Z}$,*

$$\begin{aligned} \bar{J}^1(z) &= \lim_{T\to\infty} \mathbb{E}_1^z \left[\sum_{t=1}^{T} j_1^+(Z_t)\, \mathbb{I}\{Z_t < \bar{z}\} \right] \\ \text{and } \bar{J}^0(z) &= \lim_{T\to\infty} \mathbb{E}_0^z \left[\sum_{t=1}^{T} j_0^-(Z_t)\, \mathbb{I}\{Z_t > -\bar{z}\} \right]. \end{aligned} \tag{3.17}$$

Lemma 3. *(Bellman optimality) Let $r \in (0, \bar{r})$. Given hypothesis $i \in \{0,1\}$, the market maker's inertial policy π_I, and current state $z \in \mathbb{Z}$, the type-i informed bettor does not have any incentive to deviate from the action specified by ξ_i^* in* (3.11)*. That is, $\bar{J}^i(\cdot)$ satisfies the*

following Bellman equation:

$$\begin{aligned}
\bar{J}^1(z) = \max\Big\{ & \underbrace{j_1^+(z) + \bar{J}^1(z+1)}_{\text{bettor's profit from } a_t=+1}, \\
& \underbrace{j_1^-(z) + \bar{J}^1(z-1)}_{\text{bettor's profit from } a_t=-1}, \underbrace{\bar{F}_1(\tilde{s}(z))\,\bar{J}^1(z+1) + F_1(\tilde{s}(z))\,\bar{J}^1(z-1)}_{\text{bettor's profit from } a_t=0}\Big\}, \\
\bar{J}^0(z) = \max\Big\{ & \underbrace{j_0^+(z) + \bar{J}^0(z+1)}_{\text{bettor's profit from } a_t=+1}, \quad \underbrace{j_0^-(z) + \bar{J}^0(z-1)}_{\text{bettor's profit from } a_t=-1}, \\
& \underbrace{\bar{F}_0(\tilde{s}(z))\,\bar{J}^0(z+1) + F_0(\tilde{s}(z))\,\bar{J}^0(z-1)}_{\text{bettor's profit from } a_t=0}\Big\}.
\end{aligned} \tag{3.18}$$

Verification of optimality of $\bar{J}^i(\cdot)$. To summarize, Lemma 2 implies that $\bar{J}^i(\cdot)$ defined in (3.15) is the continuation profit function of the threshold strategy ξ_i^* defined in (3.11), and Lemma 1 ensures that the total profit generated by ξ_i^* is finite. The nonnegativity of $\bar{J}^i(\cdot)$ in Lemma 1 plus the Bellman optimality in Lemma 3 imply that $\bar{J}^i(\cdot)$ is an upper bound of the informed bettor's value function.[16] Lemmas 1-3 jointly establish the optimality of threshold strategy ξ_i^*, as well as the fact that $\bar{J}^i(\cdot) = J^i(\cdot)$. This verifies the optimality of $\bar{J}^i(\cdot)$. The complete proof of Theorem 10 is in Appendix B.7.

Discussion of proof methodology. Our analysis in Theorem 10 (especially Lemma 2) builds on an exact analysis of $\{Z_t\}$. To achieve tight results, we summarize our key proof step in Lemma 4 below. We relegate its proof to Appendix B.6.

Lemma 4. *(key step for performance evaluation) Consider an arbitrary stationary discrete-time Markov chain* $\{Y_t,\ t = 1, 2, \ldots\}$ *with state space* $S \subset \mathbb{R}$ *defined on some probability measure space* $(\Omega, \mathbb{P})$*. Suppose that* $u, f : S \to \mathbb{R}$ *are functions that satisfy* $f(z) = \mathbb{E}[u(Y_2)|Y_1 = z] - u(z)$ *for all* $z \in S$*. Then,* $\mathbb{E}f(Y_1) + \mathbb{E}f(Y_2) + \cdots + \mathbb{E}f(Y_t) = \mathbb{E}u(Y_{t+1}) - \mathbb{E}u(Y_1)$ *for all* t*.*

Lemma 4 gives us a guideline on how to evaluate quantities of the form $\mathbb{E}f(Y_1)+\mathbb{E}f(Y_2)+$

16. Because we consider the total profit problem for the informed bettor, the Bellman equation (3.18) in Lemma 3 alone implies neither the optimality of ξ_i^* as the informed bettor's strategy nor the optimality of $\bar{J}^i(\cdot)$ as his value function. In fact, the solution to the Bellman equation (3.18) is not even unique: if any $\tilde{J}^i(\cdot)$ solves the Bellman equation, so does $\tilde{J}^i(\cdot) + c$, where c is an arbitrary constant.

$\cdots + \mathbb{E}f(Y_t)$ for any given function $f(\cdot)$. In the first step of this evaluation, we solve for the difference equation $f(z) = \mathbb{E}[u(Y_2)|Y_1 = z] - u(z)$ to obtain the function $u(\cdot)$. In the second step, we can replace the t-period sum with only two quantities: $\mathbb{E}u(Y_{t+1})$ and $\mathbb{E}u(Y_1)$, which are usually much easier to work with. Provided that solving the difference equation is tractable, this machinery has two main advantages over the commonly used large-deviation based arguments. First, large-deviation based arguments are sometimes not easily available because the "deterministic" part of $\{Y_t\}$ does not overwhelm the "stochastic noise" part. Second, this evaluation is *exact*, and hence tighter results can be obtained from this machinery. Readers familiar with the stochastic calculus literature may view this machinery as a discrete-time analog of Dynkin's formula (Øksendal, 2003, Theorem 7.4.1), which is commonly used to estimate various random quantities via solving differential equations (Krylov, 2002, Chapter 6.10).

In the context of our paper, we face several technical challenges: the growth of Z_t suffers from vanishing drift (see Section 3.4.4 for a more detailed discussion), and we need to evaluate the informed bettor's continuation profit *exactly* to verify the Bellman optimality. That is why we use the method in Lemma 4 to overcome our challenges. More specifically, we take $f(x) = j_1^+(x)\,\mathbb{I}\{x < \bar{z}\}$ (resp. $f(x) = j_0^-(x)\,\mathbb{I}\{x > -\bar{z}\}$) to evaluate the informed bettor's continuation profit function under ξ_1^* (resp. ξ_0^*). The same machinery is also a key step in Proposition 7, where we take other forms of $f(x)$ to show (i) the almost sure convergence of spread lines, (ii) the convergence rate of spread lines, and (iii) the logarithmic growth rate of regret under IP.

3.4.4 Discussion on the Market Maker's Regret (Theorem 11)

This subsection provides some intuition for why the market maker's regret is $O(\log T)$ under IP (as shown in Theorem 11). This performance guarantee is derived via an exact analysis of the market state $\{Z_t\}$ via Lemma 4. Roughly speaking, the market state Z_t grows in the order of $\sqrt{t}$, with the market maker's one-period regret vanishing in the order of $1/Z_t^2 = 1/t$,

and her T-period regret growing in the order of $\log T$.

Representation of the market maker's regret. The following lemma expresses the market maker's regret, $\Delta^{\pi_I}(T)$, in an additive form. We relegate its proof to Appendix B.8.

Lemma 5. *We have*

$$\Delta^{\pi_I}(T) = \sum_{t=1}^{T} \mathbb{E}_1^0[l(Z_t)], \tag{3.19}$$

where the loss function $l(\cdot) : \mathbb{Z} \to \mathbb{R}_+$ *is given by*

$$l(z) = (2-c)\rho(z)\mathbb{I}\{z < \bar{z}\} + (4-2c)\rho^2(z)\mathbb{I}\{z \geq \bar{z}\}. \tag{3.20}$$

The key advantage of Lemma 5 is that through the above additive form, we not only link the regret evaluation problem with the performance evaluation step in Lemma 4, but also have a more parsimonious way of understanding regret through an intuitive characterization (see the discussions below for more details).

Understanding the dynamics of $\{Z_t\}$. In order to make sense of the convergence of s_t and the overall regret, let us give some remarks on the dynamics of $\{Z_t\}$ (both intuitively and rigorously).

Let us first characterize $\{Z_t\}$ in a heuristic fashion to gain intuition. For simplicity, we provide a characterization under hypothesis H_1 (the reasoning for H_0 is the same). Define an auxiliary stochastic process $\{Y_t\}$ such that $Y_t := 1/\rho^2(Z_t)$ for all t. Note that for sufficiently large z (i.e., $z \geq \bar{z} \vee 1$),

$$\mathbb{E}_1^0[Y_{t+1} - Y_t | Z_t = z] = \left[\tfrac{1}{2} + \rho(z)\right](r_0 + rz + 1)^2 + \left[\tfrac{1}{2} - \rho(z)\right](r_0 + rz - 1)^2 - (r_0 + rz)^2 = 5$$

is a constant. Since $Z_t \uparrow \infty$ almost surely (as in Theorem 11), we expect Y_t to grow linearly in t, expressed as $Y_t \sim t$ for brevity. Taking the appropriate transformations, we thus expect that $\rho(Z_t) = 1/\sqrt{Y_t} \sim 1/\sqrt{t}$ and $Z_t \sim 1/\rho(Z_t) \sim \sqrt{t}$. This heuristic characterization of Z_t is related to Theorem 11 in two ways. First, through a first-order Taylor expansion of $\tilde{s}(Z_t)$ as a function of $\rho(Z_t)$, we expect that $\mathfrak{d}_t \approx \rho(Z_t) \sim 1/\sqrt{t}$.[17] Second, in light of Lemma 5,

17. Recall that Statement (T11:2) in Theorem 11 expresses that $\sum_{t=1}^{T} \mathbb{E}_1^0[\mathfrak{d}_t] = O(\sqrt{T \log T})$. This statement is a $\sqrt{\log T}$ factor weaker than the above heuristic characterization, which implies that $\sum_{t=1}^{T} \mathfrak{d}_t \sim \sqrt{T}$. While a tighter estimate is possible via Lemma 4, we did not pursue it because we only need to show that

we expect that $\Delta^{\pi_I}(T) = \sum_{t=1}^{T} \mathbb{E}_1^0[l(Z_t)] \approx \sum_{t=1}^{T} \rho^2(Z_t) \sim \log T$.

We formalize our intuition on the dynamics of $\{Z_t\}$ (under hypothesis H_1) in Proposition 7 below. We relegate the proof of Proposition 7 to Appendix B.8.

Proposition 7. *For every commission rate $c \in (0,1)$ and policy parameter $r \in (0, \bar{r})$, we have the following:*

(P7:1) For all sufficiently large $M > 0$, $\sum_t \mathbb{E}_1^0[\mathbb{I}\{Z_t \leq M\}]$ converges.

(P7:2) $\sum_t \mathbb{E}_1^0[\rho(Z_t)]$ diverges at a rate satisfying $\sum_{t=1}^{T} \mathbb{E}_1^0[\rho(Z_t)] = O\left(\sqrt{T \log T}\right)$.

(P7:3) $\sum_t \mathbb{E}_1^0[l(Z_t)]$ diverges at a rate satisfying $\sum_{t=1}^{T} \mathbb{E}_1^0[l(Z_t)] = O(\log T)$.

Proposition 7 above provides the main steps for proving Theorem 11. These steps are related to (i) the almost sure convergence of spread lines, (ii) the convergence rate of spread lines, and (iii) the logarithmic growth rate of regret under IP (see the proof in Theorem 11 below). In Section 3.6, we revisit this proposition in the case where $\rho(\cdot)$ belongs to a more general family of functions (see Proposition 15 for details).

We also make a technical remark on Proposition 7 (and hence Theorem 11). In formalizing our intuition on the dynamics of Z_t, the main technical barrier is that the growth rates of Y_t and Z_t do not necessarily overwhelm stochastic fluctuations. To see why, observe that Y_t is essentially a linearly growing process, but the (conditional) second moment of the increment of Y_t, formally defined as $\mathbb{E}_1^0[(Y_{t+1} - Y_t)^2 | Z_t = z]$, grows without bound in t. In other words, Z_t is a martingale with bounded increments, but the growth rate of Z_t is sublinear ($Z_t \sim \sqrt{t}$). This barrier makes estimating $\mathbb{E}[f(Z_t)]$ difficult for each fixed t,[18] as such estimation typically relies on large-deviation based arguments. In comparison, Proposition 7 showcases how our method in Lemma 4 estimates the partial sum $\sum_{t=1}^{T} \mathbb{E}[f(Z_t)]$ directly.

$\sum_t \mathbb{E}_1^0[\mathfrak{d}_t]$ diverges in order to demonstrate the sub-exponential convergence of spread lines. This also does not affect our main goal of characterizing the regret performance.

18. In Proposition 7, the choices for $f(x)$ are $\mathbb{I}\{x \leq M\}$ and $l(x)$.

We believe that this approach is generally helpful if quantifying $\mathbb{E}[f(Z_t)]$ is more complicated than solving the difference equation in Lemma 4.

Proof of Theorem 11. Without loss of generality, we focus our analysis on H_1 with the corresponding probability measure $\mathbb{P}_1^0(\cdot) = \mathbb{P}_1^{\pi_I, \xi_1^*}(\cdot)$; the reasoning for H_0 is the same.

Statement (T11:1) in Theorem 11 follows from Statement (P7:1) in Proposition 7. Invoking the Borel-Cantelli lemma, we conclude that $Z_t \leq M$ infinitely often with $\mathbb{P}_1^0$-probability zero. As a result, $Z_t \to \infty$ almost surely under $\mathbb{P}_1^0$. Because $\rho(\cdot)$ is asymptotically vanishing (i.e., $\rho(z) \to 0$ as $z \to \infty$), this implies that $\rho(Z_t) \to 0$ and $\mathfrak{d}_t = |s_t - m_1| = |F_1^{-1}(\frac{1}{2} - \rho(Z_t)) - F_1^{-1}(\frac{1}{2})| \to 0$ almost surely under $\mathbb{P}_1^0$.

Statement (T11:2) in Theorem 11 follows from Statement (P7:2) in Proposition 7. Note that under IP,

$$\mathfrak{d}_t = |s_t - m_1| = m_1 - s_t = \frac{1}{f_1\big(F_1^{-1}\big(\frac{1}{2} - \rho(Z_t)\big)\big)}\rho(Z_t) + o(\rho(Z_t))$$

as $\rho(Z_t) \to 0$ and $Z_t \to \infty$. By Assumption 1, the term $1/f_1\big(F_1^{-1}\big(\frac{1}{2} - \rho(Z_t)\big)\big)$ is bounded away from both zero and infinity. Therefore, $\sum_{t=1}^T \mathbb{E}_1^0[\mathfrak{d}_t] = \Theta\big(\sum_{t=1}^T \mathbb{E}_1^0[\rho(Z_t)]\big)$. Invoking Statement (P7:2), $\sum_t \mathbb{E}_1^0[\mathfrak{d}_t]$ diverges, and its growth rate is such that $\sum_{t=1}^T \mathbb{E}_1^0[\mathfrak{d}_t] = O(\sqrt{T \log T})$.

Statement (T11:3) in Theorem 11 follows from Statement (P7:3) in Proposition 7. In fact, $\Delta^{\pi_I}(T) \overset{\text{Lemma 5}}{=\!=} \sum_{t=1}^T \mathbb{E}_1^0[l(Z_t)] \overset{\text{Statement (P7:3)}}{=\!=} O(\log T)$. ■

3.5 Generalized Analysis of Bayesian Policies

To deepen our understanding of how BPs perform, this section studies BPs under two distinct generalizations of our base model. These generalizations not only demonstrate the robustness of our main results for BPs but also shed light on how BPs transition from "success" into "failure" as the informed bettor's bets become more prominent. Depending on whether the informed bettor can dominate the market (at least temporarily), he needs either $\Theta(T)$ or $o(T)$ betting opportunities to make BPs systematically fail.

3.5.1 Random Blocking by Myopic Bettors

As a generalization of our base model, assume that informed bettor's each action attempt is randomly "blocked" by myopic bettors with probability q. To be more precise, suppose that in every period t, the following events happen sequentially:

1. The market maker quotes a spread line s_t.
2. The informed bettor chooses an action $a_t \in \{+1, 0, -1\}$.
3. Nature randomly picks whether the informed bettor is blocked by a myopic bettor. We encode that by χ_t, which follows an independent and identically distributed (i.i.d.) Bernoulli sequence with mean q.
    - if $\chi_t = 1$ (i.e., the informed bettor is blocked) or $a_t = 0$ (i.e., the informed bettor chooses to wait), then bet t comes from a myopic bettor. That is, $d_t = \vartheta_t$.
    - if $\chi_t = 0$ (i.e., the informed bettor has a betting opportunity) and $a_t \neq 0$ (i.e., the informed bettor chooses to act), then bet t comes from the informed bettor. That is, $d_t = a_t$.

The random blocking model differs from our base model by introducing random blocking by myopic bettors with probability q. Note that if $q = 0$, this probabilistic blocking model reduces to our base model where the informed bettor is present. If $q = 1$, the blocking model reduces to our base model any without informed bettors. Thus, this generalization bridges two extreme cases in the base model by restricting the informed bettor's ability to bet.

We define the market maker's BP, π_B, and the informed bettor's response strategy, ξ, in the same way as before, except that the realized transaction becomes $d_t = \mathbb{I}\{\chi_t = 0 \text{ and } a_t \neq 0\}a_t + \mathbb{I}\{\chi_t = 1 \text{ or } a_t = 0\}\vartheta_t$.[19] Note that the probabilistic blocking by myopic bettors adds another source of randomness to the model. To accommodate this difference, we denote

19. Here, a_t could be interpreted as a "virtual" action that may not necessarily be executed because of random blocking.

by $\hat{\mathbb{P}}_i^{\pi_B,\xi}(\cdot)$ the probability measure governing the market statistics $\{(s_t, d_t, a_t, \chi_t)\}$ and by $\hat{\mathbb{E}}_i^{\pi_B,\xi}(\cdot)$ the corresponding expectation operator, given that the market maker's policy is π_B, the informed bettor's response strategy is ξ, and the underlying hypothesis is H_i. Accordingly, we let $\hat{V}_i^{\pi_B,\xi}(T) := \sum_{t=1}^{T} \hat{\mathbb{E}}_i^{\pi_B,\xi}\left[\mathbb{I}\{\chi_t = 0\}\left(j_i^+(s_t)\,\mathbb{I}\{a_t = +1\} + j_i^-(s_t)\,\mathbb{I}\{a_t = -1\}\right)\right]$ be the informed bettor's T-period profit. With the introduction of $\hat{\mathbb{P}}_i(\cdot), \hat{\mathbb{E}}_i[\cdot]$ and $\hat{V}_i^{\pi_B,\xi}(\cdot)$, we define the informed bettor's best response strategy $\hat{\xi}_i^*$ and the market maker's regret $\hat{\Delta}^{\pi_B}(\cdot)$ in the same way as in the base model.

Recall that $\Xi = (c, m_0, m_1, F_\epsilon)$ is the collection of all problem input parameters. Let $\hat{\Xi} := (m_0, m_1, F_\epsilon)$ be the collection of problem input parameters concerning the *distribution* information only, i.e., those in Ξ except the commission rate c. Theorem 12 below summarizes our main results for our model extension where the informed bettor is randomly blocked by myopic bettors. We relegate its proof to Appendix B.9.

Theorem 12. *Suppose that the market maker uses a Bayesian policy π_B with pricing function $s^{\pi_B}(\cdot)$. Then there exist $\underline{q}, \bar{q} \in (0,1)$, which depend only on $\hat{\Xi}$, such that for every initial belief $b_1 \in (0,1)$ and sufficiently small commission rate $c > 0$, we have the following:*

(T12:1) (low blocking probability) If $q < \underline{q}$, then for some hypothesis $i \in \{0,1\}$, with strictly positive $\hat{\mathbb{P}}_i^{\pi_B,\xi_i^*}$-probability, $\mathfrak{d}_t$ does not converge to zero. Moreover, $\hat{\Delta}^{\pi_B}(T) = \Omega(T)$.

(T12:2) (high blocking probability) If $q > \bar{q}$ and the pricing function $s^{\pi_B}(\cdot)$ is regular, then for every initial belief $b_1 \in (0,1)$ and hypothesis $i \in \{0,1\}$, $\mathfrak{d}_t$ converges to zero almost surely under $\hat{\mathbb{P}}_i^{\pi_B,\hat{\xi}_i^*}$, at a rate such that $\hat{\mathbb{E}}_i^{\pi_B,\xi_i^*}[\mathfrak{d}_t] = O\big(e^{-\lambda t}\big)$ for some constant $\lambda > 0$. Moreover, $\hat{\Delta}^{\pi_B}(T) = O(1)$.

Theorem 12 generalizes our analysis of BPs to incorporate random blocking by myopic bettors. It means that all the conclusions about the failure (Theorem 8) and success (Theorem 9) of Bayesian Policies are robust even if we perturb the blocking probability q from $\{0,1\}$ by a constant independent of T.

Theorem 12 also provides us a guidance on the transition of BPs from good to poor performance as the number of the informed bettor's betting opportunities increases. Roughly speaking, BPs display good profit performance even if the informed bettor has $(1-\bar{q})T$ betting opportunities. On the other hand, BPs suffer from a linear regret even if the market maker observes $\underline{q}T$ number of bets from myopic bettors. As a result, the critical number of informed bets that make BPs transition from success to failure is $\Theta(T)$.

3.5.2 Budget-constrained Informed Bettor

As another generalization of our base model, assume that the informed bettor can place at most K bets. In this generalization, the informed bettor's decision problem is the same as the base model except that he can place up to K bets. This setting differs from our base model due to a hard constraint on the total number of bets by the informed bettor. Note that if $K = \infty$, this budget-constrained model reduces to our base model with a informed bettor. If $K = 0$, the budget-constrained model reduces to our base model without any informed bettors. This generalization, like its counterpart in the preceding subsection, connects the two extreme cases in the base model by restricting the informed bettor's ability to bet, but there is a fundamental difference between the two generalizations. Specifically, the former generalization imposes a stochastic restriction, making it difficult for the informed bettor to place many bets without the intervention of myopic bettors. The latter generalization imposes a deterministic restriction, and hence the informed bettor can still perfectly coordinate his bets as long as they are within the budget constraint.

In the budget-constrained model, the expression for the realized transaction in period t becomes $d_t = \mathbb{I}\{\sum_{\ell=1}^{t} |a_\ell| \leq K \text{ and } a_t \neq 0\}a_t + \mathbb{I}\{\sum_{\ell=1}^{t} |a_\ell| > K \text{ or } a_t = 0\}\vartheta_t$. To account for this, we denote by $\breve{\mathbb{P}}_i^{\pi_B,\xi}(\cdot)$ the probability measure governing the market statistics $\{(s_t, d_t, a_t)\}$, and by $\breve{\mathbb{E}}_i^{\pi_B,\xi}[\cdot]$ the corresponding expectation operator in the budget-constrained model, given that the market maker's policy is π_B, the informed bettor's response strategy is ξ, and the underlying hypothesis H_i. Furthermore, we let $\breve{V}_i^{\pi_B,\xi}(T;K) :=$

$\sum_{t=1}^{T} \breve{\mathbb{E}}_i^{\pi_B,\xi}\big[\big(\mathbb{I}\{\sum_{\ell=1}^{t}|a_\ell| \le K\}\big)\big(j_i^+(s_t)\,\mathbb{I}\{a_t=+1\} + j_i^-(s_t)\,\mathbb{I}\{a_t=-1\}\big)\big]$ be the informed bettor's T-period profit when the informed bettor has K betting opportunities remaining. Given the market maker's Bayesian policy π_B, the informed bettor's adaptive strategy $\breve{\xi}_i^* = \breve{\xi}_i^*(K)$ is a best response strategy if

$$\breve{\xi}_i^* \in \begin{cases} \arg\max_\xi \liminf_{T\to\infty} \big\{\breve{V}_i^{\pi_B,\xi}(T;K)\big\} & \text{if } \sup_\xi \liminf_{T\to\infty} \big\{\breve{V}_i^{\pi_B,\xi}(T;K)\big\} < \infty, \\ \arg\max_\xi \liminf_{T\to\infty} \big\{\frac{1}{T}\breve{V}_i^{\pi_B,\xi}(T;K)\big\} & \text{if } \sup_\xi \liminf_{T\to\infty} \big\{\breve{V}_i^{\pi_B,\xi}(T;K)\big\} = \infty. \end{cases}$$

Note that in general, even if the market maker's BP is a Markov policy with the posterior belief b_t serving as a state variable, $\breve{\xi}_i^*$ may not be Markov with the same state space because $\breve{\xi}_i^*$ can depend on the number of remaining bets of the informed bettor. With the introduction of $\breve{\mathbb{P}}_i^{\pi_B,\xi}(\cdot), \breve{\mathbb{E}}_i^{\pi_B,\xi}[\cdot]$, $\breve{V}_i^{\pi_B,\xi}(\cdot)$ and $\breve{\xi}_i^*$, we define the market maker's regret $\breve{\Delta}^{\pi_B}(T) = \breve{\Delta}^{\pi_B}(T;K)$ in the same way as in the base model. In our asymptotic analysis, we study how $\breve{\Delta}^{\pi_B}(T;K)$ increases as T and K grow.

Theorem 13 below summarizes our main results for our model extension where the informed bettor is budget-constrained. We relegate its proof to Appendix B.10.

Theorem 13. *Suppose that the market maker uses a Bayesian policy π_B with pricing function $s^{\pi_B}(\cdot)$. Then for every initial belief $b_1 \in (0,1)$ and sufficiently small commission rate $c > 0$, $\breve{\Delta}^{\pi_B}(T;K) = \Omega(T \wedge K)$. If in addition, the pricing function $s^{\pi_B}(\cdot)$ is regular, then $\breve{\Delta}^{\pi_B}(T;K) = \Theta(T \wedge K)$.*

Theorem 13 states that the market maker's regret under a BP is in the order of $T \wedge K$ (under certain regularity conditions) under the budget-constrained model. In particular, the regret of a BP is unbounded as long as both T and K grow to infinity, and the regret becomes $\Omega(\log T)$ if $K = \Omega(\log T)$.

3.5.3 Discussion

Our findings for the generalized models in this section (Theorems 12 and 13) demonstrate the robustness of our main results for BPs (Theorems 8 and 9). In addition, these findings

consistently imply that the success of BPs depends on the number of betting opportunities for the informed bettor. Specifically, if the informed bettor is restricted to place up to a "small" number of bets, BPs can still achieve good revenue performance. Otherwise, the market maker should consider a policy from the IP family. Roughly speaking, in the random blocking model, the transition line from success to failure of BPs is when the informed bettor has $\Theta(T)$ betting opportunities. In the budget-constrained model, the same transition line is when the informed bettor has $o(T)$ betting opportunities.

Contrasting both models further reveals how manipulation-proofness of BPs depends on the bet arrival process beyond the volume of bets from the informed bettor. Note that the aforementioned transition line between success and failure of BPs is different in the two generalizations studied in this section. The reason is that in the budget-constrained model, the informed bettor can inject large batch of bets within the budget without the intervention of any myopic bettor. But, in the random blocking model, the informed bettor's bets are randomly mixed with myopic bettors' bets. From a managerial standpoint, the difference between transition lines can be viewed as the net value to the informed bettor of the ability to "flood" the market while still maintaining anonymity.

3.6 Generalized Analysis of Inertial Policies

This section sheds light on the general designed of IP, especially focusing on why our choice of the residual probability sequence $\rho = \{\rho(z) = \frac{1}{r_0+rz}, z \in \mathbb{Z}_+\}$ is a good one. We find that there exists a problem instance such that under mild regularity conditions, it is impossible to improve performance from logarithmic regret to bounded regret by choosing a different type of residual probability sequence.

For intuition, consider a residual probability sequence ρ. If $\rho(z)$ becomes too small as z increases, IP does not push $\{Z_t\}$ in the correct direction sufficiently, and $\{Z_t\}$ behaves like a random walk. To see this, recall that we argued in the preceding section that $\{Z_t\}$ evolves

as a birth and death chain, and its drift is essentially proportional to $\rho(Z_t)$ (except for a finite number of states). Thus, if $\rho(z) < \frac{1}{4z}$ in the limit, Z_t does not diverge to infinity and the spread line does not converge to the correct median. Regarding our original choice of ρ, this intuition suggests that for the sake of ensuring convergence, it is undesirable to pick a sequence that vanishes faster. However, if $\rho(z)$ remains too large as z increases, the market maker does not fully exploit the historical data reflected in $\{Z_t\}$. In the end, if $\rho(z) > \frac{1}{4z}$ in the limit, the market maker's T-period regret would be $O\big(\sum_{t=1}^{T} \rho(t)\big)$. For our original choice of ρ, this means that it is also undesirable to pick a sequence that vanishes slower. As a result, the choice of $\rho = \{\rho(z) = \frac{1}{r_0+rz}, z \in \mathbb{Z}_+\}$ (with an appropriate value of r) makes the residual probabilities vanish at just the right rate to regularize the dynamics of $\{Z_t\}$ so as to achieve good profit performance.

Let us now formalize the above intuition. First, since we expect $\tilde{s}(z) \to m_1$ as $z \to \infty$ and $\tilde{s}(z) \to m_0$ as $z \to -\infty$, we restrict our attention to the residual probability sequences that are *vanishing*, i.e., $\rho(z) \to 0$ as $z \to \infty$. Let us further consider the following two regimes of residual probability sequences in terms of their behaviors in the limit.

Definition 4. *We say that the sequence* $\rho = \{\rho(z) \in (0, \frac{1}{2} - \alpha), z \in \mathbb{Z}_+\}$ *is fast vanishing if* $\limsup_{z\to\infty}\{z\rho(z)\} < \frac{1}{4}$*, and slowly vanishing if* $\liminf_{z\to\infty}\{z\rho(z)\} > \frac{1}{4}$ *and* $\lim_{z\to\infty}\{\rho(z)\} = 0$*.*

In Definition 4, we compare the function $\rho(\cdot)$ with the critical function, $z \mapsto \frac{1}{4z}$. Specifically, $\{\rho(z), z \in \mathbb{Z}_+\}$ is fast (resp. slowly) vanishing if it vanishes faster (resp. more slowly) than $\frac{1}{4z}$ as $z \to \infty$. The two different cases covers all the possible vanishing residual probability sequences such that $\lim_{z\to\infty}\{z\rho(z)\}$ exists and is not equal to $\frac{1}{4}$. In particular, our original choice, $\rho = \{\rho(z) = \frac{1}{r_0+rz}, z \in \mathbb{Z}_+\}$ with a sufficiently small r, belongs to the family of slowly vanishing sequences.

Definition 5. *The residual probability sequence* $\rho = \{\rho(z), z \in \mathbb{Z}_+\}$ *is regular if there exists* $A \in \mathbb{R} \cup \{\pm\infty\}$ *such that* $\frac{\rho(z)}{\rho(z+1)} = 1 + \frac{A}{z} + o\left(\frac{1}{z}\right)$ *as* z*.*

The regularity condition above means that the sequence ρ does not alternate excessively in the limit. This is closely related to the Raabe's test of convergence (Bromwich, 1908, p. 33) applied to the series $\sum_z \rho(z)$. This condition is satisfied by many well-behaved sequences such as the popular family of "test" sequences, $\{C(a+bz)^{-\mu}, z \in \mathbb{Z}_+\}$ for given $C, a, b, \mu > 0$, which includes our original choice of ρ.

Theorem 14 below studies how inertial policies with generic residual probability sequences perform when the commission rate is sufficiently large. The proof of this result is in Appendix B.11.

Theorem 14. *Let π_I be an inertial policy with a regular residual probability sequence $\rho = \{\rho(z) : z \in \mathbb{Z}_+\}$ and the commission rate c be sufficiently large so that $\xi_i^* = \xi_\emptyset$ (i.e., the type-i informed bettor's best response strategy is to never bet) for every hypothesis $i \in \{0, 1\}$. Then, we have the following:*

- *If ρ is fast vanishing, then $\{Z_t\}$ is recurrent.*
- *If ρ is slowly vanishing, then $\Delta^{\pi_I}(T)$ diverges in T at a rate satisfying $\Delta^{\pi_I}(T) = O\big(\sum_{t=1}^{T} \rho(t)\big)$.*

In either case, $\Delta^{\pi_I}(T)$ is unbounded in T.

Theorem 14 indicates that there exists a problem instance such that, if we use almost any other type of residual probability sequence, then either $\{Z_t\}$ becomes recurrent (and thus the spread line s_t fails to converge) or the regret guarantee becomes weaker than our original result. Consequently, we cannot improve performance from logarithmic regret to bounded regret by choosing a different type of residual probability sequence. This gives a partial characterization of the best achievable regret performance. While we leave the complete characterization as an open question, we conjecture that this is generally the case, i.e., it is not possible to pick any adaptive policy such that the market maker's T-period regret is bounded in T.

It is worth emphasizing that our proof of Theorem 14 is valid in more general setting in which the informed bettor participates in the market and places bets according to a threshold strategy in the following form:

$$\xi_1^{\bar{Z}}(z) = \mathbb{I}\{z < \bar{Z}\} \text{ and } \xi_0^{\bar{Z}}(z) = -\mathbb{I}\{z > -\bar{Z}\} \text{ for every } z \in \mathbb{Z},$$

where $\bar{Z} \in \mathbb{Z} \cup \{-\infty\}$. This is a generalization of the particular threshold policy in Theorem 10. This extra level of generality in the proof of Theorem 14 reveals that the above performance results can be applied to more general problem instances, as long as the informed bettor's best response strategy is of threshold type.

3.7 Concluding Remarks

Wolfers and Zitzewitz (2006) identify the question of how the market limits manipulation as one of the five open questions about prediction markets. Partially in response to this question, we study a stylized model to analyze the pricing policies of a monopolist market maker operating a spread betting market, who is uninformed of the event outcome distribution. We demonstrate that if the market maker ignores the existence of informed and strategic bettors, an informed bettor can manipulate the market by bluffing and eventually gain an abnormal amount of profit. This (negative) finding still holds even we consider a informed bettor who is restricted in the ability to dominate the market. We propose a policy, called the inertial policy, which eliminates the informed bettor's incentive to bluff, resulting in a regret up to a logarithmic factor of the total number of bets.

There are many possibilities for future work and extensions to our model. For example, one could extend the model so that myopic agents are systematically biased. One approach would be to add a bias term. If the bias term is known to the market maker (or can be empirically estimated), then our model could be extended to incorporate this setting.

Another extension would be to relax the continuity assumption for the event outcome distribution. If the feasible set of spread lines is finite, the spread line cannot be arbitrarily

close to the correct median. We conjecture that in this setting, the same inertial policy with proper randomization would be also nearly optimal. One could also extrapolate the insights of this paper to other forms of prediction markets, e.g., the odds market and the index market (see Wolfers and Zitzewitz, 2004). The main difference between those organizations and the spread betting market is the payoff structure. We conjecture that the main insights, such as the strategic manipulation of the informed bettor, as well as the rule of thumb to be inertial, carry through to those organizations of prediction markets. Lastly, we could consider relaxing the commitment assumption by considering a perfect Bayesian equilibrium (PBE) between the market maker and the informed bettor. While it is significantly harder to characterize a PBE in a similar setting (see Routledge, 1999 for some discussions in the financial market), our Theorem 8 hints that there is no PBE that always leads the spread lines to the correct median. The reason is that in a PBE where the informed bettor eventually quits, the equilibrium condition requires the market maker to essentially use a BP studied in our paper (except that we did not cover the pathological case where $s^{\pi_B}(0+)$ or $s^{\pi_B}(1-)$ does not exist), but a BP is unable to keep the informed bettor out of the market. Hence a market maker with commitment power is needed. Two further extensions of the PBE characterization are (i) multiple market makers and informed bettors; and (ii) the endogenous formation of market makers/informed bettors/commission rates. We leave these extensions for future research.

Acknowledgements. Financial support from Duke University Fuqua School of Business and University of Chicago Booth School of Business is gratefully acknowledged.

Appendices

APPENDIX A

SUPPLEMENT TO CHAPTER 2

A.1 On the Lower Bound of the Sample Complexity of any δ-accurate Policy

In this appendix will prove Theorem 1 in a slightly more general hypothesis testing framework, motivated by that of Chernoff (1959). Besides proving this result, we will also discuss in Section A.1.4 some concrete examples of alternatives ranking and selection problems that can be cast within our proposed hypothesis framework. Finally, we provide an example in Section A.1.5 on how to apply our framework in the context of dueling bandits.

A.1.1 A General Hypothesis Testing Setting Framework

Let Θ be an arbitrary parameter space of possible states of the world and let $\theta^* \in \Theta$ be the true (unknown) state. We let $\mathcal{H} = \{H_1, H_2, \ldots, H_I\} \in 2^\Theta$ denote a collection of subsets of Θ, which we interpret as alternative "hypotheses" regarding the value of θ^*. For every $\theta \in \Theta$, we let $H(\theta) = \{H_i \colon \theta \in H_i\}$ and $\overline{H}(\theta) = \mathcal{H} \setminus H(\theta)$ be the set of true and false hypotheses, respectively, when $\theta^* = \theta$. We only make one assumption regarding the relationship between parameter space Θ and the hypothesis space $\mathcal{H}$: for all $\theta \in \Theta$, there is at least one hypothesis that is true under θ.

Assumption 2. *We assume that $H(\theta) \neq \emptyset$ for all $\theta \in \Theta$.*

The setup of Θ and $\mathcal{H}$ means that our hypothesis testing framework corresponds to a multiple composite hypothesis testing problem with hypotheses that are not mutually exclusive. More specifically, let we denote $\Theta_i := \{\theta : H(\theta) = H_i\}$ to be the set of parameters that are consistent of hypothesis H_i. We allow $|\mathcal{H}| = I > 2$ (i.e. multiple), $|\Theta_i| = \infty$ for every i (i.e., composite) and do not require $\{\Theta_i\}_i$ to be mutually exclusive.

The decision maker has access to a collection $\mathcal{S}$ of experiments that she can use to learn about the value of θ^* and decide which hypotheses are true or false. An experiment $S \in \mathcal{S}$ is a random variable taking values in an outcome space $\mathcal{X}$. (This outcome space could depend on S but we consider it as independent to ease notation). We let $\{f_\theta(X|S) : X \in \mathcal{X}\}$ denote the probability distribution over outcomes of experiment $S \in \mathcal{S}$ when $\theta^* = \theta$, where $f_\theta(X|S)$ is the probability of observing outcome X when experiment S is used. We make the following assumptions on $f_\theta(\cdot|S)$:

Assumption 3. *For all $\theta \in \Theta$,*

(B-1) (Probability Mass Function) *For any $S \in \mathcal{S}$, $\sum_{X \in \mathcal{X}} f_\theta(X|S) = 1$;*

(B-2) (Non-degeneracy) *For all $S \in \mathcal{S}$ and $X \in \mathcal{X}$, $f_\theta(X|S) > 0$;*

(B-3) (Identifiability) *For every $\theta \neq \theta' \in \Theta$, there exists $S \in \mathcal{S}$ and $X \in \mathcal{X}$ such that $f_\theta(X|S) \neq f_{\theta'}(X|S)$.*

Note that this assumption is weaker than the p-separability requirement that we impose on the set of $\mathcal{M}_p$ preferences in Definition 1. The assumption about identifiability is also weaker than typically assumed in the literature. For example, Chernoff (1959) would require for all $S \in \mathcal{S}$ and $\theta \neq \theta' \in \Theta$, there exists $X \in \mathcal{X}$ such that $f_\theta(X|S) \neq f_{\theta'}(X|S)$.

An admissible learning dynamic policy has three parts:

1. A *experimentation rule*, i.e., a sequence of probability distributions $\{\lambda_t \in \Delta(\mathcal{S})\}_{t=1}^{\infty}$, with each $\lambda_t \in \mathcal{S}$ being adapted to the filtration $\mathcal{F}_t := \sigma(S_1, X_1, \ldots, S_{t-1}, X_{t-1})$.

2. A *stopping rule*, i.e., an $\mathcal{F}_t$ stopping time τ that determines when to stop experimenting.

3. A *final selection rule*, i.e., $d_\tau \in [I]$ that identifies which hypothesis to recommend.

We let $\pi = (\{\lambda_t\}_{t=1}^{\infty}, \tau, d_\tau)$ denote an admissible policy. A parameter $\theta \in \Theta$ and an admissible policy π induce a probability distribution $\mathbb{P}_\theta^\pi(\cdot)$ over the path space $\{S_1, X_1, \ldots, S_t, X_t\}_t$. We also denote by $\mathbb{E}_\theta^\pi[\cdot]$ the expectation operator under $\mathbb{P}_\theta^\pi(\cdot)$. We say that an admissible policy π is δ-*accurate* if experimentation terminates almost surely and the probability of

selecting a false hypothesis is less than δ; that is, $\mathbb{P}_\theta^\pi(\tau < \infty) = 1$ and $\mathbb{P}_\theta^\pi(d_\tau \notin H(\theta)) \leq \delta$ for any $\theta \in \mathcal{S}$.

Note that the top-ranked selection problem in the main body of the paper is a special case of above with the following characteristics: (1) The space of parameter is the set of p-separable models, $\Theta = \mathcal{M}_p$; (2) Each hypothesis H_i corresponds to case in which product i is the top-ranked product and $H(f) = \{\sigma_f^{-1}(1)\}$; (3) The set of experiments $\mathcal{S}$ is the set of display sets; (4) The outcome space $\mathcal{X}$ of an experiment coincides with the set of versions, i.e., $\mathcal{X} = [K]$. Also, we note that the identifiability requirement in Assumption 3 is weaker than the p-Separability requirement in Definition 1.

A.1.2 Re-Statement of Theorem 1

Let us restate Theorem 1 in the context of the hypothesis testing framework above. Given any experiment $S \in \mathcal{S}$ and probability distribution $\lambda \in \Delta(\mathcal{S})$, we define the Kullback-Leibler divergence between two model parameters θ and θ' as:

$$D_S\left(\theta||\theta'\right) := \sum_{k\in\mathcal{X}} f_\theta\left(k|S\right) \log \frac{f_\theta\left(k|S\right)}{f_{\theta'}\left(k|S\right)} \quad \text{and} \quad D_\lambda\left(\theta||\theta'\right) := \sum_{S\in\mathcal{S}} D_S\left(\theta||\theta'\right) \cdot \lambda(S),$$

respectively. Given any $\theta \in \Theta$, we define $\mathcal{A}\left(\theta\right) := \{\theta' \in \Theta : H(\theta) \subseteq \overline{H}(\theta')\}$ to be the set of parameters that do not agree with θ on any hypothesis that θ treats as true. We also introduce the *information* measure $I_*(\theta)$ to be the optimal value of the following Max-Min problem:

$$I_*(\theta) = \sup_{\lambda\in\Delta(\mathcal{S})} \inf_{\theta'\in\mathcal{A}(\theta)} D_\lambda\left(\theta||\theta'\right). \tag{A.1}$$

THEOREM: (Lower Bound on $\mathbb{E}_\theta^\pi[\tau]$) *Let $\delta \in (0,1)$. For every δ-accurate policy π and $\theta \in \Theta$ such that $I_*(\theta) > 0$, we have*

$$\mathbb{E}_\theta^\pi[\tau] \geq \frac{kl(\delta, 1-\delta)}{I_*(\theta)} \qquad \textit{and} \qquad \liminf_{\delta\to 0} \frac{\mathbb{E}_\theta^\pi[\tau]}{\log\left(\frac{1}{\delta}\right)} \geq \frac{1}{I_*(\theta)},$$

where $kl(\delta, 1-\delta) = \delta \log\left(\frac{\delta}{1-\delta}\right) + (1-\delta)\log\left(\frac{1-\delta}{\delta}\right)$.

A.1.3 Proof of Theorem 1

We let $\pi = (\{\lambda_t\}_{t=1}^{\infty}, \tau, d_\tau)$ be an arbitrary δ-accurate policy and let $\mathcal{F}_t$ denote the filtration generated by the sampling history $\mathcal{H}_t := \{S_1, X_1, \ldots, S_t, X_t\}$. Without loss of generality, we will assume that $\mathbb{E}_\theta[\tau] < \infty$, for otherwise the theorem is trivially true.

For any state $\theta \in \Theta$ and any alternative state $\theta' \in \mathcal{A}(\theta)$ we define the log-likelihood function

$$L_t^{\theta,\theta'} := L^{\theta,\theta'}(\mathcal{H}_t) = \sum_{\ell=1}^{t} \log\left(\frac{f_\theta(X_\ell|S_\ell)}{f_{\theta'}(X_\ell|S_\ell)}\right).$$

As in Wald's sequential probability ratio test (SPRT), we expect that a δ-accurate policy will stop sampling when $L_t^{\theta,\theta'}$ exceeds an upper threshold under θ. In what follows, we formalize this intuition and show that

$$\mathbb{E}_\theta\left[L_\tau^{\theta,\theta'}\right] \geq kl(\delta, 1-\delta), \quad \forall\theta \in \Theta \text{ and } \forall\theta' \in \mathcal{A}(\theta). \tag{A.2}$$

The proof of (A.2) is done in two steps:

Step 1: From Lemma 19 (equation (19)) in Kaufmann et al. (2016), it follows that for any event $\mathcal{E} \in \mathcal{F}_\tau$ we have that

$$\mathbb{E}_\theta[L_\tau^{\theta,\theta'}|\mathcal{E}] \geq \log\left(\frac{\mathbb{P}_\theta(\mathcal{E})}{\mathbb{P}_{\theta'}(\mathcal{E})}\right).$$

Hence, for any partition $\mathcal{P}$ of events of $\mathcal{F}_\tau$, the previous inequality implies that

$$\mathbb{E}_\theta[L_\tau^{\theta,\theta'}] \geq \sum_{\mathcal{E}\in\mathcal{P}} \mathbb{P}_\theta(\mathcal{E}) \log\left(\frac{\mathbb{P}_\theta(\mathcal{E})}{\mathbb{P}_{\theta'}(\mathcal{E})}\right). \tag{A.3}$$

Step 2: Let us first suppose that $\delta \in (0, \frac{1}{2}]$ and consider the partition $\mathcal{P} = \{\mathcal{E}', \Omega \setminus \mathcal{E}'\}$, where $\mathcal{E}' := \{d_\tau \in H(\theta')\}$ is the event that the recommended hypothesis d_τ is true under θ'. Since $\theta' \in \mathcal{A}(\theta)$, $H(\theta) \subset \overline{H}(\theta')$, and hence $H(\theta') \subset \overline{H}(\theta)$. As a further consequence, $\mathcal{E}' \subset \{d_\tau \in \overline{H}(\theta)\}$, the event that the recommended hypothesis d_τ is false under θ. It follows

that $\mathbb{P}_\theta(\mathcal{E}') \leq \delta$ and $\mathbb{P}_{\theta'}(\mathcal{E}') \geq 1-\delta$ since π is δ-accurate. As a result,

$$\begin{aligned}
\sum_{\mathcal{E}\in\mathcal{P}} \mathbb{P}_\theta(\mathcal{E}) \log\left(\frac{\mathbb{P}_\theta(\mathcal{E})}{\mathbb{P}_{\theta'}(\mathcal{E})}\right) &= \mathbb{P}_\theta(\mathcal{E}')\log\left(\frac{\mathbb{P}_\theta(\mathcal{E}')}{\mathbb{P}_{\theta'}(\mathcal{E}')}\right) + \big(1-\mathbb{P}_\theta(\mathcal{E}')\big)\log\left(\frac{1-\mathbb{P}_\theta(\mathcal{E}')}{1-\mathbb{P}_{\theta'}(\mathcal{E}')}\right) \\
&= \mathbb{P}_\theta(\mathcal{E}')\log\big(\mathbb{P}_\theta(\mathcal{E}')\big) + \big(1-\mathbb{P}_\theta(\mathcal{E}')\big)\log\big(1-\mathbb{P}_\theta(\mathcal{E}')\big) \\
&\quad -\Big[\mathbb{P}_\theta(\mathcal{E}')\log\big(\mathbb{P}_{\theta'}(\mathcal{E}')\big) + \big(1-\mathbb{P}_\theta(\mathcal{E}')\big)\log\big(1-\mathbb{P}_{\theta'}(\mathcal{E}')\big)\Big] \\
&\geq \delta\log\left(\tfrac{\delta}{1-\delta}\right) + (1-\delta)\log\left(\tfrac{1-\delta}{\delta}\right) = kl(\delta,1-\delta). \qquad \text{(A.4)}
\end{aligned}$$

The inequality is due to the conditions $\mathbb{P}_\theta(\mathcal{E}') \leq \delta$ and $\mathbb{P}_{\theta'}(\mathcal{E}') \geq 1-\delta$ and the fact that the negative entropy function $\delta \mapsto \big[\delta\log(\delta) + (1-\delta)\log(1-\delta)\big]$ decreases in the interval $(0, 1/2]$ and increases in the interval $[1/2, 1)$. We have thus verified our claim for $\delta \in (0, 1/2]$. The same argument can be made when $\delta \in [\frac{1}{2}, 1)$ by defining $\mathcal{E}' := \{d_\tau = \theta^{-1}(1)\}$.

Combining (A.3) and (A.4) we get (A.2).

Let us now express $\mathbb{E}_\theta\left[L_\tau^{\theta,\theta'}\right]$ in terms of the value of the Kullback-Leibler divergence $D_S\left(\theta||\theta'\right)$. To this end, let $N_t(S) := \sum_{i=1}^t \mathbb{I}\{S_i = S\}$ be the number of times experiment S is used between time period 1 and t, under policy π. Also, for each $S \in \mathcal{S}$, we define a sequence $\{Y_\ell^S\}_{\ell\geq 1}$ of iid random variables with probability distribution $f_\theta(\cdot|S)$. It follows that

$$\begin{aligned}
\mathbb{E}_\theta\left[L_\tau^{\theta,\theta'}\right] &= \mathbb{E}_\theta\left[\sum_{t=1}^\tau \log\left(\frac{f_\theta(X_t|S_t)}{f_{\theta'}(X_t|S_t)}\right)\right] = \mathbb{E}_\theta\left[\sum_{S\in\mathcal{S}}\sum_{t=1}^\tau \mathbb{I}\{S_t = S\}\log\left(\frac{f_\theta(X_t|S)}{f_{\theta'}(X_t|S)}\right)\right] \\
&= \sum_{S\in\mathcal{S}} \mathbb{E}_\theta\left[\sum_{\ell=1}^{N_\tau(S)} \log\left(\frac{f_\theta(Y_\ell^S|S)}{f_{\theta'}(Y_\ell^S|S)}\right)\right] = \sum_{S\in\mathcal{S}} \mathbb{E}_\theta[N_\tau(S)]\,\mathbb{E}_\theta\left[\log\left(\frac{f_\theta(Y_1^S|S)}{f_{\theta'}(Y_1^S|S)}\right)\right] \\
&= \sum_{S\in\mathcal{S}} \mathbb{E}_\theta[N_\tau(S)]\, D_S\left(\theta||\theta'\right), \qquad \text{(A.5)}
\end{aligned}$$

where the second to last equality follows from Wald's identity (recall that we have assumed that $\mathbb{E}_\theta[\tau] < \infty$ which implies that $\mathbb{E}_\theta[N_\tau(S)] < \infty$). Combining (A.2) and (A.5), we get that

$$kl(\delta, 1-\delta) \leq \inf_{\theta'\in\mathcal{A}(\theta)} \sum_{S\in\mathcal{S}} \mathbb{E}_\theta[N_\tau(S)]\, D_S\left(\theta||\theta'\right) \qquad \text{(A.6)}$$

The final step of the proof of Theorem 1 is to show that the right-hand side of (A.6) is bounded above by $\mathbb{E}_\theta[\tau] I_*(\theta)$. Indeed,

$$\begin{aligned} & \inf_{\theta' \in \mathcal{A}(\theta)} \sum_{S \in \mathcal{S}} \mathbb{E}_\theta[N_\tau(S)] \; D_S\left(\theta || \theta'\right) \\ = \; & \mathbb{E}_\theta[\tau] \inf_{\theta' \in \mathcal{A}(\theta)} \sum_{S \in \mathcal{S}} \frac{\mathbb{E}_\theta[N_\tau(S)]}{\mathbb{E}_\theta[\tau]} D_S\left(\theta || \theta'\right) \qquad \text{(A.7)} \\ \leq \; & \mathbb{E}_\theta[\tau] \sup_{\lambda \in \Delta(\mathcal{S})} \inf_{\theta' \in \mathcal{A}(\theta)} \sum_{S \in \mathcal{S}} \lambda(S) \, D_S\left(\theta || \theta'\right) = \mathbb{E}_\theta[\tau] I_*(\theta). \qquad \text{(A.8)} \end{aligned}$$

The inequality in (A.8) follows from the fact that $\lambda(S) = \frac{\mathbb{E}_\theta[N_\tau(S)]}{\mathbb{E}_\theta[\tau]}$ is a feasible choice of $\Delta(\mathcal{S})$. Combining (A.6) and (A.8) we get that

$$\mathbb{E}_\theta[\tau] \geq \frac{kl(\delta, 1-\delta)}{I_*(\theta)},$$

which proves the first part of Theorem 1. The second part follows from noticing that $\liminf_{\delta \to 0} kl(\delta, 1-\delta)/\log(1/\delta) = 1$. ■

A.1.4 Application to Other Ranking-and-Selection Problems

The hypothesis testing framework presented in Section A.1.1 can accommodate a variety problem formulations. For example, one could change the definition of $\mathcal{H}$ to capture different objectives. For example, in a *full ranking identification* problem, the company is interested in recovering the full ranking σ_* associated to the ground-truth preference f_*. To incorporate that, we can define the set of hypoteses $\{H_\sigma \colon \sigma \in \Sigma\}$, where H_σ denotes the subset of preferences f for which $\sigma_f = \sigma$. In a *"strong" top-k identification* problem, the company wishes to identify a collection of items S with size $k = |S|$ such that all the items in S are ranked higher than those in $[K] \setminus S$. To model this setting we define hypotheses $\{H_S \colon S \subseteq \mathcal{S},\ |S| = k\}$, where H_S is the subset of prferences f such that $S = \{\sigma_f^{-1}(1), \ldots, \sigma_f^{-1}(k)\}$. In a *"weak" top-k identification* problem, the company wishes to identify a single item that is ranked k^{th} or higher. To model this problem, we use the set of hypotheses $\{H_i \colon i \in [K]\}$,

where H_i is the subset of preferences f for which product i is ranked at least top k that is, $i \in \{\sigma_f^{-1}(1), \ldots, \sigma_f^{-1}(k)\}$.

Besides considering different ranking and selection objectives, one can also adapt our framework to (i) consider additional constraints on the set $\mathcal{S}$ of available display sets (e.g., let $\mathcal{S} = \{S : |S| \leq m\}$ to incorporate capacity constraints) or (ii) use Θ to capture different forms of information structure (e.g., let Θ to be a larger set than $\mathcal{M}_p$ to relax the p-separability constraint) or (iii) use $\mathcal{X}$ to capture different feedback structures (e.g., consumers provide full rankings rather than single choices).

While we have a relatively good understanding of the lower bound of sample complexity, we leave it an open problem as of whether an MTP-based sampling algorithm still achieves a sample complexity that matches the lower bound (asymptotically). Also, we are interested in understanding the structural properties of MTP under these different formulations. For instance, we conjecture that the randomization distribution, as a result of the new Max-Min problem, is sparse in general.

A.1.5 A Dueling Bandit Example

We conclude this appendix devoted to the lower bound in Theorem 1 with an example of how to apply the result in the context of a different ranking-and-selection problem. Specifically, we consider the problem of identifying the product with the highest Borda score studied by Jamieson et al. (2015) (see also Heckel et al., 2019). In this problem, only pairwise comparisons (i.e., dueling bandits) are considered and consumers' preferences are then represented by a matrix $P = [p_{ij}]$, where p_{ij} is the probability that version $i \in [K]$ is preferred over version $j \in [K]$ when both are displayed together. The Borda score of product i is defined by

$$s_i := \tfrac{1}{K-1} \sum_{j \neq i} p_{ij},$$

that is, s_i is the probability that version i is selected when displayed with another uniformly randomly selected version. Jamieson et al. (2015) considered the fixed-confidence identification problem of finding the version $i^* = \arg\max\{s_i : i \in [K]\}$ and derived the following lower bound on the sample complexity of any δ-accurate policy.

THEOREM: (Jamieson et al., 2015) *Consider a comparison matrix P such that (i) item 1 is the Borda winner; (ii) $\frac{3}{8} \leq p_{ij} \leq \frac{5}{8}, \forall i, j \in [K]$, and (iii) $K \geq 3$. Then for $\delta \leq 0.15$, any δ-PAC dueling bandits algorithm π to find the Borda winner has*

$$\mathbb{E}_P^{\pi}[\tau] \geq \frac{1}{40}\left(\frac{K-2}{K-1}\right)^2 \left(\sum_{i=2}^{K} \frac{1}{(s_1 - s_i)^2}\right) \log \frac{1}{2\delta}.$$

The Borda winner identification problem in Jamieson et al. (2015) is closely related to our framework in Section A.1.1 with the following characteristics:

- The space of parameters is the set of comparison matrices $P = [p_{ij}]$, where the Borda winder is uniquely (and strictly) defined. That is, $\Theta^P = \{P \in \mathbb{R}_{++}^{K \times K} : p_{ij} + p_{ji} = 1 \text{ and } \exists\, i_* \in [K] \text{ such that } s_{i_*} > \max_{j \neq i_*} s_j\}$. Let us denote $i_*(P)$ to be the Borda winner matrix P;

- Each hypothesis H_i corresponds to case in which item i is the best item. That is, $\mathcal{H} = [K]$ and $H(P) = \{i_*(P)\}$;

- The set of experiments $\mathcal{S}^P$ is the set of all (unordered) pairwise display sets. That is, $\mathcal{S}^P = \{\{i, j\} : i \neq j \text{ and } i, j \in [K]\}$;

- The stochastic comparison model is such that $f_P(i|\{i, j\}) = p_{ij}$ and $f_P(j|\{i, j\}) = p_{ji}$.

The specifications above exactly matches the problem setup in Jamieson et al. (2015) except that we further allow the dueling bandits algorithm π to know the additional information that P is non-degenerate, i.e., $p_{ij} \neq \{0, 1\}$ for all i, j. Within this setting, our lower bound stated in Section A.1.2 can be transferred to the following result:

THEOREM: (Our lower bound in the setting of Jamieson et al., 2015) *Let $\delta \in (0, 1)$ and*

$P \in \Theta^P$. *Any δ-PAC dueling bandits algorithm π to find the Borda winner has*

$$\mathbb{E}_P^{\pi}[\tau] \geq \frac{kl(\delta, 1-\delta)}{\sup_{\lambda \in \Delta(\mathcal{S}^P)} \inf_{\tilde{P}: i_*(\tilde{P}) \neq i_*(P)} \sum_i \sum_{j>i} \lambda(\{i,j\})\ kl(p_{ij}, \tilde{p}_{ij})},$$

where $kl(\delta, 1-\delta) = \delta \log\left(\frac{\delta}{1-\delta}\right) + (1-\delta)\log\left(\frac{1-\delta}{\delta}\right)$.

We claim that our lower bound dominates that in Jamieson et al. (2015). On one hand, our lower bound applies to a larger collection of comparison matrices P and values of δ. On the other hand, our lower bound is tighter (i.e., larger) than that in Jamieson et al. (2015) whenever a direct comparison is possible. We formalize the second part of our claim in the result below.

Proposition 8. *Let $P \in \Theta^P$ be such that (i) item 1 is the Borda winner; (ii) $\frac{3}{8} \leq p_{ij} \leq \frac{5}{8},\ \forall i, j \in [K]$, and (iii) $K \geq 3$. Then for $\delta \leq 0.15$,*

$$\begin{aligned}&\frac{kl(\delta, 1-\delta)}{\sup_{\lambda \in \Delta(\mathcal{S}^P)} \inf_{\tilde{P}: i_*(\tilde{P}) \neq i_*(P)} \sum_i \sum_{j>i} \lambda(\{i,j\})\ kl(p_{ij}, \tilde{p}_{ij})} \\ \geq &\frac{1}{40}\left(\frac{K-2}{K-1}\right)^2 \left(\sum_{i \neq 1} \frac{1}{(s_1 - s_i)^2}\right) \log \frac{1}{2\delta}.\end{aligned}$$

Proof of Proposition 8: Let$P \in \Theta^P$ be an arbitrary comparisong matrix that satisfies the condition (i), (ii) and (iii) in the proposition. Also, select $\delta \leq 0.15$. Notice that $kl(\delta, 1-\delta) \geq \log \frac{1}{2\delta}$ when $\delta \leq 0.15$. Hence it suffices to show that

$$\sup_{\lambda \in \Delta(\mathcal{S}^P)} \inf_{\tilde{P}: i_*(\tilde{P}) \neq i_*(P)} \sum_i \sum_{j>i} \lambda(\{i,j\})\ kl(p_{ij}, \tilde{p}_{ij}) \leq \frac{1}{\frac{1}{40}\left(\frac{K-2}{K-1}\right)^2 \left(\sum_{i \neq 1} \frac{1}{(s_1 - s_i)^2}\right)}.$$

In order to show the inequality above, let us follow some the constructions in Jamieson et al.

(2015). For every $b \in \{2, \dots, K\}$, we select the alternative comparison matrix

$$\tilde{p}^b_{ij} = \begin{cases} p_{ij} + \frac{K-1}{K-2}(s_1 - s_b) + \varepsilon & \text{if } i = b \text{ and } j \neq \{1, b\} \\ p_{ij} - \frac{K-1}{K-2}(s_1 - s_b) - \varepsilon & \text{if } j = b \text{ and } i \neq \{1, b\} \\ p_{ij} & \text{otherwise.} \end{cases}$$

In other words, $\tilde{P}^b$ is constructed so that b is the Borda winner under $\tilde{P}^b$ and $\tilde{P}^b$ differs from P only in the indices $\{(b, j) : j \neq \{1, b\}\}$. Moreover, let us pick a sufficiently small ε, so that $\tilde{P}^b \in \Theta^P$ and

$$0 < \max_{j \neq \{1,b\}} \; kl(p_{bj}, \tilde{p}'_{bj}) < 20 \left(\frac{K-1}{K-2}(s_1 - s_b) \right)^2 =: \mathfrak{C}_b.$$

We refer the reader to Equations (4)-(5) in Jamieson et al. (2015) for discussions on why we are able to select such an ε. With the construction above, we notice that

$$\begin{aligned}
& \sup_{\lambda \in \Delta(\mathcal{S}^P)} \quad \inf_{\tilde{P}: i_*(\tilde{P}) \neq i_*(P)} \quad \sum_i \sum_{j>i} \lambda(\{i,j\}) \; kl(p_{ij}, \tilde{p}_{ij}) \\
= & \sup_{\lambda \in \Delta(\mathcal{S}^P)} \quad \min_{b \in \{2,\dots,K\}} \quad \inf_{\tilde{P}: i_*(\tilde{P}) = b} \quad \sum_i \sum_{j>i} \lambda(\{i,j\}) \; kl(p_{ij}, \tilde{p}_{ij}) \\
\leq & \sup_{\lambda \in \Delta(\mathcal{S}^P)} \quad \min_{b \in \{2,\dots,K\}} \quad \sum_i \sum_{j>i} \lambda(\{i,j\}) \; kl(p_{ij}, \tilde{p}^b_{ij}) \\
= & \sup_{\lambda \in \Delta(\mathcal{S}^P)} \quad \min_{b \in \{2,\dots,K\}} \quad \sum_{j \neq \{1,b\}} \lambda(\{b,j\}) \; kl(p_{bj}, \tilde{p}^b_{bj}) \\
\leq & \sup_{\lambda \in \Delta(\mathcal{S}^P)} \quad \min_{b \in \{2,\dots,K\}} \quad \sum_{j \neq \{1,b\}} \lambda(\{b,j\}) \; \mathfrak{C}_b \\
\leq & \sup_{x_b \geq 0: \, \sum_{b=2}^K x_b \leq 2} \quad \min_{b \in \{2,\dots,K\}} \quad x_b \; \mathfrak{C}_b \qquad\qquad [x_b := \sum_{j \neq \{1,b\}} \lambda(\{b,j\})] \\
= & \frac{2}{\frac{1}{\mathfrak{C}_2} + \cdots \frac{1}{\mathfrak{C}_K}} = \frac{1}{\frac{1}{40}\left(\frac{K-2}{K-1}\right)^2 \left(\sum_{i=2}^K \frac{1}{(s_1 - s_i)^2} \right)}.
\end{aligned}$$

That finishes the proof. ■

A.2 Proof of Theorem 2

This proof draws inspiration from Chernoff (1959) and is carefully adapted to the current setting.

A.2.1 Preliminaries

Let us introduce some notations. Define $\overline{\mathcal{M}}_p^{\text{F}}(f) := \overline{\mathcal{M}}_p(f) \cap \mathcal{M}_p^{\text{F}}$. Also, let us introduce $I_*^F(f) := \max_{\lambda\in\Delta(\mathcal{S})} \min_{\bar{f}\in\overline{\mathcal{M}}_p^{\text{F}}(f)} D_\lambda\left(f||\bar{f}\right) \geq \max_{\lambda\in\Delta(\mathcal{S})} \min_{\bar{f}\in\overline{\mathcal{M}}_p(f)} D_\lambda\left(f||\bar{f}\right) = I_*(f)$. Let $\lambda_*^F(f) \in \arg\max_{\lambda\in\Delta(\mathcal{S})} \min_{\bar{f}\in\overline{\mathcal{M}}_p^{\text{F}}(f_t^F)} D_\lambda\left(f_t^F||\bar{f}\right)$. If there are multiple optimal solutions, we pick a arbitrary but fixed rule to break ties (e.g., in the lexicographical order). The detailed description of the policy $\hat{\pi}$ is summarized in Algorithm 2.

Algorithm 2: Policy $\hat{\pi}$

INPUT: $\mathcal{M}_p^{\text{F}}$.

STEP 1: At each epoch t, given the history of votes $(S_1, X_1, \ldots, S_t, X_t)$, compute the most likely consensus preference by solving the MLE problem under $\mathcal{M}_p^{\text{F}}$

$$f_t^F \in \arg\max_{f\in\mathcal{M}_p^{\text{F}}} \sum_{\ell=1}^{t} \log f(X_\ell|S_\ell). \tag{A.9}$$

We break ties arbitrarily if the arg max in (MLE) is not a singleton.

STEP 2: Update the value of the generalized log-likelihood ratio process

$$\mathcal{L}_t^F = \min_{\bar{f}\in\overline{\mathcal{M}}_p^{\text{F}}(f_t^F)} L_t^{f_t^F,\bar{f}}. \tag{A.10}$$

If $\mathcal{L}_t^F \geq \beta^F := \log(|\mathcal{M}_p^{\text{F}}|) + \log\left(\frac{1}{\delta}\right)$ then stop and select the top-ranked version according to f_t^F that is, $\sigma_{f_t^F}^{-1}(1)$. Otherwise, go to Step 3.

STEP 3: If t is a perfect square number (i.e., there exists an integer i such that $t = i^2$), then pick λ_t to be the uniform distribution over all display sets, i.e., $\lambda_t(S) = \frac{1}{|\mathcal{S}|}$ for all $S \in \mathcal{S}$. Otherwise, pick $\lambda_t = \lambda_*^F(f_t^F)$. Randomly select a set using the probability distribution λ_t to be displayed to the next consumer and record her choice X_{t+1}. Go to Step 1 and iterate. □

A.2.2 Main Body of Proof

Proof of Theorem 2. Let us break the proof into two steps.

Step 1. We claim that for all $f \in \mathcal{M}_p^F$, (i) $\mathbb{E}_f[\tau] < +\infty$ for all $\delta \in (0,1)$; and (ii) $\limsup_{\delta\downarrow 0} \frac{\mathbb{E}_f^{\hat{\pi}}[\tau]}{\log\left(\frac{1}{\delta}\right)} \leq \frac{1}{I_*^F(f)}$, which implies that $\limsup_{\delta\downarrow 0} \frac{\mathbb{E}_f^{\hat{\pi}}[\tau]}{\log\left(\frac{1}{\delta}\right)} \leq \frac{1}{I_*(f)}$.

We invoke the following auxiliary lemma that gives an upper bound to the tail probability of τ, i.e., $\mathbb{P}_f(\tau \geq t)$. Section A.2.3 contains proof of this auxiliary lemma.

Lemma 6. *For all $0 < \epsilon < 1$, there exists a convergent series $\{\rho_t\}_t > 0$ (i.e. $\sum_{t=1}^{\infty} \rho_t < \infty$), which is independent of δ, such that for every $\delta \in (0,1)$, $f \in \mathcal{M}_p^{\mathrm{F}}$ and $t \geq M(\delta) := \frac{1+\epsilon}{I_*^F(f)} \log \frac{1}{\delta}$,*

$$\mathbb{P}_f(\tau \geq t) \leq \rho_t. \tag{A.11}$$

Lemma 6 above is sufficient for Step 1. To see why, pick an arbitrary $\epsilon \in (0,1)$, $f \in \mathcal{M}_p^F$, and $\{\rho_t\}_t$ as stated in the lemma.

$$\begin{aligned}
\mathbb{E}_f[\tau] = \sum_{t=1}^{\infty} \mathbb{P}_f(\tau \geq t) &\leq \sum_{t=1}^{M(\delta)} \mathbb{P}_f(\tau \geq t) + \sum_{t=M(\delta)}^{\infty} \mathbb{P}_f(\tau \geq t) && [M(\delta) \text{ is defined in Lemma 6}] \\
&\leq \sum_{t=1}^{M(\delta)} 1 + \sum_{t=M(\delta)}^{\infty} \mathbb{P}_f(\tau \geq t) && [\mathbb{P}_f(\tau \geq t) \leq 1] \\
&\leq M(\delta) + \sum_{t=M(\delta)}^{\infty} \rho_t && [\text{Lemma 6}] \\
&\leq \tfrac{1+\epsilon}{I_*^F(f)} \log \tfrac{1}{\delta} + C', && [C' := \sum_{t=1}^{\infty} \rho_t < \infty]
\end{aligned}$$

Due to Lemma 6, C' is a finite constant independent of δ. Hence $\frac{\mathbb{E}_f[\tau]}{\log \frac{1}{\delta}} \leq \frac{1+\epsilon}{I_*^F(f)} + \frac{C'}{-\log \delta} < +\infty$. Moreover, $\frac{\mathbb{E}_f[\tau]}{\log \frac{1}{\delta}} \leq \frac{1+2\epsilon}{I_*^F(f)}$ for sufficiently small δ. Take $\epsilon, \delta \to 0$, and the we finish the proof.

Step 2. We claim that $\hat{\pi}$ is $\delta(\mathcal{M}_p^{\mathrm{F}})$-accurate for every $\delta \in (0,1)$.

Given any $f \in \mathcal{M}_p^F$, $\mathbb{P}_f(\tau < \infty) = 1$ due to Lemma 6. Given every $\bar{f} \in \overline{\mathcal{M}}_p^F(f)$, let $\mathcal{E}_{\bar{f}}$ be the event that the policy $\hat{\pi}$ terminates with estimated state $f_\tau^F = \bar{f}$ (which produces a

mistake). Notice that

$$\begin{aligned}
&\mathbb{P}_f(\mathcal{E}_{\bar{f}}) \\
=&\mathbb{E}_{\bar{f}}[\mathbb{I}\{\mathcal{E}_{\bar{f}}\}\exp(-L_\tau^{\bar{f},f})] && \text{[change-of-measure]} \\
\leq&\tfrac{\delta}{|\mathcal{M}_p^{\mathrm{F}}|}. && \text{[Due to the stopping rule } \tau\text{, } L_\tau^{\bar{f},f} \geq \beta^F = \log(|\mathcal{M}_p^{\mathrm{F}}|) + \log\left(\tfrac{1}{\delta}\right)]
\end{aligned}$$

As a result, $\mathbb{P}_f(d_\tau \neq \sigma_f^{-1}(1)) = \sum_{\bar{f}\in\overline{\mathcal{M}}_p^F(f)} \mathbb{P}_f(\mathcal{E}_{\bar{f}}) \leq \frac{|\overline{\mathcal{M}}_p^F(f)|}{|\mathcal{M}_p^{\mathrm{F}}|}\delta \leq \delta$. ■

Remark 3. *The particular value of β^F is only used in Step 2 above. As a result, we can change β^F to $\tilde{C}^F + \log(1/\delta)$ for an arbitrary finite constant $\tilde{C}^F$ independent of δ, without affecting the conclusion in Step 1 above, i.e., $\limsup_{\delta\downarrow 0} \frac{\mathbb{E}_f^{\hat{\pi}}[\tau]}{\log\left(\frac{1}{\delta}\right)} \leq \frac{1}{I_*^F(f)}$ (also see Step 3 of the proof of Lemma 8).*

A.2.3 Proof of the Auxiliary Lemma 6

Before we prove Lemma 6, we first prove two technical lemmas below.

Lemma 7. *Given any $\mathbb{Z}_+$ valued random variable X and probability measure μ, the following are equivalent:*

1. $\mathbb{E}_\mu[X^2] < \infty$;
2. $\sum_{s=1}^\infty s\mathbb{P}_\mu(X \geq s) < \infty$;
3. $\sum_{t=1}^\infty \sum_{s=t}^\infty \mathbb{P}_\mu(X \geq s) < \infty$.

Proof. We prove this lemma by applying Fubini's Theorem twice. Note that all the items involved in the summation are nonnegative,

$$\begin{aligned}
\sum_{t=1}^\infty\sum_{s=t}^\infty \mathbb{P}_\mu(X\geq s) = \sum_{s=1}^\infty\sum_{t=1}^s \mathbb{P}_\mu(X\geq s) = \sum_{s=1}^\infty s\mathbb{P}_\mu(X\geq s) = \sum_{s=1}^\infty s\sum_{z=s}^\infty \mathbb{P}_\mu(X=z) \\
= \sum_{z=1}^\infty \left(\sum_{s=1}^z s\right)\mathbb{P}_\mu(X=z) \\
= \sum_{z=1}^\infty \tfrac{z(z+1)}{2}\mathbb{P}_\mu(X=z) = \tfrac{\mathbb{E}_\mu[X^2]}{2} + \tfrac{\mathbb{E}_\mu[X]}{2},
\end{aligned}$$

and hence the statement of the lemma follows. ■

Let $\hat{\tau} := \max\{t : f_t^F \neq f\}$, the last time f_t^F is not equal to f. Lemma 8 demonstrates the speed of convergence of the estimated preference f_t^F to f.

Lemma 8. *For all $f \in \mathcal{M}_p^F$, there exists $C, \epsilon > 0$, independent of δ, such that for every $t \in \{1, 2, \ldots\}$, $\mathbb{P}_f(f_t^F \neq f) \leq Ce^{-\epsilon\sqrt{t}}$. As a result, $\mathbb{E}_f[\hat{\tau}^2] < +\infty$.*

Proof of Lemma 8. Pick an arbitrary $\bar{f} \in \mathcal{M}_p^F \setminus \{f\}$. For all $t \in \mathbb{Z}_+$, $D_{\lambda_t}\left(f||\bar{f}\right) \geq 0$. Moreover, since $f \neq \bar{f}$, there exists $S \in \mathcal{S}$ such that $D_S\left(f||\bar{f}\right) > 0$, and thus $D_{\lambda_t}\left(f||\bar{f}\right) > 0$ if t is a perfect square number, due to the construction of Step 3 in the policy $\hat{\pi}$. We refer to the same argument in Lemma 1 of Chernoff (1959), and conclude that there exists $C_{\bar{f}}, \epsilon_{\bar{f}} > 0$ such that $\mathbb{P}_f(f_t^F = \bar{f}) \leq C_{\bar{f}} e^{-\epsilon_{\bar{f}}\sqrt{t}}$, for every $t \in \{1, 2, \ldots\}$. It now suffices to pick $\epsilon := \min\{\epsilon_{\bar{f}} : \bar{f} \in \overline{\mathcal{M}}_p^F\}$ and $C := |\mathcal{M}_p^F| \max\{C_{\bar{f}} : \bar{f} \in \overline{\mathcal{M}}_p^F\}$ to satisfy our claim. Noting that $e^{-\epsilon\sqrt{t}}$ decays faster than $1/t^\alpha$ for all $\alpha > 0$,

$$\sum_{t=1}^{\infty} t\mathbb{P}_f(\hat{\tau} \geq t) \leq \sum_{t=1}^{\infty} t\left(\sum_{\ell=t}^{\infty} \mathbb{P}_f(f_t^F \neq f)\right) \leq C\sum_{t=1}^{\infty} t\left(\sum_{\ell=t}^{\infty} e^{-\epsilon\sqrt{\ell}}\right) < +\infty.$$

Hence, $\mathbb{E}_f[\hat{\tau}^2] < +\infty$. In addition, the quantity does not depend on δ, because δ is only used for the stopping rule. ■

Proof of Lemma 6. Fix the $\epsilon \in (0, 1)$ and $f \in \mathcal{M}_p^F$ stated in Lemma 6 throughout the proof. Also let $\beta^F := \log(|\mathcal{M}_p^F|) + \log\left(\frac{1}{\delta}\right)$ for ease of notation. We split the discussion into three parts.

Part 1. We give an upper bound of the tail probability $\mathbb{P}_f(\tau \geq T)$. This upper bound implies that it suffices to show that the probability $\mathbb{P}_f\left(L_t^{f,\bar{f}} < \beta^F\right)$ is small (uniformly in $\bar{f} \in \overline{\mathcal{M}}_p^F(f)$) when $t \geq M(\delta)$.

For every $\bar{f} \in \overline{\mathcal{M}}_p^F(f)$, we introduce $\tau_{\bar{f}} := \max\{t : L_t^{f,\bar{f}} < \beta^F\} \in \{0, 1, \ldots, \infty\}$ to be the

final time that $L_t^{f,\bar{f}}$ is below β^F. According to definition of τ under policy $\hat{\pi}$,

$$\begin{aligned}
\mathbb{P}_f(\tau \geq t) =& \mathbb{P}_f\left(\min\left\{\ell : \min_{\bar{f}\overline{\mathcal{M}}_p^F(f_\ell^F)} L_\ell^{f_\ell^F,\bar{f}} \geq \beta^F\right\} \geq t\right) \\
\leq& \mathbb{P}_f\left(\max\left\{\ell : \min_{\bar{f}\in\overline{\mathcal{M}}_p^F(f_\ell^F)} L_\ell^{f_\ell^F,\bar{f}} < \beta^F\right\} \geq t-1\right) \\
\leq& \sum_{\ell=t-1}^{\infty} \mathbb{P}_f\left(\min_{\bar{f}\in\overline{\mathcal{M}}_p^F(f_t^F)} L_\ell^{f_\ell^F,\bar{f}} < \beta^F\right) \\
=& \sum_{\ell=t-1}^{\infty} \mathbb{P}_f\left(\min_{\bar{f}\in\overline{\mathcal{M}}_p^F(f_\ell^F)} L_\ell^{f_\ell^F,\bar{f}} < \beta^F \text{ and } f_\ell^F = f\right) \\
&+ \sum_{\ell=t-1}^{\infty} \mathbb{P}_f\left(\min_{\bar{f}\in\overline{\mathcal{M}}_p^F(f_\ell^F)} L_\ell^{f_\ell^F,\bar{f}} < \beta^F \text{ and } f_\ell^F \neq f\right) \\
\leq& \sum_{\ell=t-1}^{\infty} \mathbb{P}_f\left(\min_{\bar{f}\in\overline{\mathcal{M}}_p^F(f)} L_\ell^{f,\bar{f}} < \beta^F\right) + \sum_{\ell=t-1}^{\infty} \mathbb{P}_f(f_\ell^F \neq f) \\
\leq& \sum_{\ell=t-1}^{\infty} \mathbb{P}_f\left(L_\ell^{f,\bar{f}} < \beta^F \text{ for some } \bar{f} \in \overline{\mathcal{M}}_p^F(f)\right) + \sum_{\ell=t-1}^{\infty} \mathbb{P}_f(f_\ell^F \neq f) \\
\leq& \sum_{\bar{f}\in\overline{\mathcal{M}}_p^F(f)} \sum_{\ell=t-1}^{\infty} \mathbb{P}_f\left(L_\ell^{f,\bar{f}} < \beta^F\right) + \sum_{\ell=t-1}^{\infty} \mathbb{P}_f(f_\ell^F \neq f).
\end{aligned}$$

We claim that it suffices to construct a sequence $\{\tilde{\rho}_t\}_{t=0}^{\infty}$, independent of δ, such that the following two conditions hold:

$$\begin{aligned}
&\sum_{t=1}^{\infty}\sum_{\ell=t-1}^{\infty} \tilde{\rho}_\ell < \infty, \\
&\mathbb{P}_f\left(L_t^{f,\bar{f}} < \beta^F\right) \leq \tilde{\rho}_t, \text{ for every } \delta \in (0,1), t \geq M(\delta) \text{ and } \bar{f} \in \overline{\mathcal{M}}_p^F.
\end{aligned} \tag{A.12}$$

To see why, recall that our goal is find $\{\rho_t\}_t$ such that (i) $\sum_t \rho_t < \infty$ and (ii) $\mathbb{P}_f(\tau \geq t) \leq \rho_t$ for all $\delta \in (0,1)$ and $t \geq M(\delta)$. Given construction of $\{\tilde{\rho}_t\}_{t=0}^{\infty}$, it is easy to verify that $\rho_t := \sum_{\bar{f}\in\overline{\mathcal{M}}_p^F(f)} \sum_{\ell=t-1}^{\infty} \tilde{\rho}_t + \sum_{\ell=t-1}^{\infty} \mathbb{P}_f(f_\ell^F \neq f)$ satisfies our needs. In fact, the verification follows from the development above as well as the following two facts: (i) $\overline{\mathcal{M}}_p^F(f)$ is a finite

set; and (ii) $\sum_{t=1}^{\infty}\sum_{\ell=t-1}^{\infty}\mathbb{P}_f(f_\ell^F \neq f) \leq \sum_{t=1}^{\infty}\sum_{\ell=t-1}^{\infty} Ce^{-\epsilon\sqrt{\ell}} < +\infty$ due to Lemma 8.

Part 2. We estimate the log likelihood ratio process $L_t^{f,\bar{f}}$ to help us construct $\{\tilde{\rho}_t\}_{t=0}^{\infty}$.

Following a similar idea in Chernoff (1959) (Lemma 2 as well as Footnote 7), we may separate the likelihood ratio process into three parts: (i) the "noise" part from the choices (conditional on the sequence of display sets); (ii) the "noise" part from the randomness of displaying sets (since the algorithm decides which display set to offer at each epoch based on historical data and possible randomization); and (iii) the "deterministic" part (which captures the long-run average growth rate of the process). Formally, we introduce the one-shot log-likelihood ratio function $L^{f,\bar{f}}(X,S) : (X,S) \mapsto \log\frac{f(X|S)}{\bar{f}(X|S)}$. Also, let us write $\lambda_*^F = \lambda_*^F(f)$ for shorthand notation. Observe that

$$\begin{aligned} & L_t^{f,\bar{f}} \\ &= \sum_{\ell=1}^{t} L^{f,\bar{f}}(X_\ell, S_\ell) \quad = \underbrace{\sum_{\ell=1}^{t}\left[L^{f,\bar{f}}(X_\ell,S_\ell) - D_{S_\ell}\left(f||\bar{f}\right)\right]}_{A} + \underbrace{\sum_{\ell=1}^{t}\left[D_{S_\ell}\left(f||\bar{f}\right) - D_{\lambda_*^F}\left(f||\bar{f}\right)\right]}_{B} \\ &+ \underbrace{t\cdot D_{\lambda_*^F}\left(f||\bar{f}\right)}_{C} \end{aligned}$$

Here, Part A corresponds to the noise part from the choices; Part B corresponds to the noise part from display sets; and Part C corresponds to the deterministic part. We will show that both Part A and B diverge sub-linearly in time (in the sense that the tail probabilities $\mathbb{P}_f(A \leq -\varepsilon t)$ and $\mathbb{P}_f(B \leq -\varepsilon t)$ decay fast in t for all $\varepsilon > 0$) while Part C grows at least as fast as the deterministic linear function $I_*^F(f)\,t$.

We claim that Part A is a $\mathbb{P}_f$-martingale with bounded difference, so that it diverges

sublinearly. We first verify the martingale property. Observe that for every $t \geq 0$,

$$\begin{aligned}
&\mathbb{E}_f\left[L^{f,\bar{f}}(X_{t+1}, S_{t+1})|\mathcal{F}_t\right] \\
=&\mathbb{E}_f\left[\mathbb{E}_f\left[\log\left(\frac{f(X_{t+1}|S_{t+1})}{\bar{f}(X_{t+1}|S_{t+1})}\right)\Big|\mathcal{F}_t, S_{t+1}\right]\Big|\mathcal{F}_t\right] && \text{[tower property]} \\
=&\mathbb{E}_f\left[\mathbb{E}_f\left[\log\left(\frac{f(X_{t+1}|S_{t+1})}{\bar{f}(X_{t+1}|S_{t+1})}\right)\Big|S_{t+1}\right]\Big|\mathcal{F}_t\right] && \text{[independence of } X_{t+1} \text{ conditional on } S_{t+1}\text{]} \\
=&\mathbb{E}_f\left[D_{S_{t+1}}\left(f||\bar{f}\right)\Big|\mathcal{F}_t\right].
\end{aligned}$$

That implies that $\mathbb{E}_f\left[L^{f,\bar{f}}(X_{t+1}, S_{t+1}) - D_{S_{t+1}}\left(f||\bar{f}\right)|\mathcal{F}_t\right] = 0$, for every $t \geq 0$. Hence Part A is a $\mathbb{P}_f$-martingale. It has bounded difference, because for every $S \in \mathcal{S}$ and $k \in S$,

$$\log\frac{f(k|S)}{\bar{f}(k|S)} \leq \overline{\mathcal{L}} := \max_{S' \in \mathcal{S}; k' \in S', f', \bar{f}' \in \mathcal{M}_p^F} \log\frac{f'(k'|S')}{\bar{f}'(k'|S')} \overset{\text{(A-1)}}{<} +\infty.$$

That means $D_S\left(f||\bar{f}\right)$ is uniformly bounded by $\overline{\mathcal{L}}$ as well. Part A diverges sublinearly in the sense that due to Azuma's inequality (Chung and Lu, 2006), for every $\epsilon_2 > 0$

$$\mathbb{P}_f(A \leq -\epsilon_2 t) \leq \exp\left(\frac{-\epsilon_2^2 t}{2\overline{\mathcal{L}}^2}\right). \tag{A.13}$$

To estimate Part B, recall $\hat{\tau} = \max\{t : f_t^F \neq f\}$, the last time the estimated ranking f_t^F differs from f. In other words, for all $t \geq \hat{\tau} + 1$, $f_t^F = f$. For the sake of analysis, let $\{\tilde{S}_\ell\}_\ell$ be a sequence of i.i.d. $\mathcal{S}$-valued random variables with distribution λ_*^{OA} such that $\tilde{S}_\ell = S_\ell$ for all $t \geq \ell \geq \hat{\tau} + 1$. We may write Part B as

$$\begin{aligned}
B = &\underbrace{\sum_{\ell=1}^{t}\left[D_{\tilde{S}_\ell}\left(f||\bar{f}\right) - D_{\lambda_*^F}\left(f||\bar{f}\right)\right]}_{B_1} + \underbrace{\sum_{\ell=1}^{\hat{\tau}\wedge t}\left[D_{S_\ell}\left(f||\bar{f}\right) - D_{\tilde{S}_\ell}\left(f||\bar{f}\right)\right]}_{B_2} + \\
&\underbrace{\sum_{\ell:\hat{\tau}+1\leq\ell\leq t, \sqrt{\ell}\in\mathbb{Z}}\left[D_{S_\ell}\left(f||\bar{f}\right) - D_{\tilde{S}_\ell}\left(f||\bar{f}\right)\right]}_{B_3}
\end{aligned}$$

We will show that all of Parts B_1, B_2, B_3 diverge sublinearly: Part B_1 is sum of IID random variables with mean zero; both Part B_2 and B_3 take negligible fractions for the

time epochs. Specifically, B_1 is a partial sum of a sequence of i.i.d random variables $\left\{D_{\tilde{S}_\ell}\left(f||\bar{f}\right) - D_{\lambda_*^F}\left(f||\bar{f}\right)\right\}_\ell$, where $\tilde{S}_\ell$ is a $\mathcal{S}$-valued random variable with distribution λ_*^F. Hence B_1 is a martingale with (uniformly) bounded differences $2\overline{\mathcal{L}}$. Due to Azuma's inequality, for every $\epsilon_2 > 0$,

$$\mathbb{P}_f(B_1 \leq -\epsilon_2 t) \leq \exp\left(\frac{-\epsilon_2^2 t}{2\overline{\mathcal{L}}^2}\right). \tag{A.14}$$

Moreover, again invoking the uniformly boundedness of $D_S\left(f||\bar{f}\right)$, $B_2 \geq \sum_{\ell=1}^{\hat{\tau}\wedge t}(-2\overline{\mathcal{L}}) \geq -2\,\overline{\mathcal{L}}\,\hat{\tau}$. As a result,

$$\mathbb{P}_f(B_2 \leq -\epsilon_2 t) \leq \mathbb{P}_f(-2\,\overline{\mathcal{L}}\,\hat{\tau} \leq -\epsilon_2 t) \leq \mathbb{P}_f\Big(\hat{\tau} \geq \tfrac{\epsilon_2 t}{2\overline{\mathcal{L}}}\Big). \tag{A.15}$$

We may also obtain a bound for Part B_3, which also diverges sublinearly:

$$\mathbb{P}_f(B_3 \leq -\epsilon_2 t) \leq \mathbb{P}_f(-2\overline{\mathcal{L}}\sqrt{t} \leq -\epsilon_2 t) = 0 \text{ for all } t \geq \sqrt{2\overline{\mathcal{L}}/\epsilon_2}. \tag{A.16}$$

To estimate Part C, simply observe that

$$C = D_{\lambda_*^F}\left(f||\bar{f}\right)\, t \geq \min_{f' \overline{\mathcal{M}}_p^F(f)} D_{\lambda_*^F}\left(f||f'\right)\, t = I_*^F(f)\, t. \tag{A.17}$$

Part 3. We use the estimates of $L_t^{f,\bar{f}}$, i.e. (A.13) through (A.17), to construct proper $\{\tilde{\rho}\}_{t=0}^\infty$ that satisfies (A.12).

We pick $\epsilon_2 := \frac{\epsilon I_*}{8(1+\epsilon)}$, $C_F := \log(|\mathcal{M}_p^F|)$, and define

$$\tilde{\rho}_t = \begin{cases} 1, & t \leq \frac{2(1+\epsilon)C_F}{\epsilon I_*} \vee \sqrt{2\overline{\mathcal{L}}/\epsilon_2} \\ \mathbb{P}_f\Big(\hat{\tau} \geq \frac{\epsilon_2 t}{2\overline{\mathcal{L}}}\Big) + 2\exp\left(\frac{-\epsilon_2^2 t}{2\overline{\mathcal{L}}^2}\right), & t > \frac{2(1+\epsilon)C_F}{\epsilon I_*} \vee \sqrt{2\overline{\mathcal{L}}/\epsilon_2}. \end{cases} \tag{A.18}$$

Due to Lemma 8, the construction of $\tilde{\rho}_t$ is independent of δ. Let us verify the first line in

(A.12):

$$\begin{aligned}
&\sum_{T=1}^{\infty}\sum_{t=T-1}^{\infty}\tilde{\rho}_t=\sum_{t=0}^{\infty}(t+1)\tilde{\rho}_t\\
\leq&\sum_{t=1}^{\frac{2(1+\epsilon)C_F}{\epsilon I_*}\vee\sqrt{2\overline{\mathcal{L}}/\epsilon_2}}(t+1)+\sum_{t=1}^{\infty}(t+1)\mathbb{P}_f\Big(\hat{\tau}\geq\tfrac{\epsilon_2 t}{2\overline{\mathcal{L}}}\Big)+\sum_{t=1}^{\infty}(t+1)\cdot 2\exp\left(\frac{-\epsilon_2^2}{2\overline{\mathcal{L}}^2}\cdot t\right)
\end{aligned}$$

Of the three items on the right hand side of the inequality above, the first term is a finite sum; the second term is finite because of finite second moment of $\hat{\tau}$ in Lemma 8 as well as Lemma 7; and the third item is finite because $\exp\left(\frac{-\epsilon_2^2}{2\overline{\mathcal{L}}^2}\cdot t\right)$ decays exponentially fast in t.

Let us verify the second line in (A.12). Pick any $\delta\in(0,1), t\geq M(\delta)$ and $\bar{f}\in\overline{\mathcal{M}}_p^F$. Assume that $t\geq\frac{2(1+\epsilon)C_F}{\epsilon I_*^F(f)}\vee\sqrt{2\overline{\mathcal{L}}/\epsilon_2}$, on top of $t\geq M(\delta)$, without loss of generality in light of (A.18).

$$\begin{aligned}
&\mathbb{P}_f\Big(L_t^{f,\bar{f}}<\beta^F\Big)\\
=&\mathbb{P}_f(A+B_1+B_2+B_3+C<\beta^F)\\
\leq&\mathbb{P}_f\left(A+B_1+B_2+B_3+tI_*^F(f)<C_F+\log\tfrac{1}{\delta}\right) && \text{[due to (A.17)]}\\
\leq&\mathbb{P}_f\left(A+B_1+B_2+B_3+tI_*^F(f)<C_F+\tfrac{I_*^F(f)}{1+\epsilon}t\right) && [t\geq M(\delta)=\tfrac{1+\epsilon}{I_*^F(f)}\log\tfrac{1}{\delta}]\\
=&\mathbb{P}_f\left(A+B_1+B_2+B_3<C_F-\tfrac{\epsilon I_*^F(f)}{1+\epsilon}t\right)\\
\leq&\mathbb{P}_f\left(A+B_1+B_2+B_3<-\tfrac{\epsilon I_*^F(f)}{2(1+\epsilon)}t\right) && [t\geq\tfrac{2(1+\epsilon)C_F}{\epsilon I_*^F(f)}\Rightarrow C_F<\tfrac{\epsilon I_*^F(f)}{2(1+\epsilon)}t]\\
\leq&\mathbb{P}_f\left(A+B_1+B_2+B_3<-4\epsilon_2 t\right) && [\epsilon_2\leq\tfrac{\epsilon I_*^F(f)}{8(1+\epsilon)}]\\
\leq&\mathbb{P}_f\left(A<-\epsilon_2 t\right)+\mathbb{P}_f\left(B_1<-\epsilon_2 t\right)+\mathbb{P}_f\left(B_2<-\epsilon_2 t\right)\\
&+(B_3<-\epsilon_2 t)\\
\leq&\exp\left(\frac{-\epsilon_2^2 t}{2\overline{\mathcal{L}}^2}\right)+\exp\left(\frac{-\epsilon_2^2 t}{2\overline{\mathcal{L}}^2}\right)+\mathbb{P}_f\Big(\hat{\tau}\geq\tfrac{\epsilon_2 t}{2\overline{\mathcal{L}}}\Big)+0 && \text{[due to (A.13)-(A.16); } t\geq\sqrt{2\overline{\mathcal{L}}/\epsilon_2}]\\
=&\mathbb{P}_f\Big(\hat{\tau}\geq\tfrac{\epsilon_2 t}{2\overline{\mathcal{L}}}\Big)+2\exp\left(\frac{-\epsilon_2^2 t}{2\overline{\mathcal{L}}^2}\right)=\tilde{\rho}_t
\end{aligned}$$

Hence $\{\tilde{\rho}_t\}_t$ defined in Equation (A.18) satisfies Equation (A.12), and the proof is finished.

■

A.3 Proof of Theorem 3

A.3.1 Preliminaries

With a slight abuse of notation, let us first replace the optimization problem $\inf_{f\in\mathcal{M}_p}\sup_{\lambda\in\Delta(\mathcal{S})}\inf_{\bar{f}\in\overline{\mathcal{M}}_p(f)} D_\lambda\left(f||\bar{f}\right)$ with $\min_{f\in\mathcal{M}_p}\max_{\lambda\in\Delta(\mathcal{S})}\min_{\bar{f}\in\overline{\mathcal{M}}_p(f)} D_\lambda\left(f||\bar{f}\right)$ w.l.o.g based on the following observations: first, we study a relaxation of the original problem by taking the closures of $\mathcal{M}_p$ and $\overline{\mathcal{M}}_p(f)$ respectively, which are compact sets. In the relaxed problem, all the "inf" and "sup" can be replaced with "min" and "max" respectively because of the continuity of our objective function. Second, we may verify (ex-post) that our proposed solution (to the relaxed optimization problem) are interior points, i.e., feasible in the original optimization problem.

For a given ranking $\sigma\in\Sigma$, let us define $\mathcal{M}_p(\sigma)$ to be the subset of p-Separable preferences that are consistent with σ, that is, $\mathcal{M}_p(\sigma) := \{f\in\mathcal{M}_p\colon \sigma_f=\sigma\}$. We also define $\overline{\mathcal{M}}_p(\sigma) := \{f\in\mathcal{M}_p\colon \sigma_f(1)\neq\sigma(1)\}$. For $k\in[K]$, let $\Sigma_k := \{\sigma\in\Sigma\colon \sigma(k)=1\}$ be the set of rankings that rank version k as the top-ranked version. For each k, we distinguish one ranking $\hat{\sigma}_k\in\Sigma_k$ that satisfies

$$\hat{\sigma}_k(i) = \begin{cases} i+1 & \text{if } i=1,\dots,k-1 \\ 1 & \text{if } i=k \\ i & \text{if } i=k+1,\dots,K. \end{cases} \tag{A.19}$$

Recall (see footnote 5) that the preference f^{OA} can be identified up to permutations. So to simplify our notation, and without loss of generality, let us assume that the consensus ranking is equal to the identity ranking, i.e., $\sigma_* := (1,2,\dots,K)$ and let us compute f^{OA} with respect to this ranking.

$$\min_{f\in\mathcal{M}_p(\sigma_*)}\max_{\lambda\in\Delta(\mathcal{S})}\min_{\bar{f}\in\overline{\mathcal{M}}_p(\sigma_*)} D_\lambda\left(f||\bar{f}\right). \tag{A.20}$$

A.3.2 Main Body of the Proof

In what follows, we will derive an explicit characterization of f^{OA} by mean of two results that we have stated as lemmas. Our first intermediate result establishes a dominance structure among the rankings in Σ_k.

Lemma 9. *For any $f^* \in \mathcal{M}_p(\sigma_*)$, $\sigma_k \in \Sigma_k$ and $f \in \mathcal{M}_p(\sigma_k)$ there exists a $\hat{f} \in \mathcal{M}_p(\hat{\sigma}_k)$ such that for any $\lambda \in \Delta(\mathcal{S})$*

$$D_\lambda\left(f^*||\hat{f}\right) \leq D_\lambda\left(f^*||f\right).$$

An important implication of Lemma 9 is that it allows us to simplify the inner minimization in the definition of f^{OA} in (A.20). Indeed, instead of minimizing over the set $\overline{\mathcal{M}}_p(\sigma_*)$ we can conduct this minimization over the much smaller set $\bigcup_{k=2}^K \mathcal{M}_p(\hat{\sigma}_k)$. In other words, the optimization problem (A.20) can be rewritten as

$$\min_{f\in\mathcal{M}_p(\sigma_*)} \max_{\lambda\in\Delta(\mathcal{S})} \min_{\bar{f}\in\bigcup_{k=2}^K \mathcal{M}_p(\hat{\sigma}_k)} D_\lambda\left(f||\bar{f}\right),$$

or equivalently

$$\min_{f\in\mathcal{M}_p(\sigma_*)} \max_{\lambda\in\Delta(\mathcal{S})} \min_{k\in\{2,\ldots,K\}} \min_{\bar{f}\in\mathcal{M}_p(\hat{\sigma}_k)} D_\lambda\left(f||\bar{f}\right). \tag{A.21}$$

The reason to include explicitly the minimization over k is motivated by our next result.

Lemma 10. *For any $k \in \{2,\ldots,K\}$ and any $S \in \mathcal{S}$ the optimization problem*

$$\min_{f\in\mathcal{M}_p(\sigma_*)} \min_{\bar{f}\in\mathcal{M}_p(\hat{\sigma}_k)} D_S\left(f||\bar{f}\right) \tag{A.22}$$

admits the following solution:

$$f^*(X|S) = f^{\text{OA}}_{\sigma_*}(X|S) \quad \textit{and} \quad \bar{f}^*(X|S) = f^{\text{OA}}_{\hat{\sigma}_k}(X|S) \qquad X \in S.$$

We will show that $f^{\text{OA}}_{\sigma_*}$ in Lemma 10, which is independent of both k and S, also solves

(the outer maximization of) (A.20). We do so by coming up with variations of (A.20) below:

$$
\begin{aligned}
&\underline{L} \\
&:= \max_{\lambda\in\Delta(\mathcal{S})} \min_{k\in\{2,\ldots,K\}} D_\lambda\left(f^{\mathrm{OA}}_{\sigma_*}||f^{\mathrm{OA}}_{\hat{\sigma}_k}\right) \\
&= \max_{\lambda\in\Delta(\mathcal{S})} \min_{k\in\{2,\ldots,K\}} \min_{f\in\mathcal{M}_p(\sigma_*)} \min_{\bar{f}\in\mathcal{M}_p(\hat{\sigma}_k)} D_\lambda\left(f||\bar{f}\right) && \text{[Lemma 10]} \\
&= \max_{\lambda\in\Delta(\mathcal{S})} \min_{f\in\mathcal{M}_p(\sigma_*)} \min_{k\in\{2,\ldots,K\}} \min_{\bar{f}\in\mathcal{M}_p(\hat{\sigma}_k)} D_\lambda\left(f||\bar{f}\right) && \text{[swapping minimization]} \\
&\leq \min_{f\in\mathcal{M}_p(\sigma_*)} \max_{\lambda\in\Delta(\mathcal{S})} \min_{k\in\{2,\ldots,K\}} \min_{\bar{f}\in\mathcal{M}_p(\hat{\sigma}_k)} D_\lambda\left(f||\bar{f}\right) && \text{[Max-Min inequality; see (a) below]} \\
&\leq \min_{f\in\mathcal{M}_p(\sigma_*)} \min_{k\in\{2,\ldots,K\}} \min_{\bar{f}\in\mathcal{M}_p(\hat{\sigma}_k)} \max_{\lambda\in\Delta(\mathcal{S})} D_\lambda\left(f||\bar{f}\right) && \text{[Max-Min inequality]} \\
&= \min_{k\in\{2,\ldots,K\}} \min_{f\in\mathcal{M}_p(\sigma_*)} \min_{\bar{f}\in\mathcal{M}_p(\hat{\sigma}_k)} \max_{\lambda\in\Delta(\mathcal{S})} D_\lambda\left(f||\bar{f}\right) && \text{[swapping minimization]} \\
&= \min_{k\in\{2,\ldots,K\}} \max_{\lambda\in\Delta(\mathcal{S})} \min_{f\in\mathcal{M}_p(\sigma_*)} \min_{\bar{f}\in\mathcal{M}_p(\hat{\sigma}_k)} D_\lambda\left(f||\bar{f}\right) && \text{[see (b) below]} \\
&= \min_{k\in\{2,\ldots,K\}} \max_{\lambda\in\Delta(\mathcal{S})} D_\lambda\left(f^{\mathrm{OA}}_{\sigma_*}||f^{\mathrm{OA}}_{\hat{\sigma}_k}\right) =: \overline{U} && \text{[Lemma 10]}
\end{aligned}
$$

In the derivations above, note that (a) is a restatement of Problem (A.21), which is equivalent to Problem (A.20). Part (b) is because of the following two facts: (i) the function $D_\lambda\left(f||\bar{f}\right)$ is jointly convex in $(f,\bar{f})$ and linear in λ; and (ii) the domains $\mathcal{M}_p(\sigma_*)$ and $\mathcal{M}_p(\hat{\sigma}_k)$ are both convex. Finally, it is easy to see that $\underline{L} = \overline{U}$ because through standard arguments in finite linear programming. Hence all of the optimization problems above are equivalent. ■

A.3.3 Proof of Auxiliary Lemmas

Proof of Lemma 9. For any $S \in \mathcal{S}$, we will show that $D_S\left(f^*||\hat{f}\right) \leq D_S\left(f^*||f\right)$. Let $S = \{X_1, X_2, \ldots, X_s\}$ and let us label the versions in S such that $\sigma_*(X_i) < \sigma_*(X_{i+1})$. Also, for notational convenience, let $f^*_i = f^*(X_i|S)$ and $f_i = f(X_i|S)$ for $i = 1, \ldots, s$. Note that because of our labeling, we have $f^*_1 \geq f^*_2 \geq \cdots \geq f^*_s$. On the other hand, let $\sigma_k^{-1}(\cdot|S)$ be the inverse of the restriction of σ to S and define the permutation $\{j_i\}_{i=1}^s$ of the element in $[s]$ in such a way that $\sigma_k^{-1}(i|S) = X_{j_i}$. It follows that $f_{j_1} \geq f_{j_2} \geq \cdots \geq f_{j_s}$.

Let us define the preference $\hat{f}$ in the lemma. We identify two cases:

(i) If $k \in S$ then $\hat{f}_{j_1} = f_{j_1}$, $\hat{f}_i = f_{j_{i+1}}$ for $i < j_1$ and $\hat{f}_i = f_{j_i}$ for $i > j_1$.

(ii) If $k \notin S$ then $\hat{f}_i = f_{j_i}$ for $i = 1, \ldots, s$.

It is not hard to see that this definition of $\hat{f}$ is consistent with the requirement $\hat{f} \in \mathcal{M}_p(\hat{\sigma}_k)$. Now, the condition $D_S\left(f^*||\hat{f}\right) \leq D_S\left(f^*||f\right)$ is equivalent to

$$\sum_{i=1}^{s} f_i^* \ln(\hat{f}_i) \geq \sum_{i=1}^{s} f_i^* \ln(f_i).$$

This inequality follows from a straightforward application of the rearrangement theorem and the following three facts: (1) the sequence $\{f_i^*\}_{i=1}^s$ is nonincreasing in i; (2) the sequence $\{\ln(\hat{f}_i)\}_{i=1}^s$ is a rearrangement of the sequence $\{\ln(f_i)\}_{i=1}^s$; and (3) the sequence $\{\ln(\hat{f}_i)\}_{i=1}^s$ is either nonincreasing in i if $k \notin S$ or is nonincreasing in i after excluding the j_1^{th} term at which the two sequences coincide if $k \in S$ (since $\hat{f}_{j_1} = f_{j_1}$ in this case). ■

Proof of Lemma 10. If $k \notin S$ then by the definition of $\hat{\sigma}_k$ it follows that $\sigma_*(X|S) = \hat{\sigma}_k(X|S)$ for all $X \in S$. Thus, the proposed solution in the lemma satisfies $f(X|S) = \bar{f}(X|S)$ for all $X \in S$ and so $D_S\left(f||\bar{f}\right) = 0$, which is trivially optimal since the KL divergence is always nonnegative.

Suppose now that $k \in S$. To ease notation, let us assume that the set $S = \{1, 2, \ldots, s\}$ with $k \leq s$. Define the permutation matrix $M \in \{0,1\}^{s \times s}$ such that $M(i, i+1) = 1$ for $i = 1, \ldots, k-1$, $M(k, 1) = 1$ and $M(i, i) = 1$ for $i = k+1, \ldots s$, with all other entries $M_{ij} = 0$. We let $x_i := f(X|S)$ and $y_i = \bar{f}_i(X|S)$ for all $i = 1, \ldots, s$. We note that the requirement $f \in \mathcal{M}_p(\sigma_*)$ implies that $x_{i+1} \leq p\, x_i$ for $i = 1, \ldots, s-1$. On the other hand, the requirement that $\bar{f} \in \mathcal{M}_p(\hat{\sigma}_k)$ implies there exists a probability vector $z \in \mathbb{R}^s$ such that $y = Mz$ and $z_{i+1} \leq p\, z_i$ for $i = 1, \ldots, s-1$. It follows that the optimization problem in (A.22) can be rewritten as the following convex optimization problem (convexity follows the fact that the KL divergence is a convex function on the pair of probability distributions

(x, y)):

$$\begin{aligned} & \min_{x,z} \sum_{i=1}^{s} x_i \Big[\ln(x_i) - \ln\Big(\sum_{j=1}^{s} M_{ij}\, z_j \Big) \Big] \\ \text{subject to} \quad & x_{i+1} \le p\, x_i, \qquad i = 1, \ldots, s-1 \\ & z_{i+1} \le p\, z_i, \qquad i = 1, \ldots, s-1 \\ & \sum_i x_i = \sum_i z_i = 1 \\ & x, z \ge 0. \end{aligned} \tag{P}$$

With the convention that $\lambda_0 = \beta_0 = \lambda_s = \beta_s = 0$, we define the Lagrange function

$$\begin{aligned} & \mathscr{L}(x, z; \zeta; \eta; \theta; \gamma) \\ = & \sum_{i=1}^{s} \left[x_i \ln(x_i) - x_i \ln\left(\sum_{j=1}^{s} M_{ij}\, z_j \right) + \zeta_i \left(x_{i+1} - p x_i \right) + \eta_i \left(z_{i+1} - p\, z_i \right) + \theta x_i + \gamma z_i \right]. \end{aligned}$$

Noticing that (P) is a convex optimization problem over linear constraints, the following KKT conditions are guarantees of optimality: for all $i = 1, \ldots, s$,[1]

$$\begin{aligned} \texttt{Stationarity in } x: \quad & \frac{\partial \mathscr{L}}{\partial x_i} = \ln(x_i) + 1 - \ln\left(\sum_{\ell=1}^{s} M_{i\ell}\, z_\ell \right) + \zeta_{i-1} - p\, \zeta_i + \theta = 0 \\ \texttt{Stationarity in } z: \quad & \frac{\partial \mathscr{L}}{\partial z_i} = -\sum_{\ell=1}^{s} x_\ell\, M_{\ell i}/z_i + \eta_{i-1} - p\, \eta_i + \gamma = 0 \\ \texttt{Complementary Slackness}: \quad & \zeta_i \left(x_{i+1} - p\, x_i \right) = 0 \quad \text{and} \quad \eta_i \left(z_{i+1} - p\, z_i \right) = 0 \\ \texttt{Primal Feasibility}: \quad & x_{i+1} \le p\, x_i, \quad z_{i+1} \le p\, z_i, \quad x_i \ge 0, \quad z_i \ge 0, \\ & \text{and} \quad \sum_i x_i = \sum_i z_i = 1 \\ \texttt{Dual Feasibility}: \quad & \zeta_i \ge 0 \quad \text{and} \quad \eta_i \ge 0. \end{aligned}$$

1. Implicitly, we are setting the dual variables for the constraint "$x, z \ge 0$" to zero.

Let us revert our change of variable and set $y^* = M\, z^*$ as well as $\tilde{y}^* = M^T\, x^*$. That is:

$$y_i^* = \begin{cases} z_{i+1}^* & \text{if } i = 1, \dots, k-1 \\ z_1^* & \text{if } i = k \\ z_i^* & \text{if } i = k+1, \dots, s \end{cases} \quad \text{and} \quad \tilde{y}_i^* = \begin{cases} x_i^* & \text{if } i = 1 \\ x_{i-1}^* & \text{if } i = 2, \dots, k \\ x_i^* & \text{if } i = k+1, \dots, s. \end{cases}$$

In addition, given $i \in [s]$, let us introduce $\Lambda_i = \sum_{\ell=1}^{i} p^{\ell-1} = (1-p^i)/(1-p)$ for shorthand notations. We postulate the following solution to the KKT conditions

$$\begin{aligned} x_i^* &= z_i^* = p^{i-1}/\Lambda_s \\ \theta^* &= -\left(\sum_{\ell=1}^{s} p^{\ell-1}\, \ln \frac{x_\ell^*}{y_\ell^*} \right) /\Lambda_s - 1 \\ \zeta_i^* &= \tfrac{1}{p^i} \left[\sum_{\ell=1}^{i} p^{\ell-1}\, \ln \frac{x_\ell^*}{y_\ell^*} + (1+\theta^*)\Lambda_i \right] \\ \gamma^* &= \left(\sum_{\ell=1}^{s} p^{\ell-1}\, \frac{\tilde{y}_\ell^*}{z_\ell^*} \right) /\Lambda_s \\ \eta_i^* &= \tfrac{1}{p^i} \left[\gamma^* \Lambda_i - \sum_{\ell=1}^{i} p^{\ell-1}\, \frac{\tilde{y}_\ell^*}{z_\ell^*} \right]. \end{aligned}$$

One can verify that, by construction, the $\{x_i^*\}$ satisfy $x_{i+1}^* = p\, x_i^*$ for all $i = 1, \dots, s-1$ and $\sum_i x_i^* = 1$; and the same is true for the $\{z_i^*\}$. Hence, `Complementary Slackness` and `Primal Feasibility` are directly satisfied. Similarly, it is not hard to see that the value of θ^*, ζ_i^*, γ^* and η_i^* are chosen so that both `Stationarity` conditions are satisfied. Hence, we only need to check `Dual Feasibility` for ζ_i and η_i.

Let us first verify that $\zeta_i^* \geq 0$. It is equivalent to

$$\frac{1}{\Lambda_i} \sum_{\ell=1}^{i} p^{\ell-1}\, \ln \frac{x_\ell^*}{y_\ell^*} \geq \frac{1}{\Lambda_s - \Lambda_i} \sum_{\ell=i+1}^{s} p^{\ell-1} \ln \frac{x_\ell^*}{y_\ell^*}. \tag{A.23}$$

Invoking the expressions for x^* and y^*, we notice that

$$p^{\ell-1} \ln \frac{x_\ell^*}{y_\ell^*} = \begin{cases} p^{\ell-1} \ln \frac{1}{p} & \text{if } \ell = 1, \dots, k-1 \\ -p^{k-1}(k-1) \ln \frac{1}{p} & \text{if } \ell = k \\ 0 & \text{if } \ell = k+1, \dots, s. \end{cases}$$

Thus for $i < k$ condition (A.23) becomes

$$\ln \frac{1}{p} \geq \frac{\ln \frac{1}{p}}{\Lambda_s - \Lambda_i} \left[p^i + \cdots + p^{k-2} - (k-1)p^{k-1} \right].$$

Noticing that the term of the left is positive and that on the right is negative, one can check this inequality holds. For $i \geq k$ condition (A.23) becomes

$$\frac{\ln \frac{1}{p}}{\Lambda_i} \left[1 + \cdots + p^{k-2} - (k-1)p^{k-1} \right] \geq 0,$$

which is also satisfied.

Let us then verify that $\eta_i^* \geq 0$. Invoking the expressions for $\tilde{y}^*$ and z^*, we notice that

$$p^{\ell-1} \frac{\tilde{y}_\ell^*}{z_\ell^*} = \begin{cases} p^{k-1} & \text{if } \ell = 1 \\ p^{\ell-2} & \text{if } \ell = 2, \dots, k \\ p^{\ell-1} & \text{if } \ell = k+1, \dots, s. \end{cases}$$

In particular, the vector $\left(p^{\ell-1} \frac{\tilde{y}_\ell^*}{z_\ell^*} \right)_{\ell=1,\dots,s}$ is a permutation of the vector $\left(p^{\ell-1} \right)_{\ell=1,\dots,s}$. As a result, $\gamma^* = 1$ and $\eta_i^* \geq 0$ is equivalent to

$$\sum_{\ell=1}^{i} p^{\ell-1} \geq \sum_{\ell=1}^{i} p^{\ell-1} \frac{\tilde{y}_\ell^*}{z_\ell^*}.$$

The inequality is satisfied because the vector $\left(p^{\ell-1} \right)_{\ell=1,\dots,s}$ is a decreasing sequence. ■

A.4 Proof of Theorem 4

A.4.1 Preliminaries

Let us introduce some notation. Let us introduce $d_\lambda(\sigma) := D_\lambda\left(f^{\text{OA}}_{\sigma_*} || f^{\text{OA}}_{\sigma}\right)$, for every $\sigma \in \Sigma$, a shorthand notation for the "distance" from ranking σ to the identity mapping σ_*. Invoking Corollary 1, a key step is to observe that (2.5) is equivalent to
$\max_{\lambda \in \Delta(\mathcal{S})} \min_{\bar{f} \in \overline{\mathcal{M}}^{\text{OA}}_p(f)} D_\lambda\left(f || \bar{f}\right)$, which is further equivalent to the following LP:

$$\begin{aligned} \max_{\lambda, u} \quad & u \\ s.t. \quad & \sum_{S \in \mathcal{S}} d_S(\bar{\sigma}) \cdot \lambda(S) \geq u, \quad \forall \bar{\sigma} \in \overline{\Sigma}(\sigma_*) \\ & \sum_{S \in \mathcal{S}} \lambda(S) = 1 \\ & \lambda(S) \geq 0, \quad \forall S \in \mathcal{S} \end{aligned} \tag{LP-P}$$

In the expression above, we leverage that fact that each element $f \in \mathcal{M}^{\text{OA}}_p$ uniquely corresponds to a ranking $\sigma \in \Sigma$. Moreover, $\overline{\Sigma}(\sigma_*) := \{\sigma \in \Sigma : \sigma^{-1}(1) \neq 1\}$ is defined as the set of rankings which disagrees with σ_* in terms of the top-ranked item. We may also write out the dual problem of (LP-P) in the follows:

$$\begin{aligned} \min_{\mu, l} \quad & l \\ s.t. \quad & \sum_{\bar{\sigma} \in \overline{\Sigma}(\sigma_*)} d_S(\bar{\sigma}) \cdot \mu(\bar{\sigma}) \leq l, \quad \forall S \in \mathcal{S} \\ & \sum_{\bar{\sigma} \in \overline{\Sigma}(\sigma_*)} \mu(\bar{\sigma}) = 1 \\ & \mu(\bar{\sigma}) \geq 0, \quad \forall \bar{\sigma} \in \overline{\Sigma}(\sigma_*). \end{aligned} \tag{LP-D}$$

The dual problem is not used in the algorithm, but important in our analysis. We also introduce some other notation. Let us define, for $n \in [K]$, $s_n = (1-p)p^{n-1}$, so that the

following equalities hold:

$$\begin{aligned}\sum_{k=1}^{n-1}(s_k - s_n)\log\frac{s_k}{s_{k+1}} &= \log\left(\frac{1}{p}\right)(1-p)\left[1+p+\cdots+p^{n-2}-(n-1)p^{n-1}\right] = \mathfrak{a}_n \\ \sum_{k=1}^{n} s_k &= (1-p)(1+p+\cdots+p^{n-1}) = 1-p^n = \mathfrak{b}_n.\end{aligned} \tag{A.24}$$

Recall from the proof of Theorem 3 that $\hat{\sigma}_m := (2, 3, \ldots, m, 1, m+1, \ldots, K)$, for $m = 2, \ldots, K$. The introduction of $\mathfrak{a}_n$ and $\mathfrak{b}_n$ in (2.6) is helpful in evaluating $d_S(\hat{\sigma}_m)$, as stated in the lemma below. The proof of all technical lemmas in this subsection are in Section A.4.3.

Lemma 11. *Given any set $S \in \mathcal{S}$ and $m = 2, \ldots, K$,*

$$d_S(\hat{\sigma}_m) = \begin{cases} \mathfrak{a}_i/\mathfrak{b}_n & \text{if } m \in S \\ 0 & \text{if } m \notin S \end{cases}, \quad \text{where} \quad i = \sigma_*(m|S) \text{ and } n = |S|.$$

Ultimately we aim to solve both (LP-P) and (LP-D) in closed form. On one hand, $\lambda_*^{\mathrm{OA}}(\cdot)$ defined in (2.7) are closely related to the primal optimal solutions to (LP-P). On the other hand, we define the constants

$$\mu_m = \begin{cases} \frac{\mathfrak{b}_2}{\mathfrak{a}_2}, & \text{if } m = 2; \\ \frac{1}{\mathfrak{a}_m}(\mathfrak{b}_m - \mathfrak{b}_{m-1}), & \text{if } m = 3, \ldots, K. \end{cases} \tag{A.25}$$

and

$$\mu^*(\bar{\sigma}) = \begin{cases} \frac{\mu_m}{\mu_2+\ldots+\mu_K}, & \text{if } \bar{\sigma} = \hat{\sigma}_m,\ m = 2, \ldots, K; \\ 0, & \text{otherwise.} \end{cases} \tag{A.26}$$

The quantities $\{\mu^*(\bar{\sigma})\}_{\bar{\sigma}}$ are closely related to the dual optimal solutions to (LP-D).

In the lemmas that follows, we will show that there exist (u^*, v^*) such that $u^* = l^*$ and $(\lambda_*^{\mathrm{OA}}, u^*)$ and (μ^*, l^*) are primal and dual feasible in (LP-P) and (LP-D) respectively. This means both problems have the same objective value. By weak duality, this implies that $(\lambda_*^{\mathrm{OA}}, u^*)$ and (μ^*, l^*) are primal and dual optimal in (LP-P) and (LP-D) respectively.

Lemma 12. $(\lambda_*^{\mathrm{OA}}, u^*)$ *is feasible in* (LP-P)*, where* $u^* := \frac{1}{\lambda_2^*+\ldots+\lambda_K^*}$.

Lemma 13. (μ^*, l^*) *is feasible in* (LP-D)*, where* $l^* := \frac{1}{\mu_2+\ldots+\mu_K}$.

Lemma 14. $u^* = l^*$.

Given any $S \in \mathcal{S}$, we define the reduced cost r of the optimal solution as

$$r(S) := \sum_{\sigma \in \overline{\Sigma}(\sigma_*)} d_S(\sigma) \cdot \mu^*(\sigma) - l^*. \tag{A.27}$$

By the proof of Lemma 13 below, we can also show the following corollary regarding the reduced costs, which guarantees the uniqueness of the optimal solution λ_*^{OA}.

Corollary 5. *For every S where $\lambda_*^{\text{OA}}(S) = 0$, $r(S) < 0$.*

A.4.2 Main Body of Proof

Proof of Theorem 4. Because of primal feasibility of $(\lambda_*^{\text{OA}}, u^*)$ in Lemma 12, dual feasibility of (μ^*, l^*) in Lemma 13, and the objective values $u^* = l^*$ due to Lemma 14, we know that λ_*^{OA} is an optimal solution to (Max-Min) by weak duality (see Dantzig, 1963). Due to Corollary 5, the reduced costs outside the support of λ_*^{OA} are strictly negative. Hence the optimal solution λ_*^{OA} is unique (see Dantzig, 1963). ■

A.4.3 Proofs of Auxiliary Lemmas.

Proof of Lemma 11. The proof of this lemma is by calculation and verification. Fix an arbitrary $m \in \{2, \ldots, K\}$ and $S = \{k_1, k_2, \ldots, k_n\} \in \mathcal{S}$, where $1 \leq k_1 < k_2 < \cdots < k_n \leq K$.

To calculate $d_S(\hat{\sigma}_m)$, it suffices to explicitly write out $f_{\sigma_*}^{\text{OA}}$ and $f_{\hat{\sigma}_m}^{\text{OA}}$, the p.m.f. of votes under ranking under σ_* and $\hat{\sigma}_m$, conditional on the display set being S. Recall that the restricted ranking is defined as $\sigma(k|S) := \sum_{i \in S} \mathbb{I}\{\sigma(i) \leq \sigma(k)\}$, for all $\sigma \in \Sigma, S \in \mathcal{S}$ and $k \in S$. In particular, $\sigma_*\left(k_j|S\right) = j$ for all $j = 1, 2, \ldots, n$. Invoking the definition of f_σ^{OA} in (2.3), we can explicitly write out the p.m.f. of votes under ranking σ_* and display set S:

$$f_{\sigma_*}^{\text{OA}}\left(k_j|S\right) = \frac{(1-p)p^{j-1}}{1-p^n} = \frac{s_j}{\mathfrak{b}_n}, \quad \forall j = 1, \ldots, n. \tag{A.28}$$

The closed form expression of $f_{\sigma_*}^{\text{OA}}$ conditional on the display set being S, however, depends on the whether m is included in S:

Case 1: If $m \notin S$, S is a subset of $[K]\backslash\{m\}$. Hence $\hat{\sigma}_m(\cdot|S) = (1, 2, \cdots, n)$. Invoking (A.28),

$$f^{\text{OA}}_{\hat{\sigma}_m}\left(k_j|S\right) = f^{\text{OA}}_{\sigma_*}\left(k_j|S\right) = s_j/\mathfrak{b}_n, \quad \forall j = 1, \ldots, n. \tag{A.29}$$

That means $d_S(\hat{\sigma}_m) = \sum_{j=1}^n f^{\text{OA}}_{\sigma_*}\left(k_j|S\right) \log \frac{f^{\text{OA}}_{\sigma_*}(k_j|S)}{f^{\text{OA}}_{\hat{\sigma}_m}(k_j|S)} = 0$.

Case 2: If $m \in S = \{k_1, k_2, \cdots, k_n\}$, there exists $i \in [n]$ so that $m = k_i$. Recall that $\sigma_*\left(k_j|S\right) = j$ for all $j = 1, 2, \ldots, n$. In particular, since $m = k_i$, $\sigma_*(m|S) = \sigma_*(k_i|S) = i$ Also, the restricted ranking $\hat{\sigma}_m(\cdot|S)$ is such that $\hat{\sigma}_m(k_i|S) = 1$, $\hat{\sigma}_m\left(k_j|S\right) = j+1$ for all $j = 1, \ldots, i-1$, and $\hat{\sigma}_m\left(k_j|S\right) = j$ for all $j = i+1, \ldots, n$. Invoking (2.3), we may write $f^{\text{OA}}_{\hat{\sigma}_m}(\cdot|S)$ in closed form below:

$$f^{\text{OA}}_{\hat{\sigma}_m}\left(k_j|S\right) = \begin{cases} s_{j+1}/\mathfrak{b}_n, & \text{if } j = 1, \ldots, i-1 \\ s_1/\mathfrak{b}_n, & \text{if } j = i \\ s_j/\mathfrak{b}_n, & \text{if } j = i+1, \ldots, n \end{cases} \tag{A.30}$$

That means

$$\begin{aligned} d_S(\hat{\sigma}_m) =& D_\lambda\left(f^{\text{OA}}_{\sigma_*} || f^{\text{OA}}_{\sigma}\right) \\ =& \sum_{j=1}^n f^{\text{OA}}_{\sigma_*}\left(k_j|S\right) \log \frac{f^{\text{OA}}_{\sigma_*}\left(k_j|S\right)}{f^{\text{OA}}_{\hat{\sigma}_m}\left(k_j|S\right)} \\ =& \sum_{j=1}^{i-1} \frac{s_j}{\mathfrak{b}_n} \log \frac{s_j}{s_{j+1}} + \frac{s_i}{\mathfrak{b}_n} \log \frac{s_i}{s_1} + \sum_{j=i+1}^{n} \frac{s_j}{\mathfrak{b}_n} \log \frac{s_j}{s_j} && \text{[(A.28) and (A.30)]} \\ =& \sum_{j=1}^{i-1} \frac{s_j}{\mathfrak{b}_n} \log \frac{s_j}{s_{j+1}} - \frac{s_i}{\mathfrak{b}_n} \left(\log \frac{s_1}{s_2} + \cdots + \log \frac{s_{i-1}}{s_i}\right) + 0 && \text{[expand } \log \tfrac{s_i}{s_1}\text{]} \\ =& \sum_{j=1}^{i-1} \frac{s_j - s_i}{\mathfrak{b}_n} \log \frac{s_j}{s_{j+1}} = \frac{\mathfrak{a}_i}{\mathfrak{b}_n}. && \text{[(A.24)]} \end{aligned}$$

By combining the two cases above, we finish the proof. ■

Proof of Lemma 12. Recall from (2.7) that we have defined λ_*^{OA} as:

$$\lambda_*^{\text{OA}}(S) = \begin{cases} \frac{\lambda_n^*}{\lambda_2^*+\ldots+\lambda_K^*} & \text{if } S = [n] \text{ for some } n \in \{2,\ldots,K\} \\ 0 & \text{otherwise.} \end{cases}$$

To verify that λ_*^{OA} is primal feasible, we break the discussion into five steps. The first two steps check the first feasibility constraint in (LP-P), and rest of the steps check the remaining feasibility constraints.

Step 1. We claim that $\sum_{S\in\mathcal{S}} d_S(\hat{\sigma}_m) \cdot \lambda_*^{\text{OA}}(S) = u^* = \frac{1}{\lambda_2^*+\ldots+\lambda_K^*}, \quad \forall m = 2,\ldots,K.$

We know that λ_*^{OA} is only positive on S of form $[n] = \{1,\ldots,n\}$, where $n = 2,\ldots,K$. Hence

$$\begin{aligned}
&\sum_{S\in\mathcal{S}} d_S(\hat{\sigma}_m) \cdot \lambda_*^{\text{OA}}(S) = \sum_{n=2}^{K} d_{[n]}(\hat{\sigma}_m) \cdot \frac{\lambda_n^*}{\lambda_2^*+\cdots+\lambda_K^*} && [\lambda_*^{\text{OA}} \text{ defined in (2.7)}] \\
\overset{(a)}{=}& \sum_{n=m}^{K} \frac{\mathfrak{a}_m}{\mathfrak{b}_n} \cdot \frac{\lambda_n^*}{\lambda_2^*+\cdots+\lambda_K^*} \\
=& \frac{\mathfrak{a}_m}{\lambda_2^*+\cdots+\lambda_K^*} \left(\sum_{n=m}^{K} \frac{\lambda_n^*}{\mathfrak{b}_n} \right) \\
\overset{(b)}{=}& \frac{\mathfrak{a}_m}{\lambda_2^*+\cdots+\lambda_K^*} \cdot \left(\frac{1}{\mathfrak{a}_m} - \frac{1}{\mathfrak{a}_{m+1}} + \frac{1}{\mathfrak{a}_{m+1}} - \frac{1}{\mathfrak{a}_{m+2}} + \cdots + \frac{1}{\mathfrak{a}_K} \right) \\
=& \frac{1}{\lambda_2^*+\cdots+\lambda_K^*}.
\end{aligned}$$

In the chain of equalities, part (a) is by linking the fact that $\sigma_*(m|[n]) = m$ for $n \geq m$ to the calculation of $d_{[n]}(\hat{\sigma}_m)$ in Lemma 11. Part (b) is due to the definition of λ_n^* in (2.6).

Step 2. We claim that $\sum_{S\in\mathcal{S}} d_S(\bar{\sigma}) \cdot \lambda_*^{\text{OA}}(S) \geq u^* = \frac{1}{\lambda_2^*+\ldots+\lambda_K^*}$, for every $\bar{\sigma} \in \overline{\Sigma}(\sigma_*)$. Step 2 equivalent to the first line of constraints in Problem (LP-P).

Given any $\bar{\sigma} \in \overline{\Sigma}(\sigma_*)$, pick $m \in \{2,\ldots,K\}$ so that $\sigma(m) = 1$. Invoking the dominance

result in Lemma 9, we know that for every $S \in \mathcal{S}$, $d_S(\hat{\sigma}_m) \leq d_S(\bar{\sigma})$. As a result, we have

$$\begin{aligned}\sum_{S\in\mathcal{S}} d_S(\bar{\sigma}) \cdot \lambda_*^{\mathrm{OA}}(S) \geq & \sum_{S\in\mathcal{S}} d_S(\hat{\sigma}_m) \cdot \lambda_*^{\mathrm{OA}}(S) \\ = & \frac{1}{\lambda_2^* + \cdots + \lambda_K^*} \qquad \text{[due to Step 2].}\end{aligned}$$

Step 3. We claim that $\sum_{S\in\mathcal{S}} \lambda_*^{\mathrm{OA}}(S) = 1$. This step is trivial by (deliberate) construction of λ_*^{OA}. Step 3 is equivalent to the second line of constraints in Problem (LP-P).

Step 4. We claim that $\lambda_*^{\mathrm{OA}}(S) \geq 0$, for all $S \in \mathcal{S}$. Step 4 is equivalent to the last line of constraints in Problem (LP-P).

Due to the definition of λ_*^{OA} in (2.7), it suffices to verify that $\lambda_n^* \geq 0$, for every $n = 2, \ldots, K$. Recall that $s_n = (1-p)p^{n-1}$ and λ_n^* is given by

$$\lambda_n^* = \begin{cases} \mathfrak{b}_n \left(\frac{1}{\mathfrak{a}_n} - \frac{1}{\mathfrak{a}_{n+1}}\right) = \frac{\mathfrak{b}_n}{\mathfrak{a}_n \mathfrak{a}_{n+1}} \cdot (\mathfrak{a}_{n+1} - \mathfrak{a}_n), & \text{if } n = 2, \ldots, K-1; \\ \frac{\mathfrak{b}_K}{\mathfrak{a}_K}, & \text{if } n = K. \end{cases}$$

Notice the following three facts from (2.6):

1. $\mathfrak{b}_n = 1 - p^n > 0$;
2. $\mathfrak{a}_n = \log\left(\frac{1}{p}\right)[1 - np^{n-1} + (n-1)p^n] > 0$;
3. $\mathfrak{a}_{n+1} - \mathfrak{a}_n = np^{n-1}(1-p)^2 > 0$.

Hence $\lambda_n^* \geq 0$, and the proof is complete. ■

Proof of Lemma 13. Recall from (A.26) that we have defined μ^* as:

$$\mu^*(\bar{\sigma}) = \begin{cases} \frac{\mu_m}{\mu_2 + \ldots + \mu_K}, & \text{if } \bar{\sigma} = \hat{\sigma}_m, m = 2, \ldots, K; \\ 0, & \text{otherwise.} \end{cases}$$

We break the discussion into three steps. Each step corresponds to one line of constraints in Problem (LP-D) respectively.

Step 1. We claim that $\sum_{\bar{\sigma}\in\overline{\Sigma}(\sigma_*)} d_S(\bar{\sigma})\cdot\mu^*(\bar{\sigma}) \le l^* = \frac{1}{\mu_2+\ldots+\mu_K}$, for every $S\in\mathcal{S}$. Step 1 is equivalent to the first line of constraints in Problem (LP-D).

Fix an arbitrary $S\in\mathcal{S}$, and suppose $|S|=n$. We know that $\mu^*(\cdot)$ is only positive on $\hat{\sigma}_m$, for $m=2,\ldots,K$. As a result,

$$\begin{aligned}
&\sum_{\bar{\sigma}\in\overline{\Sigma}(\sigma_*)} d_S(\bar{\sigma})\cdot\mu^*(\bar{\sigma}) \\
=&\sum_{m=2}^{K} d_S(\hat{\sigma}_m)\cdot\frac{\mu_m}{\mu_2+\cdots+\mu_K} && [\mu^* \text{ defined in (A.26)}] \\
=&\sum_{m=2}^{n}\frac{\mathfrak{a}_{\sigma_*(m|S)}}{\mathfrak{b}_n}\cdot\frac{\mu_m}{\mu_2+\cdots+\mu_K} && [\text{Lemma 11}] \\
\le&\sum_{m=2}^{n}\frac{\mathfrak{a}_m}{\mathfrak{b}_n}\cdot\frac{\mu_m}{\mu_2+\cdots+\mu_K} && [\text{(i) } \sigma_*(m|S)\le m\text{; (ii) } \mathfrak{a}_n\uparrow \text{ in } n\ ] \\
=&\frac{1}{\mathfrak{b}_n(\mu_2+\cdots+\mu_K)}(\mathfrak{b}_2+\mathfrak{b}_3-\mathfrak{b}_2+\cdots+\mathfrak{b}_n-\mathfrak{b}_{n-1}) && [\mu_m \text{ defined in (A.25)}] \\
=&\frac{1}{\mu_2+\cdots+\mu_K}.
\end{aligned}$$

Note that $\sigma_*(m|S)=m$ if and only if $[m]\subset S$, and $\mathfrak{a}_m$ is strictly increasing in m. Hence the inequality above becomes equality if and only if $S=[n]$. This observation leads to Corollary 5 as a direct consequence.

Step 2. We claim that $\sum_{m=2}^{K}\mu^*(\hat{\sigma}_m)=1$. This step is trivial by (deliberate) construction of μ^*. Step 2 is equivalent to the second line of constraints in Problem (LP-D).

Step 3. We claim that $\mu(\hat{\sigma}_m)>0$, for all $m=2,\ldots,K$. Step 3 is equivalent to the last line of constraints in Problem (LP-D).

For every $m\in\{2,\ldots,K\}$,

$$\mu(\hat{\sigma}_m)=\mu_m=\begin{cases}\frac{\mathfrak{b}_2}{\mathfrak{a}_2}, & \text{if } m=2;\\ \frac{1}{\mathfrak{a}_m}(\mathfrak{b}_m-\mathfrak{b}_{m-1}), & \text{if } m=3,\ldots,K\end{cases}=\begin{cases}\frac{s_1+s_2}{\mathfrak{a}_2}, & \text{if } m=2;\\ \frac{s_m}{\mathfrak{a}_m}, & \text{if } m=3,\ldots,K\end{cases}>0.$$

The strict positiveness of μ_m is due to both the strict positiveness of $\mathfrak{a}_n$ and s_n, for $n,m\in\{2,\ldots,K\}$. That finishes the proof. ■

Proof of Corollary 5. This corollary is a restatement of the remark at the end of Step 1 of the proof of Lemma 13. ■

Proof of Lemma 14. Observe that

$$\begin{aligned}\sum_{n=2}^{K-1}\lambda_n^* &= \sum_{n=2}^{K-1}\mathfrak{b}_n\left(\frac{1}{\mathfrak{a}_n}-\frac{1}{\mathfrak{a}_{n+1}}\right)+\frac{\mathfrak{b}_K}{\mathfrak{a}_K}\\ &= \frac{\mathfrak{b}_2}{\mathfrak{a}_2}+\sum_{n=3}^{K}\frac{1}{\mathfrak{a}_n}(\mathfrak{b}_n-\mathfrak{b}_{n-1}) && \text{[summation by parts]}\\ &= \sum_{n=2}^{K-1}\mu_n && \text{[due to (A.25)]}.\end{aligned}$$

As a result, $u^* = \frac{1}{\sum_{n=2}^{K-1}\lambda_n^*} = \frac{1}{\sum_{n=2}^{K-1}\mu_n} = l^*$. ■

A.5 Proofs of Proposition 1 and Corollary 2

Preliminaries. Throughout this section, the context of evaluating I_*^{OA} is clear. So let us suppress the argument and write $I_* = I_*^{\text{OA}}$ for shorthand notation.

Proof of Proposition 1. We first show that

$$I_* = (1-p)\log\left(\tfrac{1}{p}\right)\left(1+\sum_{n=2}^{K}\frac{p^{n-1}}{1+2p+\cdots+(n-1)p^{n-2}}\right)^{-1}.$$

Due to Lemma 12, 13, and 14, $I_* = \frac{1}{\lambda_2^*+\cdots+\lambda_K^*} = \frac{1}{\mu_2+\cdots+\mu_K} = \frac{1}{\frac{\mathfrak{b}_2}{\mathfrak{a}_2}+\sum_{n=3}^{K}\frac{1}{\mathfrak{a}_n}(\mathfrak{b}_n-\mathfrak{b}_{n-1})}$. Plug the values of $\mathfrak{a}_n, \mathfrak{b}_n$ in, and we have

$$\begin{aligned}\frac{1}{I_*} &= \frac{\mathfrak{b}_2}{\mathfrak{a}_2}+\sum_{n=3}^{K}\frac{1}{\mathfrak{a}_n}(\mathfrak{b}_n-\mathfrak{b}_{n-1})\\ &= \frac{\mathfrak{b}_2}{\mathfrak{a}_2}+\sum_{n=3}^{K}\frac{s_n}{\mathfrak{a}_n}\\ &= \frac{1}{\log\frac{1}{p}(1-p)}\left(1+p+\sum_{n=3}^{K}\frac{p^{n-1}}{1+2p+\cdots+(n-1)p^{n-2}}\right).\end{aligned}$$

To give simple estimates for I_*, we break the rest of discussion into two steps.

Step 1. We start with the upper bound. In fact, since $0 < p < 1$,

$$\sum_{n=3}^{K} \frac{p^{n-1}}{1+2p+\cdots+(n-1)p^{n-2}} \geq 0.$$

Hence

$$I_* = \frac{\log \frac{1}{p}(1-p)}{1+p+\sum_{n=3}^{K} \frac{p^{n-1}}{1+2p+\cdots+(n-1)p^{n-2}}} \leq \frac{\log \frac{1}{p}(1-p)}{1+p}.$$

Step 2. Next we establish the lower bound. Recall that we may rewrite I_* as $I_* = \frac{\Phi(p)}{\Psi(p)}$, where $\Phi(p) = \log \frac{1}{p}(1-p)$ and $\Psi(p) = 1+p+\sum_{n=3}^{K} \frac{p^{n-1}}{1+2p+\cdots+(n-1)p^{n-2}}$. Note that

$$\begin{aligned}
\Psi(p) &= 1 + p\left(\sum_{n=2}^{K} \frac{1}{\frac{1}{p^{n-2}}+\frac{2}{p^{n-3}}\cdots+(n-1)}\right) \leq 1 + p\left(\sum_{n=2}^{K} \frac{1}{1+2+\cdots n-1}\right) \\
&= 1 + p\left(\sum_{n=2}^{K} \frac{2}{n(n-1)}\right) = 1 + 2p\left(1 - \tfrac{1}{2} + \tfrac{1}{2} - \tfrac{1}{3} + \cdots + \tfrac{1}{K-1} - \tfrac{1}{K}\right) \\
&= 1 + 2p\left(1 - \tfrac{1}{K}\right)
\end{aligned}$$

As a result, $I_* = \frac{\Phi(p)}{\Psi(p)} \geq \frac{\log \frac{1}{p}(1-p)}{1+2p(1-1/K)}$. That finishes the proof. ■

Proof of Corollary 2. Assume $K \geq 4$. We break the discussion into two steps.

Step 1. Choose an arbitrary $n \in \{2, \ldots K-2\}$ We claim that $\lambda^*_{n+1} < \lambda^*_n$. We verify our claim by evaluating the term $\frac{\lambda^*_{n+1}}{\lambda^*_n}$ below:

$$\begin{aligned}
\frac{\lambda^*_{n+1}}{\lambda^*_n} &= \frac{\mathfrak{b}_{n+1}\frac{a_{n+2}-\mathfrak{a}_{n+1}}{a_{n+2}\mathfrak{a}_{n+1}}}{\mathfrak{b}_n\frac{a_{n+1}-\mathfrak{a}_n}{\mathfrak{a}_{n+1}\mathfrak{a}_n}} \\
&= \frac{\mathfrak{b}_{n+1}}{\mathfrak{b}_n}\frac{\mathfrak{a}_{n+2}-\mathfrak{a}_{n+1}}{\mathfrak{a}_{n+1}-\mathfrak{a}_n}\frac{\mathfrak{a}_n}{\mathfrak{a}_{n+2}} \\
&= \frac{1-p^{n+1}}{1-p^n}\ \frac{(n+1)p^n}{np^{n-1}}\ \frac{1-np^{n-1}+(n-1)p^n}{1-(n+2)p^{n+1}+(n+1)p^{n+2}} \\
&\overset{(a)}{=} \frac{(n+1)\big(p+p^2+\cdots+p^{n+1}\big)\big(1+2p+\cdots+(n-1)p^{n-2}\big)}{n\big(1+p+\cdots+p^{n-1}\big)\big(1+2p+\cdots+(n+1)p^n\big)}
\end{aligned}$$

Part (a) of the derivations above is due to the following two (algebraic) facts: For every

$n \in \mathbb{Z}_+$,

$$\begin{aligned} \sum_{i=0}^{n-1} p^i &= 1 + 2 + \cdots + p^{n-1} = \tfrac{1-p^n}{1-p} \\ \sum_{i=1}^{n} ip^{i-1} &= 1 + 2p + \cdots + np^{n-1} = \tfrac{1+p+\cdots+p^{n-1}-np^n}{1-p} = \tfrac{1-(n+1)p^n+np^{n+1}}{(1-p)^2} \end{aligned} \tag{A.31}$$

Note that $\frac{\lambda^*_{n+1}}{\lambda^*_n}$ is a ratio of two polynomials of p, namely, $\frac{\lambda^*_{n+1}}{\lambda^*_n} = \frac{(n+1)P(p)}{nQ(p)}$, where

$$\begin{aligned} P(p) :=& \big(p + p^2 + \cdots + p^{n+1}\big)\big(1 + 2p + \cdots + (n-1)p^{n-2}\big) \\ Q(p) :=& \big(1 + p + \cdots + p^{n-1}\big)\big(1 + 2p + \cdots + (n+1)p^n\big). \end{aligned}$$

Let $\{\xi_k\}_{k=1}^{2n-1}$ and $\{\nu_k\}_{k=1}^{2n-1}$ be the coefficients of $P(p)$ and $Q(p)$ respectively, so that $P(p) = \sum_{k=1}^{2n-1} \xi_k p^k$ and $Q(p) = \sum_{k=1}^{2n-1} \nu_k p^k + 1$. To show that $\lambda^*_{n+1} < \lambda^*_n$, it suffices to show that for every $k \in [2n-1]$, $(n+1)\xi_k < n\nu_k$ Observe that

$$\xi_k = \begin{cases} \frac{k(k+1)}{2}, & 1 \le k \le n-1 \\ \frac{n(n-1)}{2}, & k = n \\ \frac{(k-1)(2n-k)}{2}, & n+1 \le k \le 2n-1 \end{cases} \quad \text{and} \quad \nu_k = \begin{cases} \frac{(k+1)(k+2)}{2}, & 1 \le k \le n-1 \\ \frac{n(n+3)}{2}, & k = n \\ \frac{(k+3)(2n-k)}{2}, & n+1 \le k \le 2n-1 \end{cases}$$

Hence

$$\frac{(n+1)\xi_k}{n\nu_k} = \begin{cases} \frac{(n+1)k}{n(k+2)} \le \frac{(n+1)(n-1)}{n(n+1)} < 1, & 1 \le k \le n-1 \\ \frac{(n-1)(n+1)}{(n+3)n} = \frac{n^2-1}{n^2+3n} < 1, & k = n \\ \frac{(n+1)(k-1)}{n(k-3)} \le \frac{(n+1)(2n-2)}{n(2n-4)} = \frac{(n+1)(n-1)}{n(n-2)} < 1, & n+1 \le k \le 2n-1 \end{cases}$$

By conclusion, $\lambda^*_{n+1} < \lambda^*_n$ for every $2 \le n \le K-2$.

Step 2. We claim that $\lambda^*_{K-1} < \lambda^*_K$ (thus finishing the proof). We verify our claim by

evaluating the term $\frac{\lambda^*_{K-1}}{\lambda^*_K}$ below:

$$\begin{aligned}
\frac{\lambda^*_{K-1}}{\lambda^*_K} &= \frac{\mathfrak{b}_{K-1}\frac{\mathfrak{a}_K-\mathfrak{a}_{K-1}}{\mathfrak{a}_{K-1}\mathfrak{a}_K}}{\mathfrak{b}_K/\mathfrak{a}_K} && \text{[due to (2.6)]}\\
&= \frac{\mathfrak{b}_{K-1}}{\mathfrak{b}_K}\frac{\mathfrak{a}_K-\mathfrak{a}_{K-1}}{\mathfrak{a}_{K-1}}\\
&= \frac{1-p^{K-1}}{1-p^K}\frac{(K-1)p^{K-2}(1-p)^2}{1-(K-1)p^{K-2}+(K-2)p^{K-1}}\\
&= \frac{(K-1)p^{K-2}(1+p+\cdots+p^{K-2})}{(1+p+\cdots+p^{K-1})(1+2p+\cdots+(K-2)p^{K-3})}. && \text{[due to (A.31)]}
\end{aligned}$$

Again, observe that $\frac{\lambda^*_{K-1}}{\lambda^*_K}$ is the ratio of two polynomials of p. Namely, $\frac{\lambda^*_{K-1}}{\lambda^*_K} = \frac{\tilde{P}(p)}{\tilde{Q}(p)}$, where

$$\begin{aligned}
\tilde{P}(p) :=& (K-1)p^{K-2}\left(1+p+\cdots+p^{K-2}\right)\\
\tilde{Q}(p) :=& \left(1+p+\cdots+p^{K-1}\right)\left(1+2p+\cdots+(K-2)p^{K-3}\right)
\end{aligned}$$

Denote $\{\tilde{\xi}_k\}_{k=0}^{2K-4}$ and $\{\tilde{\nu}_k\}_{k=0}^{2K-4}$ as the coefficients of $\tilde{P}(p)$ and $\tilde{Q}(p)$ to be respectively, so that $\tilde{P}(p) = \sum_{k=0}^{2K-4}\tilde{\xi}_k p^k$ and $\tilde{Q}(p) = \sum_{k=0}^{2K-4}\tilde{\nu}_k p^k$. To show $\lambda^*_{K-1} < \lambda^*_K$, it suffices to show that $\tilde{P}(p) < \tilde{Q}(p)$. By observation, one can see

$$\tilde{\xi}_k = \begin{cases} 0, & 0 \le k \le K-3\\ K-1, & K-2 \le k \le 2K-4 \end{cases} \quad \text{and} \quad \tilde{\nu}_k = \begin{cases} \frac{(k+1)(k+2)}{2}, & 0 \le k \le K-3\\ \frac{(K-1)(K-2)}{2}, & k = K-2\\ \frac{k(2K-3-k)}{2}, & K-1 \le k \le 2K-4 \end{cases}$$

Hence

$$\tilde{\nu}_k - \tilde{\xi}_k = \begin{cases} 1, & k = 0 \\ \frac{(k+1)(k+2)}{2} > 0, & 1 \le k \le K-3 \\ \frac{(K-1)(K-4)}{2} \ge 0, & k = K-2 \\ \frac{k(2K-3-k)}{2} - (K-1) \overset{(a)}{\ge} K-4 \ge 0, & K-1 \le k \le 2K-5 \\ -1, & k = 2K-4 \end{cases}$$

In the derivations above, part (a) is by minimizing the term $\frac{k(2K-3-k)}{2} - (K-1)$ (as a quadratic function of k) subject to the constraint $K-1 \le k \le 2K-5$. This term obtains its minimal value at $K-4$ when $k = 2K-5$. Finally we evaluate $\tilde{Q}(p) - \tilde{P}(p)$ below:

$$\tilde{Q}(p) - \tilde{P}(p) = \sum_{k=0}^{2K-4} \tilde{\nu}_k p^k - \sum_{k=0}^{2K-4} \tilde{\nu}_k p^k = \sum_{k=0}^{2K-4} \big(\tilde{\nu}_k - \tilde{\xi}_k\big) p^K \ge 1 - p^{2K-4} > 0$$

Hence $\lambda^*_{K-1} < \lambda^*_K$ and the proof is finished. ■

A.6 Proof of Theorem 6

The first part of the proof (i.e., proving (2.10)) is almost a repeat verbatim of that of 2 (Step 1 plus the corresponding Lemma 6), by taking $\mathcal{M}_p^F = \mathcal{M}_p^{OA}$ (which is a finite set) and taking $\beta^F = C_0 + \log(1/\delta)$ (see also Remark 3). There are, however, two differences to be mindful of.

1. In Step 3 of Algorithm 2, a uniform randomization over all display sets is implemented when the time epoch t is a perfect square number. Under the Myopic Tracking Policy, we do not need such a component since $\lambda_*^{\mathrm{OA}}([K]) > 0$ by Theorem 4.

2. In the context of Theorem 6, the sample complexity guarantee (2.10) holds for an arbitrary $f \in \mathcal{M}_p$ rather than $f \in \mathcal{M}_p^{OA}$.

Because of the aforementioned differences, we need to present a slightly modified proof compared to that in Lemma 6. Without loss, suppose $f \in \mathcal{M}_p(\sigma_*)$. Pick $f^{\text{OA}} = f^{\text{OA}}_{\sigma_*}$ (see (2.3)) and an arbitrary $\bar{f}^{\text{OA}} \in \overline{\mathcal{M}}^{\text{OA}}$. Recall that $L^{f^{\text{OA}},\bar{f}^{\text{OA}}} : X, S \mapsto \log\left(\frac{f^{\text{OA}}(X|S)}{\bar{f}^{\text{OA}}(X|S)}\right)$ is the one-stage log-likelihood ratio function. Define

$$D^f_S\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right) := \mathbb{E}_{X\sim f}\left[L^{f^{\text{OA}},\bar{f}^{\text{OA}}}(X|S)\right].$$

This notation is consistent with that of Kullback-Leibler divergence because one can verify that $D^{f^{\text{OA}}}_S(f^{\text{OA}}||\bar{f}^{\text{OA}}) = \mathbb{E}_{X\sim f^{\text{OA}}}\left[L^{f^{\text{OA}},\bar{f}^{\text{OA}}}(X|S)\right] = D_S\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right)$.

Let us present a new version of Part 2 of the proof of Lemma 6, i.e., an estimate of the log likelihood ratio process $L^{f^{\text{OA}},\bar{f}^{\text{OA}}}_t$

$$\begin{aligned}
L^{f^{\text{OA}},\bar{f}^{\text{OA}}}_t =& \sum_{\ell=1}^{t} L^{f^{\text{OA}},\bar{f}^{\text{OA}}}(X_\ell, S_\ell) \\
=& \underbrace{\sum_{\ell=1}^{t}\left[L^{f^{\text{OA}},\bar{f}^{\text{OA}}}(X_\ell, S_\ell) - D^f_{S_\ell}\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right)\right]}_{A} \\
&+ \underbrace{\sum_{\ell=1}^{t}\left[D^f_{S_\ell}\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right) - D^f_{\lambda^{\text{OA}}_*}\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right)\right]}_{B} \\
&+ \underbrace{t\cdot D^f_{\lambda^{\text{OA}}_*}\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right)}_{C}.
\end{aligned}$$

In the expression above, Part A corresponds to randomness from the choices (given the display sets). It is a $\mathbb{P}_f-$martingale with bounded difference (and hence diverges sublinearly). Part B corresponds to the randomness from display sets (since the algorithm decides which display set to offer at each epoch based on historical data and possible randomization). Part C corresponds to the deterministic part, i.e., the long-run growth rate of the process $L^{f^{\text{OA}},\bar{f}^{\text{OA}}}_t$. Similarly in Part 2 of the proof of Lemma 6, we will show that both Part A and B diverge sublinearly while Part C grows at least as fast as $I^{\text{OA}}_* t$.

Our key observation is that there exists $\hat{f}^{\text{OA}} \in \overline{\mathcal{M}}^{\text{OA}}_p$ such that for every display set

$S \in \mathcal{S}$, $D_S^f\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right) \geq D_S\left(f^{\text{OA}}||\hat{f}^{\text{OA}}\right)$. In order to see that, let us suppose $S = [s]$ without loss of generality. Let $k := \sigma_{\bar{f}^{\text{OA}}}^{-1}(1)$ be the top ranked item under preference $\bar{f}^{\text{OA}}$, and $\hat{f}^{\text{OA}} = f^{\text{OA}}_{\hat{\sigma}_k}$, where $\hat{\sigma}_k$ is defined in (A.19). In addition, let us denote $f_i = f(X_i|S)$, $f_i^{\text{OA}} = f^{\text{OA}}(X_i|S)$, $\bar{f}_i^{\text{OA}} = \bar{f}^{\text{OA}}(X_i|S)$, and $\hat{f}_i^{\text{OA}} = \hat{f}^{\text{OA}}(X_i|S)$ for shorthand notations. Notice that

$$
\begin{aligned}
& D_S^f\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right) \\
= & \sum_{i=1}^{s} f_i \left[\log\left(f_i^{\text{OA}}\right) - \log\left(\bar{f}_i^{\text{OA}}\right)\right] \\
\overset{(a)}{\geq} & \sum_{i=1}^{s} f_i \left[\log\left(f_i^{\text{OA}}\right) - \log\left(\hat{f}_i^{\text{OA}}\right)\right] \\
\overset{(b)}{\geq} & \log(1/p)\left(\sum_{i=1}^{k-1} f_i\right) + (k-1)\log(p) f_k \\
\overset{(c)}{\geq} & \log(1/p)\,\Lambda_{k-1}/\Lambda_s + (k-1)\,\log(p)\,p^{k-1}/\Lambda_s \qquad\qquad [\Lambda_n = 1 + \cdots + p^{n-1} = \tfrac{1-p^s}{1-p}] \\
\overset{(d)}{=} & \sum_{i=1}^{k-1} \log(1/p) f_i^{\text{OA}} + (k-1)\log(p) f_k^{\text{OA}} \\
= & \sum_{i=1}^{s} f_i^{\text{OA}} \left[\log\left(f_i^{\text{OA}}\right) - \log\left(\hat{f}_i^{\text{OA}}\right)\right] = D_S\left(f^{\text{OA}}||\hat{f}^{\text{OA}}\right).
\end{aligned}
$$

In the derivations above, part (a) is a result of the rearrangement theorem plus the following three facts: (i) the sequence $\{f_i\}_{i=1}^s$ is decreasing in i; (ii) the sequence $\{\hat{f}_i^{\text{OA}}\}_{i=1}^s$ is an rearrangement of the sequence $\{\bar{f}_i^{\text{OA}}\}_{i=1}^s$; and (iii) the sequence $\{\hat{f}_i^{\text{OA}}\}_{i=1}^s$ is either decreasing in i if $k \notin S$ or is nonincreasing in i after excluding the k^{th} term at which the two sequences $\{\hat{f}_i^{\text{OA}}\}_{i=1}^s$ and $\{\bar{f}_i^{\text{OA}}\}_{i=1}^s$ coincide if $k \in S$. Part (b) is by invoking the explicit expressions for f^{OA} and $\hat{f}^{\text{OA}}$ respectively. Part (c) is a result of the fact that $f \in \mathcal{M}_p(\sigma_*)$ plus the following

chain of expressions:

$$\begin{aligned}\sum_{i=1}^{k-1} f_i - (k-1)f_k &= \underbrace{\left(\sum_{i=1}^{k} f_i\right)}_{\geq \Lambda_k/\Lambda_s} \underbrace{\left(\frac{\sum_{i=1}^{k-1} f_i - (k-1)f_k}{\sum_{i=1}^{k} f_i}\right)}_{>0} \\ &\geq \frac{\Lambda_k}{\Lambda_s}\left(1 - k\,\underbrace{f_k/(f_1+\cdots+f_k)}_{\leq p^{k-1}/\Lambda_k}\right) \\ &\geq \frac{\Lambda_k}{\Lambda_s}\left(1 - k\,p^{k-1}/\Lambda_k\right) \\ &= \Lambda_{k-1}/\Lambda_s - (k-1)p^{k-1}/\Lambda_s.\end{aligned}$$

Part (d) is due to invoking the explicit expressions for f^{OA} again.

Our key observation has several implications. For example, the full display set strictly separates f^{OA} from any $\tilde{f}^{\text{OA}} \in \mathcal{M}_p^{\text{OA}} \setminus \{f^{\text{OA}}\}$, i.e., $D^f_{[K]}\left(f^{\text{OA}}||\tilde{f}^{\text{OA}}\right) \geq D_{[K]}\left(f^{\text{OA}}||\tilde{f}^{\text{OA}}\right) > 0$. As a result, f_t^{OA} converges quickly to f^{OA} in probabilityt, i.e., $\mathbb{P}_f(f_t^{\text{OA}} \neq f^{\text{OA}}) \leq Ce^{-\epsilon t}$ for some $C, \epsilon > 0$ independent of δ.[2] A further consequence is that Part B grows sublinearly. To see why, we can decomposing Part B into two parts:

$$B = \underbrace{\sum_{\ell=1}^{t}\left[D^f_{\tilde{S}_\ell}\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right) - D^f_{\lambda_*^{\text{OA}}}\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right)\right]}_{B_1} + \underbrace{\sum_{\ell=1}^{t\wedge\hat{\tau}}\left[D^f_{S_\ell}\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right) - D^f_{\lambda_*^{\text{OA}}}\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right)\right]}_{B_2},$$

where $\hat{\tau} := \max\{t : f_t^{\text{OA}} \neq f^{\text{OA}}\}$ and $\{\tilde{S}_\ell\}_\ell$ is a sequence of i.i.d. $\mathcal{S}$-valued random variables with distribution λ_*^{OA} such that $\tilde{S}_\ell = S_\ell$ for all $t \geq \ell \geq \hat{\tau}+1$. Here B_1 is a partial sum of i.i.d. random variables with mean zero, and B_2 takes a diminishing fraction of time epochs when t is large. Hence both diverge sublinearly. Finally, notice that

$$C = t \cdot D^f_{\lambda_*^{\text{OA}}}\left(f^{\text{OA}}||\bar{f}^{\text{OA}}\right) \geq t \cdot D_{\lambda_*^{\text{OA}}}\left(f^{\text{OA}}||\hat{f}^{\text{OA}}\right) \geq \min_{\bar{f}\in\overline{\mathcal{M}}_p^{\text{OA}}} D_{\lambda_*^{\text{OA}}}\left(f^{\text{OA}}||\bar{f}\right) = t \cdot I_*^{\text{OA}}.$$

2. Note that here the tail probability is improved from $Ce^{-\epsilon\sqrt{t}}$ in Lemma 8. The reason is that MTP displays a strictly separating set (i.e., $[K]$) with strictly positive probability at each time epoch. It is unclear whether this is true in the general setting. Instead, the proof relies on "forced exploration" when the time epoch is a perfect square number;see Step 3 of Algorithm 2.

The rest of the proof of (2.10) follows verbatim from the arguments in Step 1 of that of Theorem 2.

Finally, the second part of the proof, i.e., showing that (2.10) can be taken to be equality when $f \in \mathcal{M}_p^{OA}$ is a straightforward corollary of the lower bound result 1. ■

A.7 Proof of Theorem 7

Before we prove Theorem 7, let us first formally state our estimate of $\mathbb{P}_f(d_\tau \neq \sigma_f^{-1}(1))$ for an arbitrary preference $f \in \mathcal{M}_p$ in Lemma 15. Some notation is useful in the development of Lemma 15 below. Recall from (MLE) that f_t^{OA} is the most likely consensus preference at time t. For simplicity of notation, let us denote $\sigma_t := \sigma_{f_t^{\text{OA}}}$ to be ranking associated with f_t^{OA}. For every given ranking σ, let us denote

$$\mathcal{E}_\sigma := \{\sigma_\tau = \sigma\} = \cup_{t=0}^{\infty}\{\sigma_t = \sigma\} \cap \{\tau = t\} \tag{A.32}$$

the event that the aggregated ranking equals σ when the algorithm terminates.

Lemma 15. *There exists a constant $\tilde{C}_1$ (only dependent on K) such that for every preference instance $f \in \mathcal{M}_p$, $\mathbb{P}_f(d_\tau \neq \sigma_f^{-1}(1)) \leq \tilde{C}_1 e^{-\beta}$.*

Proof. Fix an arbitrary $f \in \mathcal{M}_p$ and $\bar{\sigma} \in \overline{\Sigma}(\sigma_f)$. We will show that the probability of the algorithm stops with best estimated ranking to be (mistakenly) $\bar{\sigma}$ is small, or $\mathbb{E}_f[\mathbb{I}\{\mathcal{E}_{\bar{\sigma}}\}] \leq e^{-\beta}$. That leads to our desired result because $\mathbb{P}_f(d_\tau \neq \sigma_f^{-1}(1)) \leq \sum_{\bar{\sigma}\in\overline{\Sigma}(\sigma_f)} \mathbb{P}_f(\mathcal{E}_{\bar{\sigma}}) \leq |\overline{\Sigma}(\sigma_f)|e^{-\beta}$. That finished the proof by letting $\tilde{C}_1 := |\overline{\Sigma}(\sigma_f)| = (K-1)(K-1)!$, where $n!$ represents the factorial of a positive integer n.

For ease of notation, let us properly relabel the versions (within this proof only) so that $\bar{\sigma} = \sigma_*$. Under this relabeling rule, let $k := \sigma_f^{-1}(1)$, the top-ranked item under σ. We recall from (A.19) that $\hat{\sigma}_k$ is a particular ranking such that $\hat{\sigma}_k^{-1}(1) = \sigma_f^{-1}(1) = k$. We also pick $\bar{f}^{\text{OA}} = f_{\sigma_*}^{\text{OA}}, \hat{f}^{\text{OA}} = f_{\hat{\sigma}_k}^{\text{OA}}$, the expressions of which can be found in (2.3). Finally, for simplicity of notation, recall $\Lambda_n = 1 + p + \cdots + p^{n-1} = (1-p^n)/(1-p)$ for every integer n. Notice

that

$$
\begin{aligned}
&\mathbb{E}_f[\mathbb{I}\{\mathcal{E}_{\bar{\sigma}}\}] \\
=&\mathbb{E}_{\bar{f}^{\mathrm{OA}}}\left[\mathbb{I}\{\mathcal{E}_{\bar{\sigma}}\}\exp\left(-L_\tau^{\bar{f}^{\mathrm{OA}},f}\right)\right] && \text{[change-of-measure]} \\
=&\mathbb{E}_{\bar{f}^{\mathrm{OA}}}\left[\mathbb{I}\{\mathcal{E}_{\bar{\sigma}}\}\exp\left(-L_\tau^{\bar{f}^{\mathrm{OA}},\hat{f}^{\mathrm{OA}}}\right)\exp\left(-L_\tau^{\hat{f}^{\mathrm{OA}},f}\right)\right] \\
\leq&\mathbb{E}_{\bar{f}^{\mathrm{OA}}}\left[e^{-\beta}\exp\left(-L_\tau^{\hat{f}^{\mathrm{OA}},f}\right)\right] && \left[\mathbb{I}\{\mathcal{E}_{\bar{\sigma}}\}\exp\left(-L_\tau^{\bar{f}^{\mathrm{OA}},\hat{f}^{\mathrm{OA}}}\right)\leq e^{-\beta}\text{ a.s.}\right] \\
=&e^{-\beta}\mathbb{E}_{\bar{f}^{\mathrm{OA}}}\left[\exp\left(-L_\tau^{\hat{f}^{\mathrm{OA}},f}\right)\right] \\
=&e^{-\beta}\mathbb{E}_{\bar{f}^{\mathrm{OA}}}\left[\frac{f(X_1|S_1)}{\hat{f}^{\mathrm{OA}}(X_1|S_1)}\frac{f(X_2|S_2)}{\hat{f}^{\mathrm{OA}}(X_2|S_2)}\cdots\frac{f(X_\tau|S_\tau)}{\hat{f}^{\mathrm{OA}}(X_\tau|S_\tau)}\right].
\end{aligned}
$$

Since X_t is independent of the history conditional on S_t, it suffices to show that for every display set $S\in\mathcal{S}$, $\mathbb{E}_{\bar{f}^{\mathrm{OA}}}\left[\frac{f(X|S)}{\hat{f}^{\mathrm{OA}}(X|S)}\right]\leq 1$, where the expectation is taken over X only. To verify this inequality, first notice that if $k\notin S$, $f(X|S)\equiv\hat{f}^{\mathrm{OA}}(X|S)$ and $\frac{f(X|S)}{\hat{f}^{\mathrm{OA}}(X|S)}=1$ almost surely. Otherwise, it is without loss of generality to assume that $S=[s]$. Let us introduce the notation $f_i:=f(X_i|[K])$ for all $i\in[K]$. In that case,

$$
\begin{aligned}
\mathbb{E}_{\bar{f}^{OA}}\left[\frac{f(X|S)}{\hat{f}^{\mathrm{OA}}(X|S)}\right]-1&=\sum_{X\in[K]}\bar{f}^{OA}(X|[K])\frac{f(X|[K])}{\hat{f}^{\mathrm{OA}}(X|[K])}-1 \\
&=\sum_{i=1}^{k-1}\frac{p^{i-1}}{\Lambda_K}\frac{f_i}{\frac{p^i}{\Lambda_K}}+\frac{p^{k-1}}{\Lambda_K}\frac{f_k}{\frac{1}{\Lambda_K}}+\sum_{i=k+1}^{K}\frac{p^{i-1}}{\Lambda_K}\frac{f_i}{\frac{p^{i-1}}{\Lambda_K}}-1 \\
&=\sum_{i=1}^{k-1}\frac{1}{p}f_i+p^{k-1}f_k+\sum_{i=k+1}^{K}f_i-1 \\
&=\left(\tfrac{1}{p}-1\right)\sum_{i=1}^{k-1}f_i+(p^{k-1}-1)f_k \\
&\leq\left(\tfrac{1}{p}-1\right)\tfrac{p+p^2+\cdots+p^{k-1}}{\Lambda_K}+(p^{k-1}-1)\tfrac{1}{\Lambda_K} \\
&=\tfrac{1+\cdots+p^{k-1}-1-\cdots-p^{k-1}}{\Lambda_K}=0.
\end{aligned}
$$

That concludes the proof. ■

Proof of Theorem 7. The fact that $\mathbb{P}_f(\tau<\infty)=1$ for every $f\in\mathcal{M}_p$ is ensured by the

proof of Theorem 6 Due to Lemma 15, we only need to have $\beta \geq \log(\tilde{C}_1) + \log\frac{1}{\delta}$ to ensure $\mathbb{P}_f(d_\tau \neq \sigma_f^{-1}(1)) \leq \delta$. Hence the proof is finished by letting $C_1 = \log(\tilde{C}_1)$. ■

A.8 Proof of Proposition 2, Proposition 3, and Proposition 4

Proof of Proposition 2. We first show part 1. Due to Theorem 3, we have

$$f_\sigma^{\text{OA}}(X|S) = \frac{1-p}{\left(1-p^{|S|}\right)} p^{\sigma(k|S)-1}, \quad \forall \sigma \in \Sigma, S \in \mathcal{S}, k \in S.$$

In the above expression, the relative ranking is defined as $\sigma(k|S) := \sum_{i\in S} \mathbb{I}\{\sigma(i) \leq \sigma(k)\}$. For every $f \in \mathcal{M}_p^{OA}$ there exists a σ so that $f = f_\sigma^{\text{OA}}$. For all voting history $H_t = (S_1, X_1, \ldots, S_t, X_t)$,

$$\begin{aligned}
&\sum_{\ell=1}^{t} \log f_\sigma^{\text{OA}}(X_\ell|S_\ell) \\
=&\sum_{\ell=1}^{t} \left[\log\left(\frac{1-p}{p\left(1-p^{|S_\ell|}\right)}\right) + \log p \cdot \left(\sum_{i\in S_\ell} \sigma(i) \leq \sigma(X_\ell)\right)\right] \\
=&\sum_{\ell=1}^{t} \log\left(\tfrac{1-p}{1-p^{|S_\ell|}}\right) + \log p \cdot \left(\sum_{\ell=1}^{t} \sum_{(i,j):i\neq j} \mathbb{I}\{\sigma(j) < \sigma(i)\} \cdot \mathbb{I}\{X_\ell = i\} \cdot \mathbb{I}\{i,j \in S_\ell\}\right) \\
=&\sum_{\ell=1}^{t} \log\left(\tfrac{1-p}{1-|p|^{S_\ell}}\right) + \log p \cdot \sum_{(i,j):i\neq j} \left[\mathbb{I}\{\sigma(j) < \sigma(i)\} \cdot \left(\sum_{\ell=1}^{t} \mathbb{I}\{X_\ell = i\} \cdot \mathbb{I}\{i,j \in S_\ell\}\right)\right] \\
=&\sum_{\ell=1}^{t} \log\left(\tfrac{1-p}{1-|p|^{S_\ell}}\right) + \log p \cdot \sum_{(i,j):i\neq j} \mathbb{I}\{\sigma(j) < \sigma(i)\} w_{ij}^t = \phi + \log p \cdot c(f_\sigma^{\text{OA}}, \vec{w}^t).
\end{aligned}$$

Here we define $\phi := \sum_{\ell=1}^{t} \log\left(\frac{1-p}{1-|p|^{S_\ell}}\right)$, which is independent of σ.

Turning to part 2, let $f, \bar{f} \in \mathcal{M}_p^{OA}$ be such that there exists $\sigma, \bar{\sigma}$ so that $f = f_\sigma^{OA}$ and

$\bar{f} = f_{\bar{\sigma}}^{OA}$. As a result,

$$\begin{aligned} L_t^{f_\sigma^{\text{OA}}, f_{\bar{\sigma}}^{\text{OA}}} &= \sum_{\ell=1}^{t} \log \frac{f_\sigma^{\text{OA}}(X_\ell|S_\ell)}{\bar{f}_{\bar{\sigma}}^{OA}(X_\ell|S_\ell)} \\ &= \sum_{\ell=1}^{t} \log f_\sigma^{\text{OA}}(X_\ell|S_\ell) - \sum_{\ell=1}^{t} \log f_{\bar{\sigma}}^{\text{OA}}(X_\ell|S_\ell) \\ &= \log p \cdot \left[c(f^{\text{OA}}, \vec{w}^t) - c(f_{\bar{\sigma}}^{\text{OA}}, \vec{w}^t) \right]. \end{aligned}$$

where the first line follows from part 1 of the proof. This establishes part 2. ■

Proof of Proposition 3. Recall from the proof in Theorem 3 that $\hat{\sigma}_m = (2, \ldots, m, 1, m+1, \ldots, K)$ for $m = 2, \ldots, K$. Moreover, invoking the proof in Theorem 3, we may simplify the both max-min problems in Proposition 3 by restricting the alternative preferences to the family of $\{f_{\hat{\sigma}_m}^{\text{OA}} : m = 2, \ldots, K\}$ only. More precisely:

$$\begin{aligned} &\max_{\lambda \in \Delta(\mathcal{S}^P)} \min_{\bar{f} \in \overline{\mathcal{M}}_p(f_*^{\text{OA}})} D_\lambda \left(f_*^{\text{OA}} || \bar{f} \right) \\ &= \max_{\lambda \in \Delta(\mathcal{S}^P)} \min_{m=2,\ldots,K} D_\lambda \left(f_*^{\text{OA}} || f_{\hat{\sigma}_m}^{\text{OA}} \right) \\ &= \max_{\lambda \in \Delta(\mathcal{S}^P)} \min_{m=2,\ldots,K} \sum_{S \in \mathcal{S}^P} \lambda(S) d_S \left(\hat{\sigma}_m \right); \\ &\max_{\lambda \in \Delta(\mathcal{S}^{PF})} \min_{\bar{f} \in \overline{\mathcal{M}}_p(f_*^{\text{OA}})} D_\lambda \left(f_*^{\text{OA}} || \bar{f} \right) \\ &= \max_{\lambda \in \Delta(\mathcal{S}^{PF})} \min_{m=2,\ldots,K} D_\lambda \left(f_*^{\text{OA}} || f_{\hat{\sigma}_m}^{\text{OA}} \right) \\ &= \max_{\lambda \in \Delta(\mathcal{S}^{PF})} \min_{m=2,\ldots,K} \sum_{S \in \mathcal{S}^{PF}} \lambda(S) d_S \left(\hat{\sigma}_m \right). \end{aligned}$$

We will use the shorthand notation $d_S(\sigma) = D_S \left(f_{\sigma_*}^{\text{OA}} || f_\sigma^{\text{OA}} \right)$ (also used in the proof in Theorem 4) in what follows. Given any randomization $\mu \in \Delta(\{\hat{\sigma}_m : m = 2, \ldots, K\})$, we follow the convention of defining $d_S(\mu) := \sum_{m=2}^{K} \mu(\hat{\sigma}_m) d_S(\hat{\sigma}_m)$. For the remaining of the proof, we show the optimality of $\lambda_*^{\text{P,OA}}$ and $\lambda_*^{\text{PF,OA}}$ separately.

We claim that

$$\lambda_*^{\mathrm{P,OA}} \in \arg\max_{\lambda\in\Delta(\mathcal{S}^P)} \min_{m=2,\ldots,K} \sum_{S\in\mathcal{S}^P} \lambda(S) d_S\left(\hat{\sigma}_m\right).$$

Invoking Lemma 11, for all $i<j\in[K]$,

$$d_{\{i,j\}}\left(\hat{\sigma}_m\right) = \begin{cases} \frac{\mathfrak{a}_2}{\mathfrak{b}_2} & \text{if } j=m \\ 0 & \text{otherwise.} \end{cases} = \begin{cases} \log\left(\frac{1}{p}\right)\frac{1-p}{1+p} & \text{if } j=m \\ 0 & \text{otherwise.} \end{cases}$$

As a result, for every m, $d_{\lambda_*^{\mathrm{P,OA}}}\left(\hat{\sigma}_m\right) = \frac{1}{K-1}\frac{\mathfrak{a}_2}{\mathfrak{b}_2}$. Consider the randomization $\mu_*^{\mathrm{P,OA}} \in \Delta(\{\hat{\sigma}_m : m=2,\ldots,K\})$ given by $\mu_*^{\mathrm{P,OA}}(\hat{\sigma}_m) = \frac{1}{K-1}$ for all m. Then $d_S\left(\mu_*^{\mathrm{P,OA}}\right) = \frac{1}{K-1}\frac{\mathfrak{a}_2}{\mathfrak{b}_2}$ for all $S\in\mathcal{S}^P$. Noticing that $d_{\lambda_*^{\mathrm{P,OA}}}\left(\hat{\sigma}_m\right) = d_S\left(\mu_*^{\mathrm{P,OA}}\right)$, we follow the same argument in Theorem 4 and conclude that $\lambda_*^{\mathrm{P,OA}}$ is primal optimal and $\mu_*^{\mathrm{P,OA}}$ is dual optimal for the linear program associated with the max-min problem $\max_{\lambda\in\Delta(\mathcal{S}^P)} \min_{m=2,\ldots,K} \sum_{S\in\mathcal{S}^P} \lambda(S) d_S\left(\hat{\sigma}_m\right)$. (However, unlike Theorem 3, $\lambda_*^{\mathrm{P,OA}}$ is not the unique primal optimal solution.) Correspondingly, the optimal value is

$$\frac{1}{K-1}\frac{\mathfrak{a}_2}{\mathfrak{b}_2} = (1-p)\log\left(\frac{1}{p}\right)\frac{1}{(K-1)(1+p)}.$$

We claim that $\lambda_*^{\mathrm{PF,OA}} \in \arg\max_{\lambda\in\Delta(\mathcal{S}^{PF})} \min_{m=2,\ldots,K} \sum_{S\in\mathcal{S}^P} \lambda(S) d_S\left(\hat{\sigma}_m\right)$. Note that $d_{[K]}\left(\hat{\sigma}_m\right) = \frac{\mathfrak{a}_m}{\mathfrak{b}_K}$ for all $m=2,\ldots,K$. Letting $x = \frac{\mathfrak{a}_2/\mathfrak{b}_2}{\mathfrak{a}_2/\mathfrak{b}_2+(\mathfrak{a}_3-\mathfrak{a}_2)/\mathfrak{b}_K}$, we may verify that

$$d_{\lambda_*^{\mathrm{PF,OA}}}\left(\hat{\sigma}_m\right) = \begin{cases} \frac{\mathfrak{a}_2}{\mathfrak{b}_2}(1-x) + \frac{\mathfrak{a}_2}{\mathfrak{b}_K}x & \text{if } m=2 \\ \frac{\mathfrak{a}_m}{\mathfrak{b}_K}x & \text{if } m=3,\ldots,K. \end{cases}$$

Regarding the expression above, we may verify that $\frac{\mathfrak{a}_2}{\mathfrak{b}_2}(1-x) + \frac{\mathfrak{a}_2}{\mathfrak{b}_K}x - \frac{\mathfrak{a}_3}{\mathfrak{b}_K}x = \frac{\mathfrak{a}_2}{\mathfrak{b}_2} - \left(\frac{\mathfrak{a}_2}{\mathfrak{b}_2} - \frac{\mathfrak{a}_3-\mathfrak{a}_2}{\mathfrak{b}_K}\right)x = 0$. Plus, since $\mathfrak{a}_n$ strictly increases in n, we may verify that

$$d_{\lambda_*^{\mathrm{PF,OA}}}\left(\hat{\sigma}_K\right) > \cdots > d_{\lambda_*^{\mathrm{PF,OA}}}\left(\hat{\sigma}_4\right) > d_{\lambda_*^{\mathrm{PF,OA}}}\left(\hat{\sigma}_3\right) = d_{\lambda_*^{\mathrm{PF,OA}}}\left(\hat{\sigma}_2\right) = \frac{\mathfrak{a}_3}{\mathfrak{b}_K}x = \frac{\mathfrak{a}_2\mathfrak{a}_3/\mathfrak{b}_2\mathfrak{b}_K}{\mathfrak{a}_2/\mathfrak{b}_2+(\mathfrak{a}_3-\mathfrak{a}_2)/\mathfrak{b}_K}.$$

In the meanwhile, let us consider the randomization $\mu_*^{\mathrm{PF,OA}} \in \Delta(\{\hat{\sigma}_m : m=2,\ldots,K\})$ given

by

$$\mu_*^{\mathrm{P,OA}}(\hat{\sigma}_m) = \begin{cases} \frac{\mathfrak{a}_3/\mathfrak{b}_K}{\mathfrak{a}_3/\mathfrak{b}_K+\mathfrak{a}_2(1/\mathfrak{b}_2-1/\mathfrak{b}_K)} & \text{if } m=2 \\ \frac{\mathfrak{a}_2(1/\mathfrak{b}_2-1/\mathfrak{b}_K)}{\mathfrak{a}_3/\mathfrak{b}_K+\mathfrak{a}_2(1/\mathfrak{b}_2-1/\mathfrak{b}_K)} & \text{if } m=3 \\ 0 & \text{otherwise.} \end{cases}$$

By letting $y = \frac{\mathfrak{a}_3/\mathfrak{b}_K}{\mathfrak{a}_3/\mathfrak{b}_K+\mathfrak{a}_2(1/\mathfrak{b}_2-1/\mathfrak{b}_K)}$ for shorthand notation, we may evaluate $d_S\left(\mu_*^{\mathrm{PF,OA}}\right)$ for every $S \in \mathcal{S}^{PF}$below:

$$d_S\left(\mu_*^{\mathrm{PF,OA}}\right) = \begin{cases} \frac{\mathfrak{a}_2}{\mathfrak{b}_2}y & \text{if } S=[2] \\ \frac{\mathfrak{a}_2}{\mathfrak{b}_K}y + \frac{\mathfrak{a}_3}{\mathfrak{b}_K}(1-y) & \text{if } S=[K]. \end{cases}$$

Regarding the expression above, we may verify that $\frac{\mathfrak{a}_2}{\mathfrak{b}_K}y + \frac{\mathfrak{a}_3}{\mathfrak{b}_K}(1-y) - \frac{\mathfrak{a}_2}{\mathfrak{b}_2}y = \frac{\mathfrak{a}_3}{\mathfrak{b}_K} - \left(\frac{\mathfrak{a}_3}{\mathfrak{b}_K} + \frac{\mathfrak{a}_2}{\mathfrak{b}_2} - \frac{\mathfrak{a}_2}{\mathfrak{b}_K}\right)y = 0$. That implies $d_{[2]}\left(\mu_*^{\mathrm{PF,OA}}\right) = d_{[K]}\left(\mu_*^{\mathrm{PF,OA}}\right)$. We follow the same argument in Theorem 4 and conclude that $\lambda_*^{\mathrm{PF,OA}}$ is (uniquely) primal optimal and $\mu_*^{\mathrm{PF,OA}}$ is dual optimal for the linear program associated with the max-min problem

$$\max_{\lambda \in \Delta(\mathcal{S}^{PF})} \min_{m=2,\ldots,K} \sum_{S \in \mathcal{S}^{PF}} \lambda(S) d_S\left(\hat{\sigma}_m\right).$$

Correspondingly, the optimal value is

$$\frac{\mathfrak{a}_2\mathfrak{a}_3/\mathfrak{b}_2\mathfrak{b}_K}{\mathfrak{a}_2/\mathfrak{b}_2 + (\mathfrak{a}_3-\mathfrak{a}_2)/\mathfrak{b}_K} = (1-p)\log\left(\tfrac{1}{p}\right)\frac{1+2p}{(1-p^K)/(1-p)+2p(1+p)}.$$

This finishes the proof. ■

Proof of Proposition 4. Our main observation in this proof is that we may view the policy F (resp., P and PF) as admissible solutions to our learning problem with an extra constraint about the structure of display sets. With this observation in mind, we will finish this proof by carrying over the arguments in the proofs of Theorems 1, 3, 5, 6, and 7 to the settings of pairwise, full, and pair & full display strategies.

First, let us establish the $\delta-$accuracy of F, P and PF. Note that the proof of Theorem 7 is purely based on a change-of-measure argument restricted to the OA model and then a dominance argument on the "hardness to learn" of the OA model (see also the comments right after Theorem 7). This argument does not reply on the specific structure of the display policy. Hence we may repeat the proof of Theorem 7 verbatim after replacing $\mathcal{S}$ with $\mathcal{S}^F, \mathcal{S}^P$ and $\mathcal{S}^{PF}$ respectively to show that the policies F, P, FP are all $\delta-$accurate.

Second, let us establish the sample complexity of the policies $\{\mathrm{P}, \mathrm{F}, \mathrm{PF}\}$ and conclude that (2.20) holds. Pick an arbitrary $f' \in \mathcal{M}_p^{\mathrm{OA}}$. In order to show that the estimated preference f_t^{OA} converges to f_*^{OA} fast, let us look at the following three quantities:

- $D_{[K]}\left(f_*^{\mathrm{OA}}||f'\right)$;
- $D_{\lambda_*^{\mathrm{PF,OA}}}\left(f_*^{\mathrm{OA}}||f'\right) \geq \frac{\mathfrak{a}_2/\mathfrak{b}_2}{\mathfrak{a}_2/\mathfrak{b}_2+(\mathfrak{a}_3-\mathfrak{a}_2)/\mathfrak{b}_K} \cdot D_{[K]}\left(f_*^{\mathrm{OA}}||f'\right)$;
- $D_{\lambda_*^{\mathrm{P,OA}}}\left(f_*^{\mathrm{OA}}||f'\right) = \frac{\sum_{i=1}^{K-1} \mathbb{I}\{\sigma_{f'}(i)>\sigma_{f'}(i+1)\}}{K-1} \frac{\mathfrak{a}_2}{\mathfrak{b}_2}$.

All of the three quantities above are strictly positive as long as $\sigma_{f'} \neq \sigma_*$. In any case, we know that $f_t^{\mathrm{OA}} \to f_*^{\mathrm{OA}}$ exponentially fast in probability (in the same spirit of the proof of Theorem 6). The rest of the proof is based on repeating the arguments in Theorem 6 by replacing $\mathcal{S}$ with $\mathcal{S}^F$, $\mathcal{S}^P$ and $\mathcal{S}^{PF}$ respectively.

Third, let us establish the lower bound of sample complexity and conclude that (2.21) holds. We refer to the lower bound result under the general hypothesis testing framework (see Section A.1.2) and conclude that if we restrict the collection of display sets to $\mathcal{S}^F$ and $\mathcal{S}^P$ respectively, we have that for every $f \in \mathcal{M}_p$,

$$\begin{aligned}
\liminf_{\delta\downarrow 0} \frac{\mathbb{E}_f^\pi[\tau]}{\log\left(\frac{1}{\delta}\right)} &\geq \frac{1}{\max_{\lambda\in\Delta(\mathcal{S}^F)} \min_{\bar{f}\in\overline{\mathcal{M}}_p(f)} D_\lambda\left(f||\bar{f}\right)}, \quad \forall \pi \in \mathcal{A}^F; \\
\liminf_{\delta\downarrow 0} \frac{\mathbb{E}_f^\pi[\tau]}{\log\left(\frac{1}{\delta}\right)} &\geq \frac{1}{\max_{\lambda\in\Delta(\mathcal{S}^P)} \min_{\bar{f}\in\overline{\mathcal{M}}_p(f)} D_\lambda\left(f||\bar{f}\right)}, \quad \forall \pi \in \mathcal{A}^P.
\end{aligned} \tag{A.33}$$

Specifically, if $f \in \mathcal{M}_p^{OA}$, and we take π to be F and P respectively, we have $\liminf_{\delta \downarrow 0} \frac{\mathbb{E}_f^F[\tau]}{\log\left(\frac{1}{\delta}\right)} \geq \frac{1}{I^F}$ and $\liminf_{\delta \downarrow 0} \frac{\mathbb{E}_f^P[\tau]}{\log\left(\frac{1}{\delta}\right)} \geq \frac{1}{I^P}$. Align these two inequalities with (2.20), and we conclude that (2.21) holds.[3]

Fourth, let us conclude that the policies {P, F} are worst-case asymptotically optimal for full and pairwise display policies respectively. We notice that the proof of Theorem 3 is based on a dominance argument on the "hardness to learn" of the OA model. Hence we may follow the same argument and conclude that

$$I^F \leq \max_{\lambda \in \Delta(\mathcal{S}^F)} \min_{\bar{f} \in \overline{\mathcal{M}}_p(f)} D_\lambda\left(f||\bar{f}\right), \ \forall f \in \mathcal{M}_p;$$
$$I^P \leq \max_{\lambda \in \Delta(\mathcal{S}^P)} \min_{\bar{f} \in \overline{\mathcal{M}}_p(f)} D_\lambda\left(f||\bar{f}\right), \ \forall f \in \mathcal{M}_p.$$

The inequalities above can be taken as equalities when $f \in \mathcal{M}_p^{\text{OA}}$. Hence

$$\sup_{f \in \mathcal{M}_p} \liminf_{\delta \downarrow 0} \frac{\mathbb{E}_f^\pi[\tau]}{\log\left(\frac{1}{\delta}\right)} \geq \sup_{f \in \mathcal{M}_p} \frac{1}{\max_{\lambda \in \Delta(\mathcal{S}^F)} \min_{\bar{f} \in \overline{\mathcal{M}}_p(f)} D_\lambda\left(f||\bar{f}\right)} = \frac{1}{I^F}, \quad \forall \pi \in \mathcal{A}^F;$$
$$\sup_{f \in \mathcal{M}_p} \liminf_{\delta \downarrow 0} \frac{\mathbb{E}_f^\pi[\tau]}{\log\left(\frac{1}{\delta}\right)} \geq \sup_{f \in \mathcal{M}_p} \frac{1}{\max_{\lambda \in \Delta(\mathcal{S}^P)} \min_{\bar{f} \in \overline{\mathcal{M}}_p(f)} D_\lambda\left(f||\bar{f}\right)} = \frac{1}{I^P}, \quad \forall \pi \in \mathcal{A}^P.$$

In the mean time, (2.21) implies that

$$\sup_{f \in \mathcal{M}_p} \limsup_{\delta \downarrow 0} \frac{\mathbb{E}_f^F[\tau]}{\log\left(\frac{1}{\delta}\right)} \leq \sup_{f \in \mathcal{M}_p} \frac{1}{I^F} = \frac{1}{I^F}$$
$$\sup_{f \in \mathcal{M}_p} \limsup_{\delta \downarrow 0} \frac{\mathbb{E}_f^P[\tau]}{\log\left(\frac{1}{\delta}\right)} \leq \sup_{f \in \mathcal{M}_p} \frac{1}{I^P} = \frac{1}{I^P}.$$

Hence the policies {P, F} are worst-case asymptotically optimal for full and pairwise display policies respectively.

3. Note that a similar result does not apply to the policy PF because the collection of sets $\mathcal{S} = \{[2], [K]\}$ is not closed under permutations. As a consequence, the display sets used by Policy PF may include elements other than [2] and $[K]$, which breaks down the arguments for the lower bound result.

Finally, since $\mathcal{S}^F$ is a singleton,

$$\begin{aligned} & I^F \\ =& \max_{\lambda \in \Delta(\mathcal{S}^F)} \min_{\bar{f} \in \overline{\mathcal{M}}_p(f_*^{\mathrm{OA}})} D_\lambda \left(f_*^{\mathrm{OA}} || \bar{f} \right) = \min_{\bar{f} \in \overline{\mathcal{M}}_p(f_*^{\mathrm{OA}})} D_{[K]} \left(f_*^{\mathrm{OA}} || \bar{f} \right) \\ =& \frac{\mathfrak{a}_2}{\mathfrak{b}_K} \\ =& (1-p) \log\left(\frac{1}{p}\right) \frac{1}{(1-p^K)/(1-p)}. \end{aligned}$$

The expressions for I^P and I^{PF} can be obtained in the proof of Proposition 3. ■

A.9 Running Time of the Myopic Tracking Policy

In this section, we provide more implementation details when we evaluate the computational speed of Step 1 of the Myopic Tracking Policy in Section 2.8.1.

A.9.1 Two Methods

Integer programming formulation We solve the exact integer programming formulation (2.15) associated with MTP with the Gurobi MIP solver. Because the simplex method is sometimes significantly slower than the barrier method for the root relaxation of (2.15), we always use the non-deterministic concurrent algorithm option (provided by Gurobi) to solve the root relaxation.

Our heuristic The heuristic we use combines the idea of majority tournament (with random pivoting rule) in Ailon et al. (2008) and the LP relaxation of (2.15) as also used in Van Zuylen and Williamson (2009). The LP relaxation of (2.15) is

$$\begin{aligned} \tilde{x} \in \arg\min_{\vec{x}} & \sum_{(i,j):i \neq j} x_{ji} w_{ij}^t \\ s.t. \quad & x_{ij} + x_{jk} + x_{ki} \geq 1, \quad \forall \text{ distinct } i,j,k \in [K] \\ & x_{ij} + x_{ji} = 1, \qquad \forall \text{ distinct } i,j \in [K] \\ & x_{ij} \in [0,1]. \qquad \forall \text{ distinct } i,j \in [K] \end{aligned} \tag{A.34}$$

Given the value of $\tilde{x}$, the estimated ranking is $\tilde{\sigma}^{-1}$, where $\tilde{\sigma}$ is the permutation we obtain using Algorithm 3.

Algorithm 3: LP-Rand-Pivot

INPUT: Collection of Items $\mathcal{C} \subseteq [L]$

STEP 0: Randomly pick a pivot $k \in V$

STEP 1: $\mathfrak{L} = \{i \in [K] \setminus \{k\} : \tilde{x}_{ki} \leq 0.5\}$, $\mathfrak{R} = \{i \in [K] \setminus \{k\} : \tilde{x}_{ki} > 0.5\}$

OUTPUT: LP-Rand-Pivot($\mathfrak{L}$), k, LP-Rand-Pivot($\mathfrak{R}$)
(Concatenation of left recursion, k, and right recursion.)

To get some intuition about Algorithm 3, note that a large of value of $\tilde{x}_{ij}$ represents a stronger tendency that item i is ranked higher than (or preferred to) item j. Hence $\mathfrak{L}$ (resp. $\mathfrak{R}$) in Algorithm 3 correspond to items that are ranked higher (resp. lower) than the pivot item k.

Let us also say a few words on how our heuristic is connected to the earlier literature. The difference between our heuristic and that in Ailon et al. (2008) is that the underlying tournament in our heuristic is $A_H = \{(i,j) : \tilde{x}_{ij} > \tilde{x}_{ji}\}$, while the underlying tournament in Ailon et al. (2008) is $A_M = \{(i,j) : w^t_{ij} > w^t_{ji}\}$. The main difference between our heuristic and that in Van Zuylen and Williamson (2009) (among others) is that we do not solve an underlying optimization problem to find the pivot item k. The reason we implement this heuristic is that this heuristic combines the simple implementation of Ailon et al. (2008) (e.g., random pivoting, majority tournament) and the information of the solution to the LP relaxation. In particular, if the LP returns an integral solution, our heuristic is guaranteed to achieve zero optimality gap.

A.9.2 Stimulation Details

We generate representative instances of (2.15) that arise as we implement MTP. Our stimulation is based on two parts. In the first part, we iterate over the learning problem instances by taking $p \in \{0.5, 0.9\}$ and K ranging from 25 to 150. For each problem instance (p, K), we

use our heuristic (for Step 1) and the optimal display randomization (for Step 3) to generate to generate a 500 period run of MTP five different times (allowing for random realizations of display sets and consumer choices) and track the instances of (2.15) that arise along the way. The underlying choice model is the OA model. We record the computation times and optimality gaps and aggregate them at the level of (K, p). Here the optimality gap is defined as $(v_H - v_{LP})/v_H$, where v_H is the objective function value of our heuristic and v_{LP} is the optimal value of the LP relaxation of (2.15). Note that the optimality gap relative to the LP relaxation is always an upper bound of the "actual" gap compared to the exact solution.

In the second part, we evaluate the computational performance of exactly solving the integer programming formulation. We do so by revisiting the IP instances generated in the first part with two different approaches: (i) we solve those (same) instance exactly using the Gurobi MIP solver (instead of approximately using our heuristic); and (ii) we focus on a subset of periods where $t \in \{20, 40, 60, \ldots, 500\}$ so as to speed up the stimulation process. We record the computation times and aggregate them at the level of (K, p).

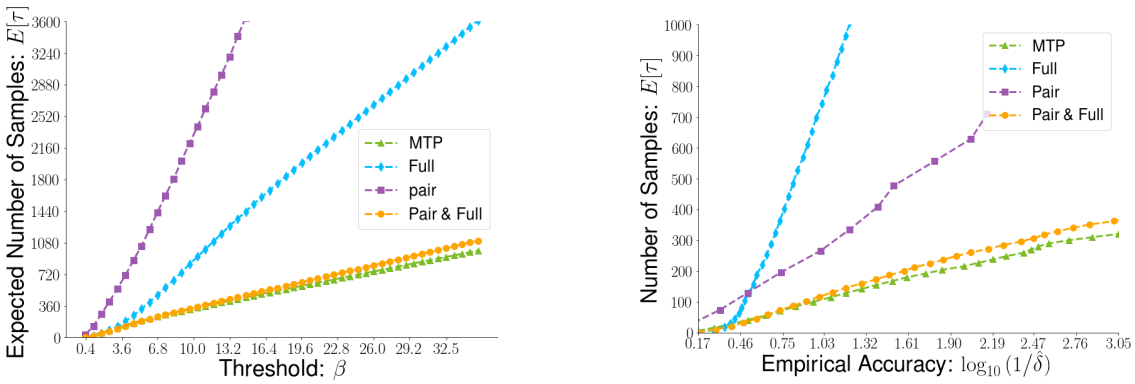
A.10 AGH Survey Data

In this appendix we test the performance of our proposed MTP policy using a real data set from a student survey conducted at AGH University of Science and Technology. Specifically, we have sampled from the (first) AGH Course survey dataset available from PREFLIB, an online library of datasets concerning preferences (see PREFLIB, 2019). This is a dataset with the complete rankings of 146 students regarding 8 course modules, collected at AGH University of Science and Technology, Krakow, in 2003.

To emulate our sequential learning setting, we randomly generate students from the data set and present them with subset of courses. We use a rather intuitive approach to convert the available complete ranking data to a choice model (see, e.g., Désir et al., 2018). Specifically, given any display set, we first (uniformly) randomly draw a complete ranking from the dataset (with replacement), and then pick the course that is ranked highest in the

display set. This conversion rule gives rise to a particular preference model f_{AGH}, where $f_{AGH}(X|S)$ represents the empirical fraction of students who rank course X at the top within display set S. We run the same four policies consider in Section 2.8 under the preference model f_{AGH} to see which uses the least amount of samples (i.e., choices) to recover the students' most preferred course with high probability. It is important to note that f_{AGH} does not satisfy Assumption (A-3) in Definition 1, i.e, the notion of "ground truth ranking" is not well defined. It does, however, satisfy the weaker condition (A-5) discussed in Section 2.9 and so lies in $\widetilde{\mathcal{M}}_p$. Indeed, we do find that there is a course i_* that is chosen with higher probability than any other courses under f_{AGH} regardless of the display set. In other words, the problem of identifying the "ground truth most preferred course" is well defined.

In Figure A.1, we look at the sample complexity of the four policies as a function of (a) the stopping threshold β and (b) the resulting empirical error probability (or the fraction of instances where the algorithm terminates with a course different from i_*). In the implementation of these experiments, we adopt the integer programming formulation of the MLE problem and set $p = 0.7$.



(a) stopping threshold β

(b) empirical error probability

Figure A.1: **Comparison of policy performance on the AGH Survey Data.**

These numerical results suggest that the Myopic Tracking Policy dominates the three benchmark policies, stopping earlier for every fixed threshold β and for every value of the empirical error probability. This suggests that the Myopic Tracking policy is effective, even

in settings more general than the one analyzed in the paper.

APPENDIX B

SUPPLEMENT TO CHAPTER 3

B.1 Facts Related to Problem Inputs (Assumption 1)

In this section, we explore immediate consequences of Assumption 1, centering around the symmetry and separability of certain problem inputs.

B.1.1 Symmetry

Much of our analysis throughout the paper is based on the ideas that we only need to analyze the system under one of the two hypotheses, and the case for the other hypothesis follows by symmetry. The following lemmas reveals the problem structure that allows us to enjoy such simplification.

Lemma 16. *(symmetry of probability distribution) For every* $x \in \mathbb{R}$, $f_0(x) = f_1(m_0+m_1-x)$ *and* $F_0(x) + F_1(m_0 + m_1 - x) = 1$.

Proof. Note that for every $i \in \{0, 1\}$, $f_i(\cdot)$ is the p.d.f. for $m_i + \epsilon$, where ϵ has a p.d.f. $f_\epsilon(\cdot)$. As a result, for every $x \in \mathbb{R}$ and $i \in \{0, 1\}$, $f_i(x) = f_\epsilon(x - m_i)$ and $F_i(x) = F_\epsilon(x - m_i)$. Due to (A1:2), which states that f_ϵ is symmetric around zero, we have $f_\epsilon(x) = f_\epsilon(-x)$. Hence, $f_0(x) = f_\epsilon(x-m_0) = f_\epsilon(m_0-x) = f_1(m_0+m_1-x)$, thereby establishing the first statement in the lemma.

To prove the second statement, we observe that for all $x \in \mathbb{R}$, $F_\epsilon(x) = \int_{-\infty}^{x} f_\epsilon(t)dt = 1-\int_{x}^{\infty} f_\epsilon(t)dt = 1-\int_{-\infty}^{-x} f_\epsilon(-t)dt = 1-\int_{-\infty}^{-x} f_\epsilon(t)dt = 1-F_\epsilon(-x)$. Thus, $F_\epsilon(x)+F_\epsilon(-x) = 1$, and $F_0(x) = F_\epsilon(x - m_0) = 1 - F_\epsilon(m_0 - x) = 1 - F_1(m_1 + m_0 - x)$. Therefore, we have the second statement in the lemma. ■

Lemma 17. *(symmetry of profit function) For every* $s \in [m_0, m_1]$,

$$j_1^+(s) = j_0^-(m_0 + m_1 - s)$$

and

$$j_1^-(s) = j_0^+(m_0 + m_1 - s).$$

Proof. We observe that $j_1^+(s) \overset{(3.1)}{=} (c-2)F_1(s)+1-c \overset{\text{Lem. } 16}{=} (c-2)[1-F_0(m_0+m_1-s)]+1-c \quad = (2-c)F_0(m_0+m_1-s)-1 \overset{(3.1)}{=} j_0^-(m_0+m_1-s)$. Similarly, note that $j_1^-(s) \overset{(3.1)}{=} (2-c)F_1(s)-1 \overset{\text{Lem. } 16}{=} (2-c)[1-F_0(m_0+m_1-s)]-1 \quad = (2-c)F_0(m_0+m_1-s)+1-c \overset{(3.1)}{=} j_0^+(m_0+m_1-s)$. ■

B.1.2 Separability

The following lemma is concerned with the separability of the hypotheses, which is a similar condition to the δ-discriminative condition in Harrison et al. (2012).

Lemma 18. *(separability) There exists $\bar{\delta} > 0$ such that for all $s \in \mathcal{S}$, (i) $F_0(s) - F_1(s) \geq \bar{\delta}$, (ii) $\log \frac{F_0(s)}{F_1(s)} \geq \bar{\delta}$, and (iii) $\log \frac{\bar{F}_1(s)}{\bar{F}_0(s)} \geq \bar{\delta}$.*

Proof. Due to (A1:1), $F_0(s) - F_1(s)$, $\log \frac{F_0(s)}{F_1(s)}$, and $\log \frac{\bar{F}_1(s)}{\bar{F}_0(s)}$ are all continuous functions of s on the compact set $\mathcal{S} = [s_L, s_H]$ and hence obtain minimal values in $\mathcal{S}$. Moreover, by (A1:3), $F_0(s) > F_1(s) > 0$ for all $s \in \mathcal{S}$. Thus, for all $s \in \mathcal{S}$, $F_0(s) - F_1(s) > 0$, $\log \frac{F_0(s)}{F_1(s)} > 0$, and $\log \frac{\bar{F}_1(s)}{\bar{F}_0(s)} > 0$. In particular, the minimal values of all three functions are strictly positive. Based on this, we complete the proof by picking $\bar{\delta} > 0$ such that $\min\left\{ \min_{s\in\mathcal{S}}\{F_0(s) - F_1(s)\}, \min_{s\in\mathcal{S}} \log \frac{F_0(s)}{F_1(s)}, \min_{s\in\mathcal{S}} \log \frac{\bar{F}_1(s)}{\bar{F}_0(s)} \right\} > \bar{\delta} > 0$. ■

B.2 Summary of Algorithms

Algorithm 4: Bayesian policy (BP)

Data: initial belief $b_1 \in (0,1)$, pricing function $s^{\pi_B} : (0,1) \to \mathcal{S}$.

Result: the spread line s_t for each bet t.

$t \leftarrow 1$;

while $t \leq T$ **do**

  $s_t \leftarrow s^{\pi_B}(b_t)$;

  observe bet $d_t \in \{-1, +1\}$;

  if $d_t = +1$ **then**

    $b_{t+1} \leftarrow \frac{b_t \bar{F}_1(s_t)}{b_t \bar{F}_1(s_t) + (1-b_t)\bar{F}_0(s_t)}$;

  else

    $b_{t+1} \leftarrow \frac{b_t F_1(s_t)}{b_t F_1(s_t) + (1-b_t)F_0(s_t)}$;

  end

  $t \leftarrow t + 1$;

end

Algorithm 5: Inertial Policy (IP)

Data: the residual probability sequence $\rho(\cdot) : \mathbb{Z}_+ \to (0, \frac{1}{2} - \alpha)$.

Result: the spread line s_t for each bet t.

$t \leftarrow 1$, $Z_1 \leftarrow 0$;

while $t \leq T$ **do**

  $s_t \leftarrow \tilde{s}(Z_t)$ according to (3.10);

  observe bet $d_t \in \{-1, +1\}$;

  if $d_t = +1$ **then**

    $Z_{t+1} \leftarrow Z_t + 1$;

  else

    $Z_{t+1} \leftarrow Z_t - 1$;

  end

  $t \leftarrow t + 1$;

end

B.3 On the Failure of Bayesian Policies (Theorem 8)

This section provides the details for the proof of Theorem 8.

B.3.1 Roadmap

We first aim to identify profitable strategies for the informed bettor when the market maker uses BPs. Our search for such strategies depends only on the values of $s^{\pi_B}(0+)$ and $s^{\pi_B}(1-)$, i.e., the limiting spread lines as the posterior beliefs converge to $\{0, 1\}$. There are two cases

of the values of $s^{\pi_B}(0+)$ and $s^{\pi_B}(1-)$, each corresponding to a profitable strategy for the informed bettor.

- Case 1 (profitable manipulation): $s^{\pi_B}(0+) \leq m_0$ and $s^{\pi_B}(1-) \geq m_1$. In this case, the informed bettor's honest bets a clear correcting power on the market maker's spread lines. We construct a policy for the informed bettor such that he can still gain a linear profit by mixing bluffing bets and honest bets in a certain manner. We formally state our result regarding Case 1 in Proposition 9 below, which is a direct generalization of Proposition 5. We present the proofs of both propositions in Appendix B.3.4.

  Proposition 9. *(bluffing) Suppose that the market maker uses a Bayesian policy π_B with pricing function $s^{\pi_B}(\cdot)$ such that $s^{\pi_B}(0+) \leq m_0$ and $s^{\pi_B}(1-) \geq m_1$. Then, there exists $\bar{c}_0 = \bar{c}_0(\hat{\Xi}) \in (0,1)$ such that for every initial belief $b_1 \in (0,1)$, hypothesis $i \in \{0,1\}$, and commission rate $c \leq \bar{c}_0$, the type-i informed bettor has a "bluffing" policy ξ_b satisfying the following:*

  1. *(belief and spread line dynamics) The posterior belief b_t converges to $(1-i)$ and the spread line s_t converges to a limit $s_\infty \neq m_i$ almost surely under $\mathbb{P}_i^{\pi_B,\xi_b}$.*
  2. *(linearly growing profit of the informed bettor) $V_i^{\pi_B,\xi_b}(T) = \Omega(T)$.*

- Case 2 (profitable honest betting): $s^{\pi_B}(0+) > m_0$ or $s^{\pi_B}(1-) < m_1$. In this case, the informed bettor's honest bets do not have a sufficiently strong correcting power on the market maker's spread lines. That is, even if a certain type of informed bettor honestly bet all the time, the spread line does not converge to the correct median. Thus the informed bettor can earn a linear profit by simply betting honestly all the time. We formally state our result regarding Case 2 in Proposition 10 below, and present the proof of this proposition in Appendix B.3.5.

  Proposition 10. *(honest betting) Suppose that the market maker uses a Bayesian policy π_B with pricing function $s^{\pi_B}(\cdot)$ such that $s^{\pi_B}(0+) > m_0$ or $s^{\pi_B}(1-) < m_1$.*

Then there exists $\bar{c}_1 = \bar{c}_1(\pi_B, \hat{\Xi})$ *such that for some hypothesis* $i \in \{0, 1\}$*, every initial belief* $b_1 \in (0, 1)$*, and every commission rate* $c \leq \bar{c}_1$*, the type-*i *informed bettor has an "honest" policy* ξ_h *satisfying the following:*

1. *(belief and spread line dynamics) The posterior belief* b_t *converges to* i *and the spread line* s_t *converges to a limit* $s_\infty \neq m_i$ *almost surely under* $\mathbb{P}_i^{\pi_B, \xi_b}$*.*
2. *(linearly growing profit of the informed bettor)* $V_i^{\pi_B, \xi_h}(T) = \Omega(T)$*.*

B.3.2 Proof of Theorem 8

Fix any policy π_B with pricing function $s^{\pi_B}(\cdot)$. Pick $\bar{c}_0 \in (0, 1)$ as in Proposition 9, and hypothesis i and $\bar{c}_1$ as in Proposition 10. Let $\bar{c} := \min\{\bar{c}_0, \bar{c}_1\} \in (0, 1)$, and pick any $c \in (0, \bar{c})$. We first claim that the market maker's regret is linear, i.e., $\liminf_{T\to\infty}\{\frac{1}{T}\Delta^{\pi_B}(T)\} > 0$. Because $0 < c < \bar{c}$, we deduce from Propositions 9 and 10 that the type-i informed bettor has a feasible policy ξ_i such that $\liminf_{T\to\infty}\{\frac{1}{T}V_i^{\pi_B, \xi_i}(T)\} > 0$. The type-$i$ bettor's best response policy, ξ_i^*, maximizes his long-run average profit. Hence, $\liminf_{T\to\infty}\{\frac{1}{T}V_i^{\pi_B, \xi_i^*}(T)\} \geq \liminf_{T\to\infty}\{\frac{1}{T}V_i^{\pi_B, \xi_i}(T)\} > 0$. From the market maker's point of view, her regret is at least the informed bettor's profit, which leads to a linear regret. In fact, we can decompose the bets into two groups: the first group comes from myopic bettors, and the second from

the informed bettor:

$$\begin{aligned}
\Delta_i^{\pi_B,\xi_i^*}(T) &\overset{(3.6)}{=} \tfrac{cT}{2} - \sum_{t=1}^{T} \mathbb{E}_i^{\pi_B,\xi_i^*} \left[\mathbb{I}\{(X-s_t)d_t < 0\} - (1-c)\mathbb{I}\{(X-s_t)d_t > 0\}\right] \\
&= \tfrac{cT}{2} - \sum_{t=1}^{T} \mathbb{E}_i^{\pi_B,\xi_i^*} \mathbb{I}\{a_t = 0\} \left[\mathbb{I}\{(X-s_t)d_t < 0\} - (1-c)\mathbb{I}\{(X-s_t)d_t > 0\}\right] \\
&\quad - \sum_{t=1}^{T} \mathbb{E}_i^{\pi_B,\xi_i^*} \mathbb{I}\{a_t \neq 0\} \left[\mathbb{I}\{(X-s_t)d_t < 0\} - (1-c)\mathbb{I}\{(X-s_t)d_t > 0\}\right] \\
&= \tfrac{cT}{2} - \sum_{t=1}^{T} \mathbb{E}_i^{\pi_B,\xi_i^*} \mathbb{I}\{a_t = 0\} \left[\mathbb{I}\{(X-s_t)\vartheta_t < 0\} - (1-c)\mathbb{I}\{(X-s_t)\vartheta_t > 0\}\right] \\
&\quad - \sum_{t=1}^{T} \mathbb{E}_i^{\pi_B,\xi_i^*} \mathbb{I}\{a_t = +1\} \left[\mathbb{I}\{X < s_t\} - (1-c)\mathbb{I}\{X > s_t\}\right] \\
&\quad - \sum_{t=1}^{T} \mathbb{E}_i^{\pi_B,\xi_i^*} \mathbb{I}\{a_t = -1\} \left[\mathbb{I}\{X > s_t\} - (1-c)\mathbb{I}\{X < s_t\}\right] \\
&= \underbrace{\mathbb{E}_i^{\pi_B,\xi_i^*} \sum_{t=1}^{T} \left[\tfrac{c}{2} - \mathbb{I}\{a_t = 0\} r_i(s_t)\right]}_{\overset{(3.4)}{\geq} 0} \\
&\quad + \underbrace{\mathbb{E}_i^{\pi_B,\xi_i^*} \sum_{t=1}^{T} \left[\mathbb{I}\{a_t = +1\} j_i^+(s_t) + \mathbb{I}\{a_t = -1\} j_i^-(s_t)\right]}_{\overset{(3.2)}{=} V_i^{\pi_B,\xi_i^*}(T)} \\
&\geq V_i^{\pi_B,\xi_i^*}(T).
\end{aligned}$$

In conclusion, the market maker's regret is asymptotically linear in T:

$$\liminf_{T\to\infty} \left\{\tfrac{1}{T}\Delta^{\pi_B}(T)\right\} \geq \liminf_{T\to\infty} \left\{\tfrac{1}{T}\Delta_i^{\pi_B,\xi_i^*}(T)\right\} \geq \liminf_{T\to\infty} \left\{\tfrac{1}{T}V_i^{\pi_B,\xi_i^*}(T)\right\} > 0.$$

We next claim that for some $i \in \{0,1\}$, with positive $\mathbb{P}_i^{\pi_B,\xi_i^*}$-probability, s_t does not converge to m_i. Suppose towards a contradiction that for all $i \in \{0,1\}$, s_t converges to m_i almost surely. It implies that the informed bettor, who makes a linear profit, bets finite times (i.e.,

$\sum_{t=1}^{\infty} \mathbb{I}\{a_t \neq 0\} < \infty$) almost surely. Thus,

$$0 < \liminf_{T\to\infty} \{\tfrac{1}{T} V_i^{\pi_B,\xi_i^*}(T)\} \overset{(a)}{\leq} \lim_{T\to\infty} \left\{ \tfrac{1}{T} \mathbb{E}_i^{\pi_B,\xi_i^*} \left[\textstyle\sum_{t=1}^{T} \mathbb{I}\{a_t \neq 0\} \right] \right\}$$
$$\overset{(b)}{=} \mathbb{E}_i^{\pi_B,\xi_i^*} \left[\lim_{T\to\infty} \tfrac{1}{T} \textstyle\sum_{t=1}^{T} \mathbb{I}\{a_t \neq 0\} \right] = 0,$$

where (a) follows because the informed bettor's profit per bet is at most 1, and (b) follows from the bounded convergence theorem. We have thus arrived at a contradiction. ■

B.3.3 Main Proof Idea: One-stage Analysis

The above proof is based on what we call a "one-stage analysis." That is, if the informed bettor places one (possibly randomized) bet, we characterize the expected impact on the market maker's belief as well as the informed bettor's profit from this single bet. Two functions are of interest throughout the proof. First, define

$$D(b, \mathfrak{p}) := (1 - \mathfrak{p}) \log \left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))} \right) + \mathfrak{p} \log \left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))} \right) \tag{B.1}$$

as the expected increment (i.e., drift) of the market maker's log-likelihood process L_t after a single bet if (i) the current belief state is b and (ii) the informed bettor bets positively with probability $\mathfrak{p}$ and negatively with probability $1 - \mathfrak{p}$. In (B.1), the expectation is taken over the randomized action of the informed bettor. The informed bettor misleads the market maker if $D(b, \mathfrak{p}) < 0$ under H_1 and $D(b, \mathfrak{p}) > 0$ under H_0. Second, let

$$R_i(b, \mathfrak{p}) := (1 - \mathfrak{p}) j_i^-(s^\pi(b)) + \mathfrak{p} j_i^+(s^\pi(b)) \tag{B.2}$$

be the informed bettor's expected profit from a single bet under H_i if (i) the current belief state is b and (ii) the informed bettor bets positively with probability $\mathfrak{p}$ and negatively with probability $1 - \mathfrak{p}$. In (B.2), the expectation is taken over the randomized action of the informed bettor and the final realization of the event outcome X. The type-i informed bettor makes a profit in expectation if $R_i(b, \mathfrak{p}) > 0$.

The following result demonstrates how we utilize the aforementioned one-stage analysis in our proofs. It builds on standard large-deviation based arguments. With a slight abuse

of notation, we let $\{\mathfrak{p}(b)\}$ denote the informed bettor's following *behavioral strategy*: he randomly (and independently) chooses to bet positively with probability $\mathfrak{p}(b)$ and negatively with probability $1-\mathfrak{p}(b)$ given the belief state b.

Lemma 19. *Let $i \in \{0,1\}$. Suppose that the market maker uses a Bayesian policy π_B and the type-i informed bettor's policy ξ is given by the behavioral strategy $\{\mathfrak{p}(b)\}$. Then, we have the following:*

1. *If there exists $\delta > 0$ such that $\mathbb{E}_i^{\pi_B,\xi}[L_{t+1} - L_t | b_t = b] = D(b,\mathfrak{p}(b)) < -\delta$ for all $b \in (0,1)$ and $t \in \mathbb{Z}_+$, then (i) $\mathbb{E}_i^{\pi_B,\xi}[b_t] = O(e^{-\lambda t})$ for some $\lambda > 0$, and (ii) $b_t \to 0$, $L_t \to -\infty$, $s_t \to s^\pi(0+)$ almost surely. If in addition, there exists $\bar{b} \in (0,1)$ such that $R_i(b,\mathfrak{p}(b)) > \delta$ for all $b \in (0,\bar{b}]$, then $V_i^{\pi_B,\xi}(T) = \Omega(T)$.*

2. *If there exists $\delta > 0$ such that $\mathbb{E}_i^{\pi_B,\xi}[L_{t+1} - L_t | b_t = b] = D(b,\mathfrak{p}(b)) > \delta$ for all $b \in (0,1)$ and $t \in \mathbb{Z}_+$, then (i) $\mathbb{E}_i^{\pi_B,\xi}[1 - b_t] = O(e^{-\lambda t})$ for some $\lambda > 0$, and (ii) $b_t \to 1$, $L_t \to \infty$, $s_t \to s^\pi(1-)$ almost surely. If in addition, there exists $\bar{b} \in (0,1)$ such that $R_i(b,\mathfrak{p}(b)) > \delta$ for all $b \in [\bar{b},1)$, then $V_i^{\pi_B,\xi}(T) = \Omega(T)$.*

In our proofs, we employ our one-stage analysis in different settings, and combine it with Lemma 19 to obtain desired results. Specifically, we use the one-stage analysis in Lemma 20, in Step 1 in the proof of Proposition 10 in Appendix B.3.5, and in Step 1 in the proof of Theorem 9 in Appendix B.4.1. We combine these instances of the one-stage analysis with Lemma 19 to prove Propositions 9 and 10 as well as Theorem 9.

Proof of Lemma 19. Without loss of generality, suppose that $D(b,\mathfrak{p}(b)) < -\delta$ for all $b \in (0,1)$. The proof for the other case follows by repeating the same arguments verbatim. We complete the proof in two steps.

Step 1: concentration and convergence for $\{L_t\}$. We claim that there exists $\varepsilon > 0$ such that $\mathbb{P}_i^{\pi,\xi}(L_t \geq -\frac{\delta t}{2}) \leq \exp(-\varepsilon t)$ for $t \in \mathbb{Z}_+$, in which case the following hold: (i) $\mathbb{E}_i^{\pi_B,\xi}[b_t] = O(e^{-\lambda t})$ for some $\lambda > 0$; (ii) $L_t \to -\infty$ almost surely; (iii) $b_t \to 0$ almost surely; and (iv) $s_t \to s^\pi(0+)$ almost surely.

Observe that the market maker's pricing policy π_B is Markovian with respect to the belief state b_t. The informed bettor's behavioral strategy $\{\mathfrak{p}(b)\}$ is also Markovian. As a result, b_t is a Markov chain under $\mathbb{P}_i^{\pi_B,\xi}$, and so is the market maker's log-likelihood ratio process L_t.

We apply Doob's decomposition to the process L_t. That is, we define

$$\mathfrak{A}_t := \sum_{\ell=1}^{t} \mathbb{E}_i^{\pi_B,\xi}[L_\ell - L_{\ell-1}|L_{\ell-1}] \text{ and } \mathfrak{M}_t := \sum_{\ell=1}^{t} \left(L_\ell - \mathbb{E}_i^{\pi_B,\xi}[L_\ell|L_{\ell-1}]\right)$$

so that $L_t = \mathfrak{A}_t + \mathfrak{M}_t$ for all t. We may interpret $\mathfrak{A}_t$ as the "drift" of L_t and $\mathfrak{M}_t$ as its "noise." Because $D(b, \mathfrak{p}(b)) < -\delta$ for all $b \in (0,1)$, $\mathfrak{A}_t < -\delta t$ almost surely. Moreover, $\mathfrak{M}_t$ is a martingale with bounded increments: note that

$$\begin{aligned}
&|\mathfrak{M}_\ell - \mathfrak{M}_{\ell-1}| \\
\leq &\left|L_\ell - \mathbb{E}_i^{\pi_B,\xi}[L_\ell|L_{\ell-1}] - L_{\ell-1} + \mathbb{E}_i^{\pi_B,\xi}[L_{\ell-1}|L_{\ell-2}]\right| \\
= &\left|L_\ell - \mathbb{E}_i^{\pi_B,\xi}[L_\ell - L_{\ell-1}|L_{\ell-1}] - 2L_{\ell-1} + \mathbb{E}_i^{\pi_B,\xi}[L_{\ell-1} - L_{\ell-2}|L_{\ell-2}] + L_{\ell-2}\right| \\
\leq &|L_\ell - L_{\ell-1}| + |L_{\ell-1} - L_{\ell-2}| + \mathbb{E}_i^{\pi_B,\xi}[|L_\ell - L_{\ell-1}||L_{\ell-1}] + \mathbb{E}_i^{\pi_B,\xi}[|L_{\ell-1} - L_{\ell-2}||L_{\ell-2}] \\
\overset{(a)}{\leq} &4M,
\end{aligned}$$

where (a) follows by defining the constant

$$M := \max\left\{\sup_{s\in\mathcal{S}} \log\left(\frac{F_0(s)}{F_1(s)}\right), \sup_{s\in\mathcal{S}} \log\left(\frac{\bar{F}_1(s)}{\bar{F}_0(s)}\right)\right\} \leq \max\left\{\log\frac{1}{F_0(s_L)}, \log\frac{1}{F_1(s_H)}\right\},$$

which is finite due to Assumption (A1:3). As a result, L_t is a Markov chain with a non-vanishing drift and bounded increments.

Finally, we deduce the following for all $t \in \mathbb{Z}_+$:

$$
\begin{aligned}
&\mathbb{P}_i^{\pi_B,\xi}(L_t \geq -\tfrac{\delta t}{2}) \\
&\quad = \mathbb{P}_i^{\pi_B,\xi}\left(\mathfrak{A}_t + \mathfrak{M}_t \geq -\tfrac{\delta t}{2}\right) \\
&\quad \leq \mathbb{P}_i^{\pi_B,\xi}\left(\mathfrak{M}_t \geq \tfrac{\delta t}{2}\right) && [\mathfrak{A}_t < -\delta t \text{ almost surely}] \\
&\quad \leq \exp\left(-\tfrac{t^2}{2t(64M^2/\delta^2)}\right) && [\text{by Azuma-Hoeffding inequality; } \left|\tfrac{2\mathfrak{M}_t}{\delta} - \tfrac{2\mathfrak{M}_{t-1}}{\delta}\right| \leq \tfrac{8M}{\delta}\ ] \\
&\quad = \exp\left(-\tfrac{t}{128M^2/\delta^2}\right) \\
&\quad = \exp\left(-\varepsilon t\right) && [\varepsilon := \tfrac{\delta^2}{128M^2} > 0]
\end{aligned}
$$

The four facts mentioned in the beginning of this step, namely, (i) $\mathbb{E}_i^{\pi_B,\xi}[b_t] = O(e^{-\lambda t})$ for some $\lambda > 0$; (ii) $L_t \to -\infty$ almost surely; (iii) $b_t \to 0$ almost surely; and (iv) $s_t \to s^\pi(0+)$ almost surely, follow from the above analysis in a standard manner; see, e.g., the proof of Proposition 7 in Harrison et al. (2012).

Step 2: profit evaluation. Suppose that $R_i(b, \mathfrak{p}(b)) > \delta$ for all $b \in (0, \bar{b}]$. We claim that $\liminf_{T\to\infty}\{\frac{1}{T}V_i^{\pi_B,\xi}\} > 0$. Pick $\bar{L} := \log\left(\frac{\bar{b}}{1-\bar{b}}\right) - \log\left(\frac{b_1}{1-b_1}\right)$ so that $L_t \geq \bar{L}$ if and only if $b_t \geq \bar{b}$. Note that $\bar{L}$ is finite because $b_1 \in (0,1)$ and $\bar{b} \in (0,1)$. Now, let us evaluate the

type-i informed bettor's payoff:

$$
\begin{aligned}
&V_i^{\pi_B,\xi}(T) \\
&= \sum_{t=1}^{T} \Big(\mathbb{E}_i^{\pi_B,\xi}[R_i(b_t, \mathfrak{p}(b_t))] \mathbb{I}\{b_t \le \bar{b}\} \\
&+ \mathbb{E}_i^{\pi_B,\xi}[R_i(b_t, \mathfrak{p}(b_t))] \mathbb{I}\{b_t > \bar{b}\} \Big) \\
&\ge \sum_{t=1}^{T} \Big(\mathbb{E}_i^{\pi_B,\xi}[\delta\, \mathbb{I}\{b_t \le \bar{b}\}] + \mathbb{E}_i^{\pi_B,\xi}[(-1)\mathbb{I}\{b_t > \bar{b}\}] \Big) \qquad [R_i(b, \mathfrak{p}(b)) > \delta \ \forall b \in (0, \bar{b}]] \\
&= \delta T - (1+\delta) \sum_{t=1}^{T} \mathbb{P}_i^{\pi_B,\xi}(b_t > \bar{b}) \\
&\ge \delta T - (1+\delta) \sum_{t=1}^{\infty} \mathbb{P}_i^{\pi_B,\xi}(b_t > \bar{b}) \\
&\ge \delta T - (1+\delta) \sum_{t=1}^{\infty} \mathbb{P}_i^{\pi_B,\xi}(L_t \ge \bar{L})
\end{aligned}
$$

We deduce from Step 1 that $(1+\delta)\sum_{t=1}^{\infty} \mathbb{P}_i^{\pi_B,\xi}(L_t \ge \bar{L})$ is a finite constant that is independent of T. Therefore, $\liminf_{T\to\infty}\{\frac{1}{T} V_i^{\pi_B,\xi}\} \ge \delta > 0$. ■

B.3.4 Profitable Manipulation (Proofs of Propositions 5 and 9)

Let π_B be the market maker's Bayesian policy satisfying $s^{\pi_B}(0+) \le m_0$ and $s^{\pi_B}(1-) \ge m_1$. The existence of $s^{\pi_B}(0+)$ and $s^{\pi_B}(1-)$ are guaranteed by the definition of a Bayesian policy in our setting. Roughly speaking, the proofs of Propositions 5 and 9 rely on the construction of a strategy for the informed bettor that randomizes between bluffing and honest betting. Under such a strategy, the informed bettor keeps misleading the market maker while making profits. To formalize this idea, we recall that $\Xi = (c, m_0, m_1, F_\epsilon)$ is the collection of problem input parameters and $\alpha = F_1(m_0)$, and state the following auxiliary result.

Lemma 20. *(one-stage analysis for manipulation) There exist $\bar{c}_0, \mathfrak{p}_0 \in (0,1)$, which depend only on α, such that for all $c \in (0, \bar{c}_0]$, there exist $\bar{b} = \bar{b}(\Xi, \mathfrak{p}_0) \in (0,1)$ and $\delta = \delta(\Xi, \mathfrak{p}_0) > 0$ satisfying the following:*

1. *(global manipulability) For all* $b \in (0,1)$, $D(b,0) < -\delta$ *and* $D(b,1) > \delta$.

2. *(local profitable manipulation; type-*1*) For all* $b \in (0, \bar{b}]$, $D(b, \mathfrak{p}_0) < -\delta$ *and* $R_1(b, \mathfrak{p}_0) > \delta$.

3. *(local profitable manipulation; type-*0*) For all* $b \in [1-\bar{b}, 1)$, $D(b, 1-\mathfrak{p}_0) > \delta$ *and* $R_0(b, 1-\mathfrak{p}_0) > \delta$.

Proof of Proposition 9. We focus on the type-1 bettor, as the proof for the type-0 bettor follows from the same arguments. Let $b_1 \in (0,1)$, and choose $\bar{c}_0, c, \mathfrak{p}_0, \bar{b}, \delta$ as in Lemma 20. In particular, $\bar{c}_0$ depends only on α (and hence on $\hat{\Xi}$). Consider the following (behavioral) betting strategy ξ_b for the type-1 informed bettor: under ξ_b, $\mathfrak{p}(b) = \mathfrak{p}_0 \mathbb{I}\{b \leq \bar{b}\}$. That is, the probability that he bets positively is $\mathfrak{p}_0$ if $b_t \leq \bar{b}$ and 0 otherwise. Lemma 20 implies that $\mathbb{E}_i^{\pi_B, \xi_b}[L_{t+1} - L_t | b_t = b] = D(b, \mathfrak{p}(b)) = D(b, \mathfrak{p}_0)\, \mathbb{I}\{b \leq \bar{b}\} + D(b,0)\, \mathbb{I}\{b > \bar{b}\} < -\delta$ for all $b \in (0,1)$ and $t \in \mathbb{Z}_+$, and $R_1(b, \mathfrak{p}(b)) > \delta$ for all $b \in (0, \bar{b}]$. Hence, we are in the first case in the statement of Lemma 19. As a result, $b_t \to 0$, $L_t \to -\infty$, and $s_t \to s^\pi(0+)$ almost surely, as well as $V_1^{\pi_B, \xi}(T) = \Omega(T)$. ■

Proof of Proposition 5. Proposition 5 is a special case of Proposition 9, because we focus on the case where $s^\pi(0+) = m_0$ and $s^\pi(1-) = m_1$ in Proposition 5 while we consider all possible cases satisfying $s^\pi(0+) \leq m_0$ and $s^\pi(1-) \geq m_1$ in Proposition 9. ■

Proof of Lemma 20. We complete the proof in four steps.

Step 1. We claim that there exists $\delta_1 = \delta_1(m_0, m_1, F_\epsilon)$ such that (i) $D(b,0) < -\delta_1$ and (ii) $D(b,1) > \delta_1$ for all $b \in (0,1)$. By Lemma 18, there exists $\bar{\delta} > 0$ such that $D(b,0) = \log\left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) \leq \sup_{s \in \mathcal{S}} \log\left(\frac{F_1(s)}{F_0(s)}\right) = -\inf_{s \in \mathcal{S}} \log\left(\frac{F_0(s)}{F_1(s)}\right) \leq -\bar{\delta}$ for all $b \in [0,1]$. Similarly, $D(b,1) = \log\left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right) \geq \inf_{s \in \mathcal{S}} \log\left(\frac{\bar{F}_1(s)}{\bar{F}_0(s)}\right) \geq \bar{\delta}$. To prove our claim in this step, we choose $\delta_1 = \frac{\bar{\delta}}{2}$.

Step 2. We claim that $\bar{c}_0, \mathfrak{p}_0 \in (0,1)$, which depend only on α, satisfying the following for all $c \in (0, \bar{c}_0]$:

- $D(0+, \mathfrak{p}_0) < 0$ and $R_1(0+, \mathfrak{p}_0) > 0$;

- $D(1-, 1-\mathfrak{p}_0) > 0$ and $R_0(1-, 1-\mathfrak{p}_0) > 0$.

In other words, the type-1 (resp. type-0) informed bettor can enjoy profitable manipulation when the market maker's belief is close to 0 (resp. 1). To prove this claim, let us first introduce some constants. Define $\kappa := \frac{-\log 2\alpha}{\log 2(1-\alpha)}$, $\bar{c}_0 := \frac{(\kappa-1)(1-2\alpha)}{2(\kappa-1)(1-\alpha)+1}$, and $\hat{\kappa} := \frac{(\bar{c}_0-2)\alpha+1}{(\bar{c}_0-2)\alpha+1-\bar{c}_0}$. By definition, all of the three constants depend only on α. The following auxiliary result below summarizes the relationship among these constants.

Lemma 21. *(ranges of and relations between $\kappa, \hat{\kappa}$ and $\bar{c}_0$) We have $\bar{c}_0 \in (0,1)$. Moreover, for all $c \in (0, \bar{c}_0]$, $\kappa > \hat{\kappa} \geq \frac{(c-2)\alpha+1}{(c-2)\alpha+1-c} > 1$.*

In light of the result above, we choose $\mathfrak{p}_0$ so that $\kappa > \frac{\mathfrak{p}_0}{1-\mathfrak{p}_0} > \hat{\kappa}$. For example, we can choose $\mathfrak{p}_0$ as the solution to the equation $\frac{\mathfrak{p}}{1-\mathfrak{p}} = \frac{\kappa+\hat{\kappa}}{2}$. Such a construction is valid because $\kappa > \hat{\kappa}$ and the mapping $\mathfrak{p} \mapsto \frac{\mathfrak{p}}{1-\mathfrak{p}}$ maps $(0,1)$ onto $(0,\infty)$. Since κ and $\hat{\kappa}$ depend only on α, so does $\mathfrak{p}_0$. Intuitively, we may interpret $\mathfrak{p}_0$ as the probability of honest betting (instead of bluffing), which means positive betting for the type-1 informed bettor and negative betting for the type-0 informed bettor. Similarly, we may interpret $\frac{\mathfrak{p}_0}{1-\mathfrak{p}_0}$ as the probability ratio of honest betting over bluffing. The constants κ and $\hat{\kappa}$ are respectively upper and lower benchmarks for this ratio: if bluffing is sufficiently frequent (i.e., $\frac{\mathfrak{p}_0}{1-\mathfrak{p}_0} < \kappa$) then manipulation happens, and if honest betting is sufficiently frequent (i.e., $\frac{\mathfrak{p}_0}{1-\mathfrak{p}_0} > \hat{\kappa}$) then manipulation is profitable for the informed bettor. A more detailed derivation is presented below. The fact that there is a strict gap between κ and $\hat{\kappa}$ is a key construct in this proof, ensuring that the strategy ξ_b achieves profitable manipulation.

Second, we verify that $D(0+, \mathfrak{p}_0) < 0$ and $D(1-, 1-\mathfrak{p}_0) > 0$. In other words, as the informed bettor bluffs with high probability (i.e., the ratio of honest betting over bluffing

less than κ), he misleads the market maker. To see this, observe that

$$
\begin{aligned}
& D(0+, \mathfrak{p}_0) \\
= \; & (1-\mathfrak{p}_0) \log\left(\tfrac{F_1(s^\pi(0+))}{F_0(s^\pi(0+))}\right) + \mathfrak{p}_0 \log\left(\tfrac{\bar{F}_1(s^\pi(0+))}{\bar{F}_0(s^\pi(0+))}\right) \\
< \; & \tfrac{1}{1+\kappa} \log\left(\tfrac{F_1(s^\pi(0+))}{F_0(s^\pi(0+))}\right) + \tfrac{\kappa}{1+\kappa} \log\left(\tfrac{\bar{F}_1(s^\pi(0+))}{\bar{F}_0(s^\pi(0+))}\right) && [\mathfrak{p}_0 < \tfrac{\kappa}{\kappa+1}] \\
= \; & H\big(F_1\left(s^\pi\left(0+\right)\right)\big) - H\big(F_0\left(s^\pi\left(0+\right)\right)\big) && [H(x) := \tfrac{1}{\kappa+1}\log(x) + \tfrac{\kappa}{\kappa+1}\log(1-x)] \\
\leq \; & H(x_0) - H\big(F_0\left(s^\pi\left(0+\right)\right)\big) && [x_0 := F_1\left(s^\pi\left(0+\right)\right) \leq \alpha] \\
\overset{(a)}{=} \; & \big(H(x_0) - H(\tfrac{1}{2})\big) \vee 0 \\
\overset{(b)}{\leq} \; & \big(H(\alpha) - H(\tfrac{1}{2})\big) \vee 0 \overset{(c)}{=} 0 \vee 0 = 0.
\end{aligned}
$$

To derive part (a) above, we use the following two facts: (i) $H(\cdot)$ increases in the region $(0, \frac{1}{\kappa+1})$ and decreases in the region $(\frac{1}{\kappa+1}, 1)$, and hence is a quasi-concave function; and (ii) $x_0 = F_1\left(s^\pi\left(0+\right)\right) \leq F_0\left(s^\pi\left(0+\right)\right) \leq F_0\left(m_0\right) = \frac{1}{2}$. These two facts imply that $H\big(F_0\left(s^\pi\left(0+\right)\right)\big) \geq H(x_0) \wedge H(\frac{1}{2})$. Rearranging terms, we deduce that $H(x_0) - H\big(F_0\left(s^\pi\left(0+\right)\right)\big) \leq \big(H(x_0) - H(\frac{1}{2})\big) \vee 0$. For part (b), we use two facts as well. First, $H(\alpha) - H(\frac{1}{2}) = \frac{1}{\kappa+1}\log(2\alpha) + \frac{\kappa}{\kappa+1}\log(2(1-\alpha)) = 0$, implying that $H(\cdot)$ is increasing in the region $(0, \alpha)$. Second, $x_0 = F_1\left(s^\pi\left(0+\right)\right) \leq F_1(m_0) = \alpha$. Thus, $H(x_0) \leq H(\alpha)$. Part (c) follows because $H(\alpha) = H(\frac{1}{2})$. Similarly,

$$
\begin{aligned}
& D(1-, 1-\mathfrak{p}_0) \\
& \quad = \; \mathfrak{p}_0 \log\left(\tfrac{F_1(s^\pi(1-))}{F_0(s^\pi(1-))}\right) + (1-\mathfrak{p}_0) \log\left(\tfrac{\bar{F}_1(s^\pi(1-))}{\bar{F}_0(s^\pi(1-))}\right) \\
& \quad > \; \tfrac{\kappa}{1+\kappa} \log\left(\tfrac{F_1(s^\pi(1-))}{F_0(s^\pi(1-))}\right) + \tfrac{1}{1+\kappa} \log\left(\tfrac{\bar{F}_1(s^\pi(1-))}{\bar{F}_0(s^\pi(1-))}\right) && [\mathfrak{p}_0 < \tfrac{\kappa}{\kappa+1}] \\
& \quad = \; H\big(\bar{F}_1\left(s^\pi\left(1-\right)\right)\big) - H\big(\bar{F}_0\left(s^\pi\left(1-\right)\right)\big) \\
& \quad \geq \; \big(H(\tfrac{1}{2}) \vee H(y_0)\big) - H(y_0) && [y_0 := \bar{F}_0\left(s^\pi\left(1-\right)\right) \leq \alpha] \\
& \quad = \; \big(H(\tfrac{1}{2}) - H(y_0)\big) \vee 0 \\
& \quad \geq \; \big(H(\tfrac{1}{2}) - H(\alpha)\big) \vee 0 = 0.
\end{aligned}
$$

Third, we verify that for all $c \in (0, \bar{c}_0]$, $R_1(0+, \mathfrak{p}_0) > 0$ and $R_1(1-, 1-\mathfrak{p}_0) > 0$. In other

words, If the informed bettor bets honestly with high probability (i.e., the ratio of honest betting over bluffing higher than $\hat{\kappa}$), he makes profits from every bet in expectation. To see this point, note that

$$\begin{aligned}
R_1(0+,\mathfrak{p}_0) &= (1-\mathfrak{p}_0)j_1^-(s^\pi(0+)) + \mathfrak{p}_0 j_1^+(s^\pi(0+)) \\
&= (1-\mathfrak{p}_0)\left[(2-c)F_1\left(s^\pi(0+)\right)-1\right] + \mathfrak{p}_0\left[(c-2)F_1\left(s^\pi(0+)\right)+1-c\right] \quad \text{[by (3.1)]} \\
&\overset{(d)}{\geq} (1-\mathfrak{p}_0)\left[(2-c)F_1\left(m_0\right)-1\right] + \mathfrak{p}_0\left[(c-2)F_1\left(m_0\right)+1-c\right] \\
&= (1-\mathfrak{p}_0)\left[(2-c)\alpha-1\right] + \mathfrak{p}_0\left[(c-2)\alpha+1-c\right] \overset{(e)}{>} 0,
\end{aligned}$$

where (d) follows because $\frac{\partial}{\partial F_1(s^\pi(0+))}R_1(0+,\mathfrak{p}_0) = (2-c)(1-2\mathfrak{p}_0) < 0$ as $\frac{\mathfrak{p}_0}{1-\mathfrak{p}_0} > \hat{\kappa} > 1$ by construction, and (e) follows because $\frac{\mathfrak{p}_0}{1-\mathfrak{p}_0} > \hat{\kappa} = \frac{(\bar{c}_0-2)\alpha+1}{(\bar{c}_0-2)\alpha+1-\bar{c}_0} \geq \frac{(c-2)\alpha+1}{(c-2)\alpha+1-c}$ (see Lemma 21). Similarly,

$$\begin{aligned}
&R_0(1-,1-\mathfrak{p}_0) \\
&= \mathfrak{p}_0 j_0^-(s^\pi(1-)) + (1-\mathfrak{p}_0)j_0^+(s^\pi(1-)) \\
&= \mathfrak{p}_0\left[(2-c)F_0\left(s^\pi(1-)\right)-1\right] + (1-\mathfrak{p}_0)\left[(c-2)F_0\left(s^\pi(1-)\right)+1-c\right] \qquad \text{[by (3.1)]} \\
&\geq \mathfrak{p}_0\left[(2-c)F_0\left(m_1\right)-1\right] + (1-\mathfrak{p}_0)\left[(c-2)F_0\left(m_1\right)+1-c\right] \\
&= \mathfrak{p}_0\left[(2-c)(1-\alpha)-1\right] + (1-\mathfrak{p}_0)\left[(c-2)(1-\alpha)+1-c\right] \\
&= \mathfrak{p}_0\left[(c-2)\alpha+1-c\right] + (1-\mathfrak{p}_0)\left[(2-c)\alpha-1\right] > 0.
\end{aligned}$$

The preceding derivations confirm that $D(0+,\mathfrak{p}_0)$, $D(1-,1-\mathfrak{p}_0)$, $R_1(0+,\mathfrak{p}_0)$, and $R_0(1-,1-\mathfrak{p}_0)$ are all well-defined as $s^\pi(0+)$ and $s^\pi(1-)$ exist.

Step 3. By Step 2, there exists $\bar{b}, \delta_2, \delta_3, \delta_4, \delta_5 > 0$, all of which depend only on $\mathfrak{p}_0$ and Ξ, such that

- *(local profitable manipulation; type-1)* $D(b,\mathfrak{p}_0) < -\delta_2$ and $R_1(b,\mathfrak{p}_0) > \delta_3$ for all $b \in (0,\bar{b}]$;
- *(local profitable manipulation; type-0)* $D(b,1-\mathfrak{p}_0) > \delta_4$ and $R_0(b,1-\mathfrak{p}_0) > \delta_5$ for all $b \in (1-\bar{b},1]$.

The existence is guaranteed by the local continuity of $D(b, \mathfrak{p}_0)$ and $R_1(b, \mathfrak{p}_0)$ with respect to b at 0+, as well as that of $D(b, 1-\mathfrak{p}_0)$ and $R_0(b, 1-\mathfrak{p}_0)$ with respect to b at 1−.

Step 4. Based on Steps 1 and 3, we complete the proof by choosing

$$\delta := \min\{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5\}.$$

■

Proof of Lemma 21. To prove that $\bar{c}_0 > 0$, it suffices to verify that $\kappa > 1$. Note that $\alpha \in (0, \frac{1}{2})$, and $\log\big(2(1-\alpha)\big) > 0 > \log 2\alpha$. Moreover, due to Jensen's inequality, $\log 2(1-\alpha) + \log 2\alpha < 2\log\frac{2(1-\alpha)+2\alpha}{2} = 0$. Thus, $\kappa = \frac{-\log 2\alpha}{\log 2(1-\alpha)} > 1$, and $\bar{c}_0 > 0$. To see why $\bar{c}_0 < 1$, note that $\bar{c}_0 = \frac{(\kappa-1)(1-2\alpha)}{2(\kappa-1)(1-\alpha)+1} < \frac{(\kappa-1)(1-2\alpha)}{(\kappa-1)(1-\alpha)+1} < \frac{(\kappa-1)(1-\alpha)}{(\kappa-1)(1-\alpha)+1} < 1$. We have thus verified that $\bar{c}_0 \in (0,1)$. Now, we choose $c \in (0, \bar{c}_0]$ to verify that $\kappa > \hat{\kappa} \geq \frac{(c-2)\alpha+1}{(c-2)\alpha+1-c} > 1$. To see why $\frac{(c-2)\alpha+1}{(c-2)\alpha+1-c} > 1$, note that $c \leq \bar{c}_0 < \frac{1-2\alpha}{1-\alpha} \Rightarrow (1-c) + (c-2)\alpha > 0$. Moreover, $\big[(c-2)\alpha+1\big] - \big[(c-2)\alpha+1-c\big] = c > 0$. Hence, $\frac{(c-2)\alpha+1}{(c-2)\alpha+1-c} > 1$. To see why $\hat{\kappa} \geq \frac{(c-2)\alpha+1}{(c-2)\alpha+1-c}$, observe that the function $c \mapsto \frac{(c-2)\alpha+1}{(c-2)\alpha+1-c}$ increases in c because $\alpha < 1$. As a result, $c \leq \bar{c}_0$, which implies that $\frac{(c-2)\alpha+1}{(c-2)\alpha+1-c} \leq \frac{(\bar{c}_0-2)\alpha+1}{(\bar{c}_0-2)\alpha+1-\bar{c}_0} = \hat{\kappa}$. Finally, let us verify that $\kappa > \hat{\kappa}$. Note that $\bar{c}_0 = \frac{(\kappa-1)(1-2\alpha)}{2(\kappa-1)(1-\alpha)+1} < \frac{(\kappa-1)(1-2\alpha)}{(\kappa-1)(1-\alpha)+1}$. By rearranging terms, we see that

$$\begin{aligned}
(\kappa-1)(1-2\alpha) > [(1-\alpha)(\kappa-1)+1]\bar{c}_0 &\implies (1-2\alpha)\kappa + 2\alpha - 1 > (1-\alpha)\bar{c}_0\kappa + \alpha\bar{c}_0 \\
&\implies [(1-\bar{c}_0) + (\bar{c}_0-2)\alpha]\kappa > 1 + (\bar{c}_0-2)\alpha \\
&\implies \kappa > \tfrac{1+(\bar{c}_0-2)\alpha}{(1-\bar{c}_0)+(\bar{c}_0-2)\alpha} = \hat{\kappa}.
\end{aligned}$$

We have thus completed the proof. ■

B.3.5 Profitable Honest Betting (Proof of Proposition 10)

Proof of Proposition 10. Choose

$$\bar{c}_1 := \min\left\{\max\left\{\tfrac{2F_0(s^{\pi_B}(0+))-1}{2F_0(s^{\pi_B}(0+))}, \tfrac{1-2F_1(s^{\pi_B}(1-))}{2F_1(s^{\pi_B}(1-))}\right\}, \tfrac{1}{2}\right\}. \tag{B.3}$$

By construction, $\bar{c}_1 \in (0,1)$, because either $s^{\pi_B}(0+) > m_0$ or $s^{\pi_B}(1-) < m_1$. Let $c \in (0, \bar{c}_1]$.

First, suppose that $s^{\pi_B}(0+) > m_0$. We define the type-0 informed bettor's "honest"

strategy ξ_h as always betting negatively, i.e., under ξ_h, $\mathfrak{p}(b) = 0$ for all $b \in (0,1)$. We claim that there exists $\delta, \bar{b} > 0$ such that $\mathbb{E}_0^{\pi_B,\xi_h}[L_{t+1} - L_t | b_t = b] = D(b,0) < -\delta$ for all $b \in (0,1)$ and $t \in \mathbb{Z}_+$, and $R_0(b, \mathfrak{p}(b)) = R_0(b,0) > 0$ for all $b \in (0, \bar{b}]$. To that end, we deduce from Lemma 18 that there exists $\bar{\delta} > 0$ such that $D(b,0) = \log \frac{F_1(s^\pi(b))}{F_0(s^\pi(b))} \leq -\bar{\delta}$. Note that $F_0(s^{\pi_B}(0+)) > F_0(m_0) = \frac{1}{2}$, and $0 < c < \frac{2F_0(s^{\pi_B}(0+))-1}{F_0(s^{\pi_B}(0+))}$. In that case, $R_0(0+,0) = j_0^-(s^{\pi_B}(0+)) = (2-c)F_0(s^{\pi_B}(0+)) - 1 = [2F_0(s^{\pi_B}(0+)) - 1] - cF_0(s^{\pi_B}(0+)) > 0$. Hence, there exist $\varepsilon > 0$ and $\bar{b}$ (independent of T) such that $R_0(b,0) \geq \varepsilon$ for all $b \in (0, \bar{b}]$. Choosing $\delta = \min\{\bar{\delta}, \varepsilon\}$, we deduce from Lemma 19 that $b_t \to 0$ and $s_t \to s^{\pi_B}(0+)$ almost surely, and that $V_0^{\pi_B,\xi_h}(T) = \Omega(T)$.

In the case where $s^{\pi_B}(1-) < m_1$, our analysis is similar. In fact, type-1 informed bettor's honest policy ξ_h is specified as always betting positively. Notice that $0 < c < \frac{1-2F_1(s^{\pi_B}(1-))}{F_1(s^{\pi_B}(1-))}$; thus, $R_1(0+,0) = j_1^+(s^{\pi_B}(1-)) = (c-2)F_1(s^{\pi_B}(1-)) + 1 - c = [1 - 2F_1(s^{\pi_B}(1-))] - c\bar{F}_1\big(s^{\pi_B}(1-)\big) > 0$. The rest of the proof for this case follows by repeating the above arguments. ■

B.4 On the Success of Bayesian Policies (Theorem 9)

This section provides the details for the proof of Theorem 9, as well as additional discussions on the myopic Bayesian policy (MBP) as a special case of BP. In what follows, we use the following little-o notation: for all functions f, g defined in a neighborhood around zero, we say that $f(x) = o(g(x))$ if $\lim_{x\to 0} \frac{f(x)}{g(x)} = 0$.

B.4.1 Proof of Theorem 9

Let $s^{\pi_B}(\cdot)$ be a regular pricing function and $b_1 \in (0,1)$. We assume without loss of generality that $i = 0$ (the analysis that follows can be repeated verbatim for the case where $i = 1$). Throughout this proof, we also denote $\mathbb{P}_0^{\pi_B,\xi_\emptyset}(\cdot)$ and $\mathbb{E}_0^{\pi_B,\xi_\emptyset}[\cdot]$ as $\mathbb{P}_0(\cdot)$ and $\mathbb{E}_0[\cdot]$ for brevity. We complete the proof in three steps.

Step 1. We claim that there exists $\delta > 0$ such that $\mathbb{E}_0[L_{t+1} - L_t | b_t = b] < -\delta$ for all $b \in (0,1)$ and $t \in \mathbb{Z}_+$. To prove this claim, we deduce from Lemma 18 that π_B is a $\bar{\delta}$-discriminative policy (in the sense of Harrison et al., 2012) for some $\bar{\delta} > 0$. By Lemma A.2 in Harrison et al. (2012), $\mathbb{E}_0[L_{t+1} - L_t | b_t = b] < -2\bar{\delta}^2$ for all $b \in (0,1)$ and $t \in \mathbb{Z}_+$. As a result, Lemma 19 implies that $\mathbb{E}_0[b_t] = O(e^{-\lambda t})$ for some $\lambda > 0$, and $s_t \to s^\pi(0+)$ almost surely.

Step 2. We claim that $\mathfrak{d}_t \to 0$ almost surely, and $\mathbb{E}_0[\mathfrak{d}_t] = O\big(e^{-\lambda t}\big)$, thus verifying Statements (T9:1) and (T9:2) in Theorem 9. To that end, choose $\mu_0 > 0$ such that $\mathbb{E}_0[b_t] \leq \mu_0 e^{-\lambda t}$ for all $t \in \mathbb{Z}_+$. Since $s^{\pi_B}(\cdot)$ is regular, $\limsup_{b \downarrow 0} \frac{|s^{\pi_B}(b) - m_0|}{b} < \infty$. That is, there exists $C > 0$ such that $|s^{\pi_B}(b) - m_0| < bC$ for all $b \in (0, \delta)$. Hence, $\mathfrak{d}_t = |s^{\pi_B}(b_t) - m_0| \to 0$ $\mathbb{P}_0$-almost surely. Moreover,

$$\begin{aligned}
\mathbb{E}_0[\mathfrak{d}_t] &= \underbrace{\mathbb{E}_0[\mathfrak{d}_t \big| b_t \geq \delta]}_{\leq s_H - s_L} \underbrace{\mathbb{P}_0(b_t \geq \delta)}_{\leq \frac{\mathbb{E}_0[b_t]}{\delta}} + \underbrace{\mathbb{E}_0[\mathfrak{d}_t | b_t \leq \delta]}_{\leq \mathbb{E}_0[b_t] C} \underbrace{\mathbb{P}_0(b_t \leq \delta)}_{\leq 1} \\
&\leq \tfrac{(s_H - s_L)\mathbb{E}_0[b_t]}{\delta} + \mathbb{E}_0[b_t] C = \big[\tfrac{s_H - s_L}{\delta} + C\big] \mathbb{E}_0[b_t] \leq \big[\tfrac{s_H - s_L}{\delta} + C\big] \mu_0 e^{-\lambda t}.
\end{aligned}$$

Therefore, $\mathbb{E}_0[\mathfrak{d}_t] = O\big(e^{-\lambda t}\big)$.

Step 3. We claim that $\Delta_0^{\pi_B, \xi_\emptyset}(T) = O(1)$, verifying Statement (T9:3) in Theorem 9. For this purpose, observe that $\frac{c}{2} - r_0(s^\pi(b)) = o(b)$ and $\frac{c}{2} - r_1(s^\pi(b)) = o(1-b)$. To see this, recall from (3.4) that $r_i(s) = (2c-4)\left(F_i(s) - \frac{1}{2}\right)^2 + \frac{c}{2}$ for $i \in \{0,1\}$. Hence, $r_i'(s) = (2c-4)(2F_i(s) - 1) f_i(s)$. In particular, $r_0'(m_0) = r_1'(m_1) = 0$. By Taylor's theorem, we have $r_0(s) = \frac{c}{2} - o(s - m_0)$ and $r_1(s) = \frac{c}{2} - o(m_1 - s)$. Because the BP in question is regular, $\max\left\{\limsup_{b \downarrow 0} \frac{|s^{\pi_B}(b) - m_0|}{b},\ \limsup_{b \uparrow 1} \frac{|s^{\pi_B}(b) - m_1|}{1-b}\right\} < \infty$. Thus, $r_0(s^\pi(b)) - \frac{c}{2} = o\big(s^{\pi_B}(b) - m_0\big) = o(b)$ and $r_1(s^\pi(b)) - \frac{c}{2} = o\big(m_1 - s^{\pi_B}(b)\big) = o(1-b)$. Based on this, and repeating the arguments in Step 2, we deduce that there exists $\mu_2 > 0$ such that $\mathbb{E}_0[\frac{c}{2} - r_0(s^\pi(b_t))] \leq \mu_2 e^{-\lambda t}$. Consequently, $\Delta_0^{\pi_B, \xi_\emptyset}(T) = \frac{cT}{2} - \mathbb{E}_i\big[\sum_{t=1}^T \mathbb{I}\{(X - s_t) d_t < 0\} - (1-c)\mathbb{I}\{(X - s_t) d_t > 0\}\big] = \sum_{t=1}^T \mathbb{E}_0\left[\frac{c}{2} - r_0(s^\pi(b_t))\right] \leq \sum_{t=1}^T \mu_2 e^{-\lambda t} < \sum_{t=1}^\infty \mu_2 e^{-\lambda t} = O(1)$. ■

B.4.2 Discussion on an Equivalent Interpretation of the Absence of the Informed Bettor

The informed bettor's vacuous strategy $\xi_\emptyset$ is equivalent to his best response strategy when the commission rate c is sufficiently high. We formalize this observation in the result below.

Lemma 22. *For every policy π of the market maker, hypothesis $i \in \{0,1\}$, and sufficiently large c that depends only on $F_0(\cdot)$ and $F_1(\cdot)$, $\xi_i^* = \xi_\emptyset$.*

Proof of Lemma 22. Let $\tilde{\alpha} := \min\{F_1(s_L), 1 - F_0(s_H)\}$, which is strictly positive by Assumption 1. Let $c > \frac{1-2\tilde{\alpha}}{1-\tilde{\alpha}}$ so that $\tilde{\alpha} < \frac{1-c}{2-c}$ (this choice of c is feasible because $\frac{1-2\tilde{\alpha}}{1-\tilde{\alpha}} < 1$). For all $s \in \mathcal{S}$,

$$\begin{aligned}
j_1^+(s) =& (c-2)F_1(s) + 1 - c \leq (c-2)\tilde{\alpha} + 1 - c < 0;\\
j_0^+(s) =& (c-2)F_0(s) + 1 - c \leq (c-2)F_1(s) + 1 - c < 0;\\
j_0^-(s) =& (2-c)F_0(s) - 1 \leq (2-c)(1-\tilde{\alpha}) - 1 = (1-c) - \tilde{\alpha}(2-c) < 0;\\
j_1^-(s) =& (2-c)F_1(s) - 1 \leq (2-c)F_0(s) - 1 < 0.
\end{aligned}$$

That is, the informed bettor finds it (strictly) better off not to bet at all, regardless of the market maker's spread line s. In that case, it is easy to verify that the informed bettor's best response strategy is to quit the market, regardless the market maker's policy. ■

B.4.3 Discussion on the Myopic Bayesian Policy (MBP)

In this section, we briefly discuss the MBP. The main purpose of this discussion is to connect to the previous dynamic pricing and learning literature (in particular, Harrison et al., 2012 and Chen and Wang, 2016), and demonstrate that the MBP satisfies the additional regularity condition in Theorem 9. For all $b \in [0,1]$, we denote the market maker's myopic expected profit function by

$$r_b(s) := br_1(s) + (1-b)r_0(s) = (2c-4)\left[b(F_1(s) - \tfrac{1}{2})^2 + (1-b)(F_0(s) - \tfrac{1}{2})^2\right] + \tfrac{c}{2}. \quad \text{(B.4)}$$

The MBP uses the following policy function:

$$s^{\pi_B}(b) := s^\dagger(b) = \sup \arg\max_{s\in\mathbb{R}}\{r_b(s)\}. \tag{B.5}$$

The supremum operator in Equation (B.5) is introduced to ensure that $s^\dagger(b)$ is uniquely defined. In the following series of results, we establish the key properties of the MBP.

Lemma 23. *For $i \in \{0,1\}$, $\arg\max_{s\in\mathbb{R}} r_i(s) = m_i$.*

Proof. Observe that $r_0(s) = (2c-4)\left(F_0(s) - \frac{1}{2}\right)^2 + \frac{c}{2}$ and $r_1(s) = (2c-4)\left(F_1(s) - \frac{1}{2}\right)^2 + \frac{c}{2}$. The statement follows by noticing that $F_0(\cdot)$ and $F_1(\cdot)$ have unique medians m_0 and m_1, respectively, due to (A1:3). ■

Lemma 24. *For $b \in [0,1]$, $s^\dagger(b) \in [m_0, m_1]$. Moreover, for $b \in [0,1]$, $s^\dagger(b)$ strictly increases in b.*

Proof. To prove the first statement, note that for $i \in \{0,1\}$, $r_i(\cdot)$ has a unique maximizer, m_i. Thus, it suffices to consider $b \in (0,1)$. Observe that $r_b'(s) = b(2c-4)f_1(s)\big(2F_1(s)-1\big) + (1-b)(2c-4)f_0(s)\big(2F_0(s)-1\big)$. Hence, for all $s \geq m_1$, $r_b'(s) \leq b(2c-4)f_1(m_1)\big(2F_1(m_1)-1\big) + (1-b)(2c-4)f_0(m_1)\big(2F_0(m_1)-1\big)$. By (A1:3), $F_0(m_1) > F_0(x) > F_0(m_0) = \frac{1}{2} = F_1(m_1) > F_1(x) > F_1(m_0)$ for all $x \in (m_0, m_1)$. As a result, $r_b'(s) \leq 0$ and $r_b(s) \leq s_b(m_1)$ for all $s \geq m_1$. Moreover, $r_b'(m_1) = (1-b)(2c-4)f_0(m_1)\big(2F_0(m_1)-1\big) < 0$, because $b < 1$, $c < 1$, $f_0(m_1) > 0$ by (A1:3), and $2F_0(m_1) - 1 > 0$. Thus, $r_b(s) < s_b(m_1)$ for all $s > m_1$. Consequently, by a similar argument, $r_b(s) < s_b(m_0)$ for all $s < m_0$.

To prove the second statement, we deduce from the first statement that it suffices to consider the case where $s \in [m_0, m_1]$. Note that, for all $b \in [0,1]$ and $s \in [m_0, m_1]$,

$$\begin{aligned}\frac{\partial^2 r_b(s)}{\partial b \partial s} = \frac{\partial}{\partial b}[r_b'(s)] &= \frac{\partial}{\partial b}\big[b(2c-4)f_1(s)(2F_1(s)-1) + (1-b)(2c-4)f_0(s)(2F_0(s)-1)\big] \\ &= \underbrace{(2c-4)}_{<0}\big[\underbrace{f_1(s)(2F_1(s)-1) + f_0(s)(1-2F_0(s))}_{<0}\big] > 0,\end{aligned}$$

where the strict inequality is due to (A1:3). Thus, $r_b(s)$ is a strictly supermodular function of (b,s). Consequently, we deduce from Topkis's theorem (Topkis, 1978) that $s^\dagger(b)$ is non-decreasing in b. To see why $s^\dagger(b)$ is strictly increasing in b, note that $s^\dagger(b)$ satisfies the first

order condition $r_b'(s) = 0$, which is equivalent to

$$b = \frac{f_0(s)(2F_0(s)-1)}{f_0(s)(2F_0(s)-1)+f_1(s)(1-2F_1(s))} =: \mathcal{G}(s). \tag{B.6}$$

For all $s \in [m_0, m_1]$, the denominator of $\mathcal{G}(s)$ is strictly positive and hence well-defined. Suppose that there exist $b_x, b_y \in [0,1]$ such that $s^\dagger(b_x) = s^\dagger(b_y)$. Then, the first order condition states that $b_x = \mathcal{G}(s^\dagger(b_x)) = \mathcal{G}(s^\dagger(b_y)) = b_y$. Hence, $s^\dagger(b)$ must be strictly increasing in b. ■

Proposition 11. *There exist $C_0, C_1 > 0$ such that $s^\dagger(b) = m_0 + C_0 b + o(b) = m_1 + C_1(b-1) + o(1-b)$. That is, $s^\dagger(b)$ is a regular pricing function.*

Proof. Recall that $s^\dagger(\cdot)$ satisfies the first order condition in Equation (B.6). By Lemma 23, m_0 and m_1 are the unique solutions to $\mathcal{G}(s) = 0$ and $\mathcal{G}(s) = 1$, respectively. Moreover, it is straightforward verify that

$$\mathcal{G}'(m_0) = \frac{2f_0^2(m_0)}{f_1(m_0)(1-2F_1(m_0))} > 0 \text{ and } \mathcal{G}'(m_1) = \frac{2f_1^2(m_1)}{f_0(m_1)(2F_0(m_1)-1)} > 0.$$

Note that the strict positivity is guaranteed by (A1:3). Define $C_0 := \frac{1}{\mathcal{G}'(m_0)} > 0$ and $C_1 := \frac{1}{\mathcal{G}'(m_1)} > 0$. By the inverse function theorem (Rudin, 1976, Theorem 9.24), $\mathcal{G}^{-1}(b)$ is uniquely defined in $[0,\varepsilon) \cup (1-\varepsilon, 1]$ for some $\varepsilon > 0$, with $(\mathcal{G}^{-1})'(i) = C_i$ for $i \in \{0,1\}$. Thus, by Taylor's theorem, for $b \in [0,\varepsilon) \cup (1-\varepsilon, 1]$, $s^\dagger(b) = \mathcal{G}^{-1}(b) = m_0 + C_0 b + o(b) = m_1 + C_1(b-1) + o(1-b)$. For $b \in [\varepsilon, 1-\varepsilon]$, $s^\dagger(b)$ is uniquely defined, and the statement of this lemma trivially holds. ■

B.5 Residual Probability Representation of Inertial Policies (Proposition 6)

In this section, we provide the details for the proof of Proposition 6, as well as additional discussions on the extension of the function $\rho(\cdot)$ from $\mathbb{Z}_+$ to $\mathbb{Z}$.

B.5.1 Proof of Proposition 6

We first extend $\rho(\cdot)$ from $\mathbb{Z}_+$ to $\mathbb{Z}$ as follows:

$$\rho(z) = \begin{cases} \frac{1}{2} - F_1\left(\frac{m_0+m_1}{2}\right) & \text{if } z = 0, \\ \frac{1}{2} - F_1 \circ F_0^{-1}\left(\frac{1}{2} + \rho(-z)\right) & \text{if } z \in \mathbb{Z}_-. \end{cases} \tag{B.7}$$

In particular, we find it useful to combine (B.7) with our specific construction of the residual probability sequence $\{\rho(z) = \frac{1}{r_0+rz}, z \in \mathbb{Z}_+\}$, and write out the extended version of $\rho(\cdot)$ as the following:

$$\rho(z) = \begin{cases} \frac{1}{r_0+rz} & \text{if } z \in \mathbb{N}, \\ \frac{1}{2} - F_1 \circ F_0^{-1}\left(\frac{1}{2} + \frac{1}{r_0-rz}\right) & \text{if } z \in \mathbb{Z}_-. \end{cases} \tag{B.8}$$

We complete the rest of proof in three steps.

Step 1. We claim that both $\tilde{s}(\cdot)$ in (3.10) and the extension of $\rho(\cdot)$ in (B.7) are well-defined (we prove this statement to verify that the inverse functions in (3.10) and (B.7) both exist). Because $\rho(z) \in (0, \frac{1}{2} - \alpha)$ for all $z \in \mathbb{Z}_+$, it suffices to verify that (i) $F_0^{-1}(\cdot)$ exists in $(\frac{1}{2}, 1-\alpha)$, and (ii) $F_1^{-1}(\cdot)$ exists in $(\alpha, \frac{1}{2})$. Moreover, as $F_\epsilon(\cdot) = F_i(\cdot + m_i)$ for $i \in \{0, 1\}$, it is sufficient to show that $F_\epsilon^{-1}(\cdot)$ exists in $(\alpha, 1-\alpha)$. Note that $F_\epsilon(m_0 - m_1) = F_1(m_0) = \alpha$, and $F_\epsilon(m_1 - m_0) = F_0(m_1) = 1 - F_1(m_0) = 1 - \alpha$, where the second equality follows from Lemma 16. Since $|m_1 - m_0| = m_1 - m_0 \leq s_H - m_0$, we deduce from (A1:3) that F_ϵ is strictly increasing in the interval $[m_0 - m_1, m_1 - m_0]$. This means that $F_\epsilon^{-1}(\cdot)$ exists and strictly increases in $(\alpha, 1-\alpha)$.

Step 2. We now claim that the construction of $\tilde{s}(\cdot)$ and the extension of $\rho(\cdot)$ satisfy (3.9). First, let $z \in \mathbb{Z}_+$. Then, $F_0\left(\tilde{s}(-z)\right) \overset{(3.10)}{=} F_0\left(F_0^{-1}\left(\frac{1}{2} + \rho(z)\right)\right) = \frac{1}{2} + \rho(z)$. Furthermore, $F_1\big(\tilde{s}(z)\big) \overset{(3.10)}{=} F_1\left(F_1^{-1}\left(\frac{1}{2} - \rho(z)\right)\right) = \frac{1}{2} - \rho(z)$. Now, let $z = 0$. In this case, $F_0(\tilde{s}(0)) \overset{(3.10)}{=} F_0\left(\frac{m_0+m_1}{2}\right) \overset{\text{Lem. } 16}{=} 1 - F_1\left(\frac{m_0+m_1}{2}\right) \overset{(B.7)}{=} \frac{1}{2} + \rho(0)$. Moreover, $F_1(\tilde{s}(0)) \overset{(3.10)}{=} F_1\left(\frac{m_0+m_1}{2}\right) \overset{(B.7)}{=} \frac{1}{2} - \rho(0)$. Finally, let $z \in \mathbb{Z}_-$. To analyze this case, we use the following result, which states that the pricing function $\tilde{s}(\cdot)$ in (3.10) is symmetric around the point $\left(0, \frac{m_0+m_1}{2}\right)$.

Lemma 25. *For all $z \in \mathbb{Z}$, $\tilde{s}(z) + \tilde{s}(-z) = m_0 + m_1$.*

Thus, $F_0(\tilde{s}(-z)) \overset{\text{Lem. } 16 \,\&\, 25}{=} 1 - F_1(\tilde{s}(z)) \overset{(3.10)}{=} 1 - F_1 \circ F_0^{-1}\left(\frac{1}{2} + \rho(-z)\right) \overset{(B.7)}{=} \frac{1}{2} + \rho(z)$. In addition, $F_1\big(\tilde{s}(z)\big) \overset{(3.10)}{=} F_1\left(F_0^{-1}\left(\frac{1}{2} + \rho(-z)\right)\right) \overset{(B.7)}{=} \frac{1}{2} - \rho(z)$.

Step 3. Lastly, we claim that given $\{\rho(z), z \in \mathbb{Z}_+\}$, the construction of $\tilde{s}(\cdot)$ and the extension of $\rho(\cdot)$ that satisfy (3.9) are both unique. The proof of uniqueness also provides us some intuition for the choices of $\tilde{s}(\cdot)$ and $\rho(\cdot)$. Note that both $F_0(\cdot)$ and $F_1(\cdot)$ are strictly increasing by Step 1. Thus, we first uniquely determine the values of $\tilde{s}(\cdot)$ from (3.9). In fact, given the residual probability sequence $\{\rho(z), z \in \mathbb{Z}_+\}$,

- $\{\tilde{s}(z), z \in \mathbb{Z}_+\}$ is uniquely determined by the relationship $F_1\big(\tilde{s}(z)\big) = \frac{1}{2} - \rho(z)$ for all $z \in \mathbb{Z}_+$ (this corresponds to the zone where the betting sequence is in favor of H_1, and hence $\tilde{s}(\cdot)$ is closer to m_1),
- $\{\tilde{s}(z), z \in \mathbb{Z}_-\}$ is uniquely determined by the relationship $F_0\big(\tilde{s}(-z)\big) = \frac{1}{2} - \rho(z)$ for all $z \in \mathbb{Z}_+$ (this corresponds to the zone where the betting sequence is in favor of H_0, and hence $\tilde{s}(\cdot)$ is closer to m_0),
- $\tilde{s}(0)$ is uniquely determined by the relationships $F_1\big(\tilde{s}(0)\big) = \frac{1}{2} - \rho(0)$ and $F_0\big(\tilde{s}(0)\big) = \frac{1}{2} - \rho(0)$, which imply that $F_1\big(\tilde{s}(0)\big) + F_0\big(\tilde{s}(0)\big) = 1$ (this corresponds to the zone where the betting sequence is in favor of neither hypothesis, and hence $\tilde{s}(0) = \frac{m_0+m_1}{2}$).

Because the values of $\tilde{s}(\cdot)$ are uniquely determined, the value of $\rho(z)$ for every $z \in \mathbb{Z}_- \cup \{0\}$ is uniquely determined by (3.9). ■

Proof of Lemma 25. Let $z \in \mathbb{Z}$. If $z = 0$, then we deduce from (3.10) that $\tilde{s}(0) + \tilde{s}(0) = m_0 + m_1$, and the claim holds. On the other hand, if $z \in \mathbb{Z}_+$, then we note that $F_0\left(m_0 + m_1 - F_1^{-1}\left(\frac{1}{2} - \rho(z)\right)\right) \overset{\text{Lem. } 16}{=} 1 - F_1 \circ F_1^{-1}\left(\frac{1}{2} - \rho(z)\right) = \frac{1}{2} + \rho(z)$. In this case, by Step 1 of the proof of Proposition 6, the inverse of F_0 is well-defined. Thus, $F_0^{-1}\left(\frac{1}{2} + \rho(z)\right) = m_0 + m_1 - F_1^{-1}\left(\frac{1}{2} - \rho(z)\right)$. As a result,

$$\tilde{s}(z) + \tilde{s}(-z) \overset{(3.10)}{=} F_1^{-1}\left(\frac{1}{2} - \rho(z)\right) + F_0^{-1}\left(\frac{1}{2} + \rho(z)\right) = m_0 + m_1.$$

■

B.5.2 Discussions

Let us now explore general properties of the extended residual probability sequence $\{\rho(z), z \in \mathbb{Z}\}$ in (B.7).

Lemma 26. *(upper bound)* $\rho(z) < \frac{1}{2} - \alpha$ *for all* $z \in \mathbb{Z}$.

Proof. By definition, $\rho(z) < \frac{1}{2} - \alpha$ for all $z \in \mathbb{Z}_+$. Note that $\rho(0) = \frac{1}{2} - F_1\left(\frac{m_0+m_1}{2}\right) < \frac{1}{2} - F_1(m_0) = \frac{1}{2} - \alpha$ by the (strict) monotonicity of $F_1(\cdot)$. For all $z \in \mathbb{Z}_-$, $\rho(z) = \frac{1}{2} - F_1 \circ F_0^{-1}\left(\frac{1}{2} + \rho(-z)\right) < \frac{1}{2} - F_1 \circ F_0^{-1}\left(\frac{1}{2}\right) = \frac{1}{2} - \alpha$ by the (strict) monotonicity of $F_1(\cdot)$ and $F_0(\cdot)$. ■

Lemma 27. *(lower bound)* $\rho(z) > 0$ *for all* $z \in \mathbb{Z}$. *Moreover, if* $\sup\{\rho(z) : z \in \mathbb{Z}_+\} < \frac{1}{2} - \alpha$, *then* $\inf\{\rho(z) : z \leq M\} > 0$ *for all* $M \in \mathbb{Z}$.

Proof. By definition, $\rho(z) > 0$ for all $z \in \mathbb{Z}_+$. Moreover, $\rho(0) = \frac{1}{2} - F_1\left(\frac{m_0+m_1}{2}\right) > \frac{1}{2} - F_1(m_1) = 0$ by the (strict) monotonicity of $F_1(\cdot)$. Choose $\delta \geq 0$ so that $\rho(z) \leq \frac{1}{2} - \alpha - \delta$ for all $z \in \mathbb{Z}_+$. For all $z \in \mathbb{Z}_-$, $\rho(z) = \frac{1}{2} - F_1 \circ F_0^{-1}\left(\frac{1}{2} + \rho(-z)\right) \geq \frac{1}{2} - F_1 \circ F_0^{-1}(1 - \alpha - \delta) \geq 0$, where the last inequality is strict if $\delta > 0$ by the (strict) monotonicity of $F_1(\cdot)$ and $F_0(\cdot)$. If $\delta > 0$, $\rho(z) \geq \min\{\frac{1}{2} - F_1 \circ F_0^{-1}(1 - \alpha - \delta), \rho(1), \ldots, \rho(M)\} > 0$ for all $z \leq M$. ■

B.6 Key Proof Steps for the Results in Section 3.4

Proof of Lemma 4. Since Y_t is time-homogeneous, we have $f(z) = \mathbb{E}[u(Y_2)|Y_1 = z] - u(z) = \mathbb{E}[u(Y_{t+1})|Y_t = z] - u(z)$ for all z, t. Consequently,

$$\begin{aligned} & f(Y_1) + f(Y_2) + \cdots + f(Y_t) \\ &= \mathbb{E}[u(Y_2|Y_1)] - u(Y_1) + \mathbb{E}[u(Y_3|Y_2)] - u(Y_2) + \cdots + \mathbb{E}[u(Y_{t+1}|Y_t)] - u(Y_t) \\ &= \underbrace{\sum_{i=1}^{t-1} \Big(\mathbb{E}[u(Y_{i+1})|Y_i] - u(Y_{i+1}) \Big)}_{\mathcal{M}_t} + \mathbb{E}[u(Y_{t+1})|Y_t] - u(Y_1) \end{aligned}$$

for all t, where $\mathcal{M}_t$ is a martingale with respect to the σ-algebra $\mathcal{F}_t = \sigma(Y_1, \cdots, Y_t)$ because

$$\begin{aligned} & \mathbb{E}[\mathcal{M}_{t+1}|\mathcal{F}_t] \\ \overset{(a)}{=} & \mathbb{E}\left[\mathcal{M}_t + \mathbb{E}[u(Y_{t+1})|Y_t] - u(Y_{t+1})|\mathcal{F}_t\right] \overset{(b)}{=} \mathcal{M}_t + \mathbb{E}[u(Y_{t+1})|Y_t] - \mathbb{E}[u(Y_{t+1})|\mathcal{F}_t] \overset{(c)}{=} 0. \end{aligned}$$

In the preceding chain of equalities, (a) follows from the definition of $\mathcal{M}_t$, (b) from the fact that both $\mathcal{M}_t$ and $\mathbb{E}[u(Y_{t+1})|Y_t]$ are $\mathcal{F}_t$-measurable, and (c) from the Markov property of the Markov chain Y_t. Thus,

$$\mathbb{E}[f(Y_1)] + \mathbb{E}[f(Y_2)] + \cdots + \mathbb{E}[f(Y_t)] = \mathbb{E}\mathcal{M}_t + \mathbb{E}\mathbb{E}[u(Y_{t+1})|Y_t] - \mathbb{E}u(Y_1) = \mathbb{E}u(Y_{t+1}) - \mathbb{E}u(Y_1).$$

That finishes the proof. ■

The following lemma is another key proof step for the results in Section 3.4.

Lemma 28. *For all constants $\delta \in \mathbb{R}, \hat{z} \in \mathbb{Z}$, and any two sequences $x(\cdot), p(\cdot) : \{\hat{z}, \hat{z}+1, \ldots\} \to \mathbb{R}$, consider the difference equation* (B.9) *below:*

$$\begin{cases} y(\hat{z}-1) = 0, \ y(\hat{z}) = \delta, \\ p(z)y(z+1) + \bar{p}(z)y(z-1) - y(z) = x(z) \ \ \textit{for all } z \geq \hat{z}, \end{cases} \tag{B.9}$$

where $\bar{p}(z) := 1 - p(z)$. If $p(\cdot) \notin \{0, 1\}$, the difference equation (B.9) *above admits the*

following solution $y_\delta^{\hat{z}}(\cdot) : \{\hat{z}-1, \hat{z}, \ldots\} \to \mathbb{R}$*:*

$$y_\delta^{\hat{z}}(z) = \begin{cases} 0 & \text{if } z = \hat{z}-1, \\ \left(1 + \sum_{n=\hat{z}}^{z-1} \prod_{m=\hat{z}}^{n} \frac{\bar{p}(m)}{p(m)}\right)\delta + \sum_{n=\hat{z}}^{z-1} \sum_{k=\hat{z}}^{n} \left(\prod_{m=k}^{n} \frac{\bar{p}(m)}{p(m)}\right) \frac{x(k)}{\bar{p}(k)} & \text{if } z \geq \hat{z}. \end{cases} \tag{B.10}$$

In the notation above, we use the convention that $\sum_{k=n}^{n-1}(\cdot) := 0$*, and* $\prod_{k=n}^{n-1}(\cdot) := 1$*.*

Proof. Let $\delta \in \mathbb{R}$, $\hat{z} \in \mathbb{Z}$, and $x(\cdot)$ be a function from $\{\hat{z}, \hat{z}+1, \ldots\}$ to $\mathbb{R}$. By construction, $y_\delta^{\hat{z}}(\cdot)$ satisfies the boundary conditions $y_\delta^{\hat{z}}(\hat{z}-1) = 0$ and $y_\delta^{\hat{z}}(\hat{z}) = \delta$. To verify the inductive relation, we first evaluate the term $y_\delta^{\hat{z}}(z+1) - y_\delta^{\hat{z}}(z)$ for $z \geq \hat{z}-1$:

$$y_\delta^{\hat{z}}(z+1) - y_\delta^{\hat{z}}(z) = \begin{cases} \delta & \text{if } z = \hat{z}-1, \\ \delta \prod_{m=\hat{z}}^{z} \frac{\bar{p}(m)}{p(m)} + \sum_{k=\hat{z}}^{z} \left(\prod_{m=k}^{z} \frac{\bar{p}(m)}{p(m)}\right) \frac{x(k)}{\bar{p}(k)} & \text{if } z \geq \hat{z}. \end{cases}$$

Next, we evaluate the term $p(z)y_\delta^{\hat{z}}(z+1) + \bar{p}(z)y_\delta^{\hat{z}}(z-1) - y_\delta^{\hat{z}}(z)$: for all $z \geq \hat{z}$,

$$\begin{aligned} & p(z)y_\delta^{\hat{z}}(z+1) + \bar{p}(z)y_\delta^{\hat{z}}(z-1) - y_\delta^{\hat{z}}(z) \\ &= p(z)[y_\delta^{\hat{z}}(z+1) - y_\delta^{\hat{z}}(z)] - \bar{p}(z)[y_\delta^{\hat{z}}(z) - y_\delta^{\hat{z}}(z-1)] \\ &= \delta \left(\prod_{m=\hat{z}}^{z-1} \frac{\bar{p}(m)}{p(m)}\right) \left(p(z)\frac{\bar{p}(z)}{p(z)} - \bar{p}(z)\right) + p(m)\frac{\bar{p}(m)}{p(m)}\frac{x(z)}{\bar{p}(z)} \\ &+ \left[\sum_{k=\hat{z}}^{z-1} \left(\prod_{m=k}^{z-1} \frac{\bar{p}(m)}{p(m)}\right) \frac{x(k)}{\bar{p}(k)}\right] \left(p(z)\frac{\bar{p}(z)}{p(z)} - \bar{p}(z)\right) \\ &= 0 + x(z) + 0 = x(z). \quad \blacksquare \end{aligned}$$

B.7 The Informed Bettor's Best Response to IP (Theorem 10)

In this section, we provide the details for the proof of Theorem 10.

B.7.1 Summary of Intuition

Lemmas 1 and 3 correspond to two separate mechanisms through which the informed bettor's profit may be unbounded. The first mechanism is that the threshold strategy ξ_i^* itself

generates an infinite amount of profit for the informed bettor. In the context of the threshold strategy ξ_i^* defined in (3.11), this happens when the market state Z_t behaves so noisily that the event of severe mispricing (i.e., the spread line being sufficiently far away from the true median) occurs infinitely often. In such cases, the market maker is effectively not collecting information from the market. Our Myopic Tracking policy guards against this mechanism by preventing the learning rate (formally defined as the drift of $\{Z_t\}$) from vanishing too fast, in which case $\{Z_t\}$ diverges in the right direction and the spread line s_t converges to the correct median almost surely.

The second mechanism is that the informed bettor may have an incentive to deviate from ξ_i^*. To mathematically verify that IP guards against this mechanism, it suffices to only algebraically verify the Bellman equation (3.18). But to see it intuitively, let us separately discuss the following two cases, each corresponding to a different type of deviation:

- (*case 1*) The informed bettor may deviate from ξ_i^* by bluffing. IP prevents this type of deviation because under IP, a pair of positive-negative bets have no net impact on the market state Z_t. Thus, the informed bettor can only push the market state Z_t to the wrong direction by bluffing more often than honest betting. But as discussed in Section 3.3.2, he also needs to bet honestly more often than bluffing to gain a positive net profit. Based on these two contradicting facts, we reach the conclusion that the informed bettor does not have an incentive to bluff.

- (*case 2*) The informed bettor may also deviate from ξ_i^* by not following the threshold structure of betting honestly versus waiting. IP induces the informed bettor to bet according to a threshold strategy, because we can marginally change the action from betting ($a_t = +1$) to waiting ($a_t = 0$) at every state $z \in \mathbb{Z}$, and evaluate the difference in his continuation profits. It turns out this difference function only crosses zero once, leading to a threshold structure.[1]

1. Algebraically, this is ultimately reduced to verifying that the function $j_1^+(z-1)\left[\left(\frac{1}{2}-\rho(z)\right)+\left(\frac{1}{2}-\rho(z)\right)^2+\cdots\right]-j_1^+(z)=j_1^+(z-1)\,\frac{\frac{1}{2}-\rho(z)}{\frac{1}{2}+\rho(z)}-j_1^+(z)$ crosses zero only once.

B.7.2 Preliminaries

First, we provide the explicit expressions for $\bar{z}$ and $\bar{r}$. Let

$$\bar{z} := \inf\left\{z : \frac{\rho(z-1)}{2\rho(z)} - \rho(z) - \rho(z-1) > \frac{1}{2} - \frac{c}{2-c}\right\} \tag{B.11}$$

be the threshold used in the informed bettor's optimal policy ξ_i^* in (3.11). We follow the common convention that $\bar{z} = \infty$ if the set inside the infimum in (B.11) is empty, and $\bar{z} = -\infty$ if this set is unbounded from below. By Lemma 30 below, $\bar{z}$ is finite if and only if $j_1^+(-\infty) > 0$ (i.e., the type-1 informed bettor finds it profitable to act at some point in time). Otherwise, $\bar{z} = -\infty$, which corresponds to the informed bettor's policy of never betting according to (3.11). Given strictly positive constants $\bar{r}_0, \bar{r}_1$ that depend only on $m_0, m_1, F_\epsilon(\cdot)$, and c, let

$$\bar{r} = \min\{2, \bar{r}_0, \bar{r}_1\} \tag{B.12}$$

be the upper bound of r for Theorem 10 to hold. The closed form expressions for $\bar{r}_0$ and $\bar{r}_1$ are as follows:

$$\bar{r}_0 := \frac{cr_0}{(2-c)\zeta_0}, \quad \text{where } \zeta_0 := \max\left\{1, \max_{s\in[m_0,m_1]} \frac{f_1(s)}{f_0(s)}\right\}, \tag{B.13}$$

and

$$\bar{r}_1 := \sup\{r : \zeta_1 r + 2r_0 > 0\} = \begin{cases} \frac{2r_0}{-\zeta_1} & \text{if } \zeta_1 < 0, \\ \infty & \text{if } \zeta_1 \geq 0, \end{cases} \quad \text{where } \zeta_1 := \min_{s\in[m_0,m_1]} \left\{\left(\frac{f_1'(s)}{f_1(s)} - \frac{f_0'(s)}{f_0(s)}\right)\frac{1}{f_0(s)}\right\}. \tag{B.14}$$

Remark 4. *Theorem 10 requires $\bar{r}$ to be strictly positive. It is straightforward to verify the strict positivity of $\bar{r}$. By (A1:1) and (A1:3), $f_i(\cdot)$ is continuous and strictly positive in the interval $[m_0, m_1]$. Hence, ζ_0 is strictly positive and finite (which implies that $\bar{r}_0 > 0$), and ζ_1 is finite (which implies that $\bar{r}_1 > 0$).*

B.7.3 Auxiliary Lemmas

We employ the following auxiliary lemmas to prove Theorem 10, deferring their proofs to Appendix B.7.6. The first auxiliary lemma summarizes the properties of $\rho(\cdot)$ in (B.8).

Lemma 29. *The (extended) residual probability sequence* $\{\rho(z), z \in \mathbb{Z}\}$ *in* (B.8) *satisfies the following:*

(L29-1) The natural further extension of $\rho(z)$ *from domain* $\mathbb{Z}$ *to domain* $\mathbb{R}$*, defined as*

$$\rho(x) = \begin{cases} \frac{1}{r_0 + rx} & \text{if } x \geq 0, \\ \frac{1}{2} - F_1 \circ F_0^{-1}(\frac{1}{2} + \frac{1}{r_0 - rx}) & \text{if } x < 0, \end{cases} \tag{B.15}$$

is a continuous and strictly decreasing function. Moreover, $\rho(\cdot)$ *is twice differentiable in* $\mathbb{R} \setminus \{0\}$*.*

(L29-2) For all $z \in \mathbb{Z}$*,* $\frac{1}{2} - \alpha = \rho(-\infty) > \rho(z) > \rho(\infty) = 0$*.*

(L29-3) For all $r \in (0, 4)$ *and* $z_0 \in \mathbb{Z}$*,* $\sum_{n=0}^{\infty} \prod_{k=0}^{n} \frac{\frac{1}{2} - \rho(z_0 + k)}{\frac{1}{2} + \rho(z_0 + k)} < \infty$*.*

(L29-4) Suppose that $0 < r < \bar{r}_0$*. Then, for every integer* $z \in \mathbb{Z}$*,* $\frac{\rho(z+1)}{\rho(z)} > 1 - \frac{c}{2-c}$*. Therefore,*

(a) $\rho(z) - \rho(z+1) < \frac{c}{2-c}$*,*

(b) $\frac{\frac{1}{2} - \rho(z)}{\frac{1}{2} + \rho(z)} < \frac{(2-c)\rho(z+1) + \frac{c}{2}}{(2-c)\rho(z) + \frac{c}{2}}$*.*

In Lemma 29 above, (L29-1)-(L29-2) are (intuitive) regularity conditions for $\rho(\cdot)$. Property (L29-3) is closely related to the convergence of $\sum_{n=1}^{\infty} \Lambda_n$, which ensures that Z_t diverges to infinity with probability one. The last property, (L29-4), ensures that $\rho(z+1)$ is "close enough" to $\rho(z)$. This "inertia" property eliminates the informed bettor's incentive to bluff; see the proof of Lemma 3 for further details. The following lemma summarizes the properties of the threshold $\bar{z}$, which is defined in (B.11).

Lemma 30. *Suppose that* $0 < r < \min\{2, \bar{r}_1\}$*. Then,* $\bar{z}$ *possesses the following properties:*

(L30-1) (finiteness) If $j_1^+(-\infty) > 0$*, then* $\bar{z}$ *is finite. Otherwise,* $\bar{z} = -\infty$*.*

(L30-2) (single-crossing property) $j_1^+(z)\left[\frac{1}{2} + \rho(z)\right] < j_1^+(z-1)\left[\frac{1}{2} - \rho(z)\right]$ *if and only if* $z \geq \bar{z}$*.*

(L30-3) (profitable action) $j_1^+(z) > 0$ *for all* $z < \bar{z}$.

The following lemma characterizes the summation of the probabilities that Z_t hits the region $(-\infty, M]$ up to period T under the probability measure $\mathbb{P}_1^z$.

Lemma 31. *For all* $r \in (0, \bar{r}), \bar{z} \in \mathbb{Z} \cup \{-\infty\}$, *and* $M \in \mathbb{Z}$ *satisfying* $M > \bar{z} - 2$, *there exists an increasing function* $\tilde{u} : \{z \in \mathbb{Z} : z > \bar{z} - 2\} \to \mathbb{R}$ *such that*

$$\sum_{t=1}^{T} \mathbb{E}_1^z \mathbb{I}\{Z_t \leq M\} = \mathbb{E}_1^z \tilde{u}(Z_{T+1}) - \tilde{u}(z) \quad \textit{for all } z \in \mathbb{Z} \textit{ satisfying } z > \bar{z} - 2 \textit{ and } T \in \mathbb{Z}_+. \tag{B.16}$$

The closed-form expression for $\tilde{u}(\cdot)$ *is as follows:*

$$\tilde{u}(z) = \begin{cases} \left(1 + \sum\limits_{n=z+1}^{M} \prod\limits_{m=n}^{M} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)}\right)\tilde{\beta} + \sum\limits_{n=z+1}^{M} \sum\limits_{k=n}^{M} \frac{1}{\frac{1}{2}+\rho(k)} \prod\limits_{m=n}^{k} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} & \textit{if } \bar{z} - 2 < z \leq M, \\ 0 & \textit{if } z = M+1, \\ \left(1 + \sum\limits_{n=M+2}^{z-1} \prod\limits_{m=M+2}^{n} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\right)\beta & \textit{if } z \geq M+2, \end{cases} \tag{B.17}$$

where $\beta > 0$ *and* $\tilde{\beta} < 0$ *are finite constants given by:*

$$\begin{cases} \tilde{\beta} = -\prod_{m=\bar{z}}^{M} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} - \sum_{k=\bar{z}}^{M} \frac{1}{\frac{1}{2}-\rho(k)} \prod_{m=k}^{M} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}, \\ \beta = -\frac{\frac{1}{2}-\rho(M+1)}{\frac{1}{2}+\rho(M+1)}\tilde{\beta}. \end{cases} \tag{B.18}$$

B.7.4 Main Body of the Proof of Theorem 10

For the market maker, fix an arbitrary Myopic Tracking policy π_I with tuning parameter $r \in (0, \bar{r})$, where $\bar{r}$ is as in (B.12). We make two assumptions without loss of generality. First, we assume that $i = 1$, because the analysis for the case where $i = 0$ can be repeated verbatim. Second, we assume that $j_1^+(-\infty) > 0$ (which implies that $\bar{z} > -\infty$ by (L30-1)) because otherwise the type-1 informed bettor never finds it profitable to act and the theorem's statement holds trivially.

Given the function $\bar{J}^1(\cdot)$ defined in (3.15), we have the following for all $T \in \mathbb{Z}_+$ and any admissible response policy ξ_1 of type-1 informed bettor:

$$
\begin{aligned}
&0 \\
&\leq \mathbb{E}_1^{\pi_I,\xi_1}[\bar{J}^1(Z_{T+1})] && \text{[by Lemma 1]} \\
&= \mathbb{E}_1^{\pi_I,\xi_1}[\bar{J}^1(Z_1)] + \sum_{t=1}^{T}\Big[\mathbb{E}_1^{\pi_I,\xi_1}[\bar{J}^1(Z_{t+1})] - \mathbb{E}_1^{\pi_I,\xi_1}[\bar{J}^1(Z_t)]\Big] \\
&= \bar{J}^1(0) \\
&+ \mathbb{E}_1^{\pi_I,\xi_1}\sum_{t=1}^{T}\Bigg[\mathbb{I}\{a_t=+1\}\big(\underbrace{\bar{J}^1(Z_t+1)-\bar{J}^1(Z_t)}_{\overset{\text{Lem. 3}}{\leq} -j_1^+(Z_t)}\big) \\
&\qquad + \mathbb{I}\{a_t=-1\}\big(\underbrace{\bar{J}^1(Z_t-1)-\bar{J}^1(Z_t)}_{\overset{\text{Lem. 3}}{\leq} -j_1^-(Z_t)}\big) \\
&\qquad + \mathbb{I}\{a_t=0\}\Big(\underbrace{\Big[\tfrac{1}{2}+\rho(Z_t)\Big]\bar{J}^1(Z_t+1)+\Big[\tfrac{1}{2}-\rho(Z_t)\Big]\bar{J}^1(Z_t-1)-\bar{J}^1(Z_t)}_{\overset{\text{Lem. 3}}{\leq} 0}\Big)\Bigg] \\
&\leq \bar{J}^1(0) + \mathbb{E}_1^{\pi_I,\xi_1}\sum_{t=1}^{T}\Big[\mathbb{I}\{a_t=+1\}[-j_1^+(Z_t)] + \mathbb{I}\{a_t=-1\}[-j_1^-(Z_t)]\Big] \\
&= \bar{J}^1(0) - V_1^{\pi_I,\xi_1}(T). && \text{[by (3.2)]}
\end{aligned}
$$

As a result, under the strategy profile (π_I, ξ_1), the informed bettor's continuation profit function is bounded above by the constant $\bar{J}^1(0)$, which is independent of T. That is, $V_1^{\pi_I,\xi_1}(T) \leq \bar{J}^1(0)$. Taking the limit infimum over T on the term on both sides and invoking Lemma 2, we reach the following chain of relations:

$$
\liminf_{T\to\infty} V_1^{\pi_I,\xi_1}(T) \leq \bar{J}^1(0) \overset{\text{Lem. 2}}{=} \lim_{T\to\infty} \mathbb{E}_1^0\left[\sum_{t=1}^{T} j_1^+(Z_t)\,\mathbb{I}\{Z_t \leq \bar{z}-1\}\right] \overset{(3.2)}{=} \lim_{T\to\infty} V_1^{\pi_I,\xi_1^*}(T).
$$

Thus, $\xi_1^* \in \arg\max_\xi \liminf_{T\to\infty} V_i^{\pi_I,\xi}(T)$. By Lemma 1, we deduce that the informed bettor's total profit is finite: $\sup_\xi \liminf_{T\to\infty} V_i^{\pi_I,\xi}(T) = \lim_{T\to\infty} V_1^{\pi_I,\xi_1^*}(T) = \bar{J}^1(0) \overset{\text{Lem. 1}}{<} \infty$. Consequently, ξ_1^* is the type-1 informed bettor's best response strategy in the sense of

(3.3). Moreover, $\bar{J}^1(\cdot) = J^1(\cdot)$ is the optimal value function. ■

B.7.5 Proofs of Lemma 1-3

Proof of Lemma 1. Recall from (B.12) that $\bar{r} \leq 2 < 4$. Hence, when $r \in (0, \bar{r})$,

$$\sum_{n=0}^{\infty} \Lambda_n \overset{(3.16)}{=} \sum_{n=0}^{\infty} \prod_{k=0}^{n} \frac{\frac{1}{2} - \rho(\bar{z}+k)}{\frac{1}{2} + \rho(\bar{z}+k)} \overset{\text{(L29-3)}}{<} \infty.$$

Recalling (3.15), we note that the finiteness and nonnegativity of $\bar{J}^i(\cdot)$ follows from the convergence of $\sum_{n=0}^{\infty} \Lambda_n$ as well as the nonnegativity of $j_1^+(z)$ for all $z < \bar{z}$ because of (L30-3). ■

Proof of Lemma 2. To verify that (3.17) holds, let us assume that $\bar{z} > -\infty$ without loss of generality. Otherwise, both sides of the equation are zero trivially. We complete the proof in three steps.

Step 1. We claim that if $\bar{J}^1(\cdot)$ satisfies (3.17), where the limit term on the right-hand side exists, then $\bar{J}^0(\cdot)$ also satisfies (3.17), and the limit term on the right-hand side exists. In other words, we may assume that $i = 1$ without loss of generality. To prove this claim, note that for all $z, \breve{z} \in \mathbb{Z}$, $\mathcal{P}^1_{z,\breve{z}} = \mathcal{P}^0_{-z,-\breve{z}}$. Therefore, $\{Z_t\}$ under $\mathbb{P}_1^z$ has the same law as $\{-Z_t\}$ under $\mathbb{P}_0^{-z}$. Formally speaking, given any $T \in \mathbb{Z}_+$, measurable function $f : \mathbb{R}^T \to \mathbb{R}$, and $z \in \mathbb{Z}$, we have

$$\mathbb{E}_1^z f(Z_1, Z_2, \ldots, Z_T) = \mathbb{E}_0^{-z} f(-Z_1, -Z_2, \ldots, -Z_T). \tag{B.19}$$

As a result, if $\bar{J}^1(\cdot)$ satisfies (3.17), we have

$$\begin{aligned}
&\bar{J}^0(z) \\
&= \bar{J}^1(-z) && \text{[by (3.15)]} \\
&= \lim_{T\to\infty} \mathbb{E}_1^{-z}\left[\sum_{t=1}^{T} j_1^+(Z_t)\,\mathbb{I}\{Z_t \le \bar{z}-1\}\right] && \text{[(3.17); the lim. exists by assumption]} \\
&= \lim_{T\to\infty} \mathbb{E}_0^{z}\left[\sum_{t=1}^{T} j_1^+(-Z_t)\,\mathbb{I}\{-Z_t \le \bar{z}-1\}\right] && \text{[by (B.19)]} \\
&= \lim_{T\to\infty} \mathbb{E}_0^{z}\left[\sum_{t=1}^{T} j_0^-(Z_t)\,\mathbb{I}\{Z_t \ge 1-\bar{z}\}\right] && \text{[by (3.14)]}
\end{aligned}$$

Step 2. We claim that for all $z \in \mathbb{Z}$, there exists a bounded function $u_0(\cdot)$ such that for all $T \ge \bar{z} - z$,

$$\begin{aligned}
&\mathbb{E}_1^z \sum_{t=1}^{T} j_1^+(Z_t)\,\mathbb{I}\{Z_t \le \bar{z}-1\} \\
&= \begin{cases} j_1^+(\bar{z}-1)\left[\mathbb{E}_1^z u_0(Z_{T+1}) - u_0(z)\right] & \text{if } z \ge \bar{z}, \\ \sum_{i=1}^{\bar{z}-z} j_1^+(\bar{z}-i) + j_1^+(\bar{z}-1)\mathbb{E}_1^{\bar{z}} u_0(Z_{T+1-\bar{z}+z}) & \text{if } z \le \bar{z}-1. \end{cases}
\end{aligned} \tag{B.20}$$

Here, $u_0(\cdot)$ is as follows:

$$u_0(z) = \begin{cases} -1 & \text{if } z = \bar{z}-1, \\ 0 & \text{if } z = \bar{z}, \\ \sum_{n=0}^{z-\bar{z}-1} \Lambda_n & \text{if } z \ge \bar{z}+1. \end{cases} \tag{B.21}$$

Let us first consider the case where $z \ge \bar{z}$. Starting with initial value z, the Markov chain Z_t is restrained to the region $[\bar{z}-1, \infty)$. Consider a special case of Lemma 31 where $\bar{z} > -\infty$

and $M = \bar{z} - 1$. In that special case, the corresponding $\tilde{u}(\cdot)$ function simplifies to:

$$\tilde{u}(z) = \begin{cases} -1 & \text{if } z = \bar{z} - 1, \\ 0 & \text{if } z = \bar{z}, \\ \left(1 + \sum_{n=\bar{z}+1}^{z-1} \prod_{m=\bar{z}+1}^{n} \frac{\frac{1}{2} - \rho(m)}{\frac{1}{2} + \rho(m)}\right) \frac{\frac{1}{2} - \rho(\bar{z})}{\frac{1}{2} + \rho(\bar{z})} & \text{if } z \geq \bar{z} + 1, \end{cases} \qquad \text{[by (B.17)]}$$

$$= \begin{cases} -1 & \text{if } z = \bar{z} - 1, \\ 0 & \text{if } z = \bar{z}, \\ \sum_{n=\bar{z}}^{z-1} \prod_{m=\bar{z}}^{n} \frac{\frac{1}{2} - \rho(m)}{\frac{1}{2} + \rho(m)} & \text{if } z \geq \bar{z} + 1, \end{cases}$$

$$= \begin{cases} -1 & \text{if } z = \bar{z} - 1, \\ 0 & \text{if } z = \bar{z}, \\ \sum_{n=0}^{z-\bar{z}-1} \Lambda_n & \text{if } z \geq \bar{z} + 1, \end{cases} \qquad \text{[by (3.16)]}$$

$$= u_0(z). \qquad \text{[by (B.21)]}$$

Thus, $u_0(\cdot)$ is a special case of the function $\tilde{u}(\cdot)$ in Lemma 31. By the conclusion of Lemma 31, $u_0(\cdot)$ is an increasing function such that $\sum_{t=1}^{T} \mathbb{E}_1^z \mathbb{I}\{Z_t \leq \bar{z} - 1\} = \mathbb{E}_1^z \tilde{u}(Z_{T+1}) - \tilde{u}(z) = \mathbb{E}_1^z u_0(Z_{T+1}) - u_0(z)$ for all $z \geq \bar{z} - 1$. Moreover, since $\sum_n \Lambda_n$ is a convergent series (see the derivations in the proof of Lemma 1), the function $u_0(\cdot)$ is bounded. Finally, $\mathbb{E}_1^z \sum_{t=1}^{T} j_1^+(Z_t)\, \mathbb{I}\{Z_t \leq \bar{z} - 1\} = \mathbb{E}_1^z \sum_{t=1}^{T} j_1^+(\bar{z} - 1)\, \mathbb{I}\{Z_t \leq \bar{z} - 1\}$
$= j_1^+(\bar{z} - 1) \sum_{t=1}^{T} \mathbb{E}_1^z \mathbb{I}\{Z_t \leq \bar{z} - 1\} = j_1^+(\bar{z} - 1)[\mathbb{E}_1^z u_0(Z_{T+1}) - u_0(z)]$.

Now, let us consider the case where $z < \bar{z}$. Note that Z_t increases with certainty until it

hits $[\bar{z}, \infty)$. Thus,

$$
\begin{aligned}
&\mathbb{E}_1^z \sum_{t=1}^{T} j_1^+(Z_t)\, \mathbb{I}\{Z_t \le \bar{z}-1\} \\
&= \mathbb{E}_1^z \sum_{t=1}^{\bar{z}-z} j_1^+(Z_t)\, \mathbb{I}\{Z_t \le \bar{z}-1\} \\
&\quad + \mathbb{E}_1^z \sum_{t=\bar{z}-z+1}^{T} j_1^+(Z_t)\, \mathbb{I}\{Z_t \le \bar{z}-1\} \\
&= \sum_{i=1}^{\bar{z}-z} j_1^+(\bar{z}-i) + \mathbb{E}_1^z \sum_{t=\bar{z}-z+1}^{T} j_1^+(\bar{z}-1)\, \mathbb{I}\{Z_t \le \bar{z}-1\} \\
&= \sum_{i=1}^{\bar{z}-z} j_1^+(\bar{z}-i) \\
&\quad + \mathbb{E}_1^{\bar{z}} \sum_{t=1}^{T-\bar{z}+z} j_1^+(\bar{z}-1)\, \mathbb{I}\{Z_t \le \bar{z}-1\} && \text{[by the Markov property of } Z_t\text{]} \\
&\overset{(a)}{=} \sum_{i=1}^{\bar{z}-z} j_1^+(\bar{z}-i) + \mathbb{E}_1^{\bar{z}} \sum_{t=1}^{T-\bar{z}+z} j_1^+(Z_t)\, \mathbb{I}\{Z_t \le \bar{z}-1\} \\
&= \sum_{i=1}^{\bar{z}-z} j_1^+(\bar{z}-i) + j_1^+(\bar{z}-1)\mathbb{E}_1^{\bar{z}} u_0(Z_{T-\bar{z}+z+1}). && \text{[by the case for } z \ge \bar{z};\ u_0(\bar{z}) = 0\text{]}
\end{aligned}
$$

In the derivations above, (a) follows since Z_t increases by one with certainty as soon as it hits $(-\infty, \bar{z}-1]$. Combining our results for the cases where $z \ge \bar{z}$ and $z < \bar{z}$, we finish this step.

Step 3. We evaluate the term $\bar{J}^1(z)$ for all $z \in \mathbb{Z}$:

$$
\begin{aligned}
&\bar{J}^1(z) \\
&= \begin{cases} j_1^+(\bar{z}-1)\sum_{n=z-\bar{z}}^{\infty}\Lambda_n & \text{if } \bar{z} \le z, \\ \sum_{i=1}^{\bar{z}-z} j_1^+(\bar{z}-i) + j_1^+(\bar{z}-1)\sum_{n=0}^{\infty}\Lambda_n & \text{if } z \le \bar{z}-1, \end{cases} && \text{[by (3.15)]} \\
&= \begin{cases} j_1^+(\bar{z}-1)\left[u_0(\infty)-u_0(z)\right] & \text{if } \bar{z} \le z, \\ \sum_{i=1}^{\bar{z}-z} j_1^+(\bar{z}-i) + j_1^+(\bar{z}-1)u_0(\infty) & \text{if } z \le \bar{z}-1, \end{cases} && \text{[by (B.21)]} \\
&\overset{(b)}{=} \lim_{T\to\infty} \begin{cases} j_1^+(\bar{z}-1)\left[\mathbb{E}_1^z u_0(Z_{T+1}) - u_0(z)\right] & \text{if } \bar{z} \le z, \\ \sum_{i=1}^{\bar{z}-z} j_1^+(\bar{z}-i) + j_1^+(\bar{z}-1)\mathbb{E}_1^{\bar{z}} u_0(Z_{T+1-\bar{z}+z}) & \text{if } z \le \bar{z}-1, \end{cases} \\
&= \lim_{T\to\infty} \mathbb{E}_1^z \sum_{t=1}^{T} j_1^+(Z_t)\,\mathbb{I}\{Z_t \le \bar{z}-1\}, && \text{[by (B.20)]}
\end{aligned}
$$

where (b) follows from the bounded convergence theorem (Rudin, 1976, Theorem 11.32) and the following two facts: (i) $Z_t \uparrow \infty$ almost surely (by Statement (T11:1) in Theorem 11, which is proven independently), and (ii) $u_0(\cdot)$ in (B.21) is a bounded function (see Step 2). As a result, $\lim_{T\to\infty} \mathbb{E}_1^z \sum_{t=1}^{T} j_1^+(Z_t)\,\mathbb{I}\{Z_t \le \bar{z}-1\}$ exists and equals $\bar{J}^1(z)$. Combining Steps 1-3, we complete the proof. ■

Proof of Lemma 3. We first assume that $\bar{z} > -\infty$ without loss of generality. To see this, suppose that $\bar{z} = -\infty$. By (L30-1), $j^+(-\infty) \le 0$. As a result, we have $j_1^+(z) \le 0$ and $j_1^-(z) \le 0$ for all $z \in \mathbb{Z}$, and the Bellman equation (3.18) holds trivially. We complete the rest of the proof in four steps.

Step 1. We claim that we can assume that $i = 1$ without loss of generality. To be more accurate, $\bar{J}^0(\cdot)$ satisfies the Bellman equation (3.18) if $\bar{J}^1(\cdot)$ does. Note that if $\bar{J}^1(\cdot)$ satisfies

(3.18), then we have the following for all $z \in \mathbb{Z}$:

$$
\begin{aligned}
&\bar{J}^0(z) \\
&= \bar{J}^1(-z) && \text{[by (3.15)]} \\
&= \max\Big\{ \underbrace{j_1^+(-z)}_{\overset{(a)}{=} j_0^-(z)} + \underbrace{\bar{J}^1(-z+1)}_{\overset{(b)}{=} \bar{J}^0(z-1)}, \quad \underbrace{j_1^-(-z)}_{\overset{(a)}{=} j_0^+(z)} + \underbrace{\bar{J}^1(-z-1)}_{\overset{(b)}{=} \bar{J}^0(z+1)}, \\
&\qquad \underbrace{\bar{F}_1(\tilde{s}(-z))}_{\overset{(c)}{=} F_0(\tilde{s}(z))} \underbrace{\bar{J}^1(-z+1)}_{\overset{(b)}{=} \bar{J}^0(z-1)} + \underbrace{F_1(\tilde{s}(-z))}_{\overset{(c)}{=} \bar{F}_0(\tilde{s}(z))} \underbrace{\bar{J}^1(-z-1)}_{\overset{(b)}{=} \bar{J}^0(z+1)} \Big\} && \text{[by (3.18)]} \\
&= \max\left\{ j_0^-(z) + \bar{J}^0(z-1), j_0^+(z) + \bar{J}^0(z+1), F_0(\tilde{s}(z))\bar{J}^0(z-1) + \bar{F}_0(\tilde{s}(z))\bar{J}^0(z+1) \right\}
\end{aligned}
$$

where (a), (b), and (c) follow from (3.14), (3.15), and (3.9) respectively. Therefore, $\bar{J}^0(\cdot)$ satisfies the Bellman equation (3.18) as well.

Step 2. We claim that the following equation holds:

$$
\bar{J}^1(z) = \begin{cases} j_1^+(z) + \bar{J}^1(z+1) & \text{if } z \le \bar{z} - 1, \\ \bar{F}_1(\tilde{s}(z))\bar{J}^1(z+1) + F_1(\tilde{s}(z))\bar{J}^1(z-1) & \text{if } z \ge \bar{z}. \end{cases} \tag{B.22}
$$

The intuition for (B.22) is that the function $\bar{J}^1(\cdot)$ satisfies the above recursive relation if the type-1 bettor follows the threshold strategy ξ_1^*. This step is a direct consequence of Lemma 2, the transition rule of Z_t, and the Markov property of Z_t.

Step 3. We build on Step 2, and claim that for all $z \in \mathbb{Z}$, the following holds:

$$
\bar{J}^1(z) = \max\left\{ j_1^+(z) + \bar{J}^1(z+1), \bar{F}_1(\tilde{s}(z))\bar{J}^1(z+1) + F_1(\tilde{s}(z))\bar{J}^1(z-1) \right\}. \tag{B.23}
$$

Equation (B.23) can be interpreted as a "weakened" version of the Bellman equation for the decision problem where bluffing (or $a_t = -1$) is not a feasible action for the type-1 bettor. Invoking (B.22), we deduce that it suffices to verify that

$$
\underbrace{j_1^+(z) + \bar{J}^1(z+1) - \left[\bar{F}_1(\tilde{s}(z))\bar{J}^1(z+1) + F_1(\tilde{s}(z))\bar{J}^1(z-1) \right]}_{(*)} \ge 0 \iff z \le \bar{z} - 1.
$$

Let us evaluate the term $(*)$ above. For all $z \in \mathbb{Z}$,

$$
\begin{aligned}
&(*)\\
&= j_1^+(z) + \bar{J}^1(z+1) - \left[\bar{F}_1\left(\tilde{s}(z)\right)\bar{J}^1(z+1) + F_1\left(\tilde{s}(z)\right)\bar{J}^1(z-1)\right]\\
&= j_1^+(z) + \bar{J}^1(z+1) - \left[\tfrac{1}{2} + \rho(z)\right]\bar{J}^1(z+1) - \left[\tfrac{1}{2} - \rho(z)\right]\bar{J}^1(z-1)\\
&= j_1^+(z) + \left[\tfrac{1}{2} - \rho(z)\right]\left[\bar{J}^1(z+1) - \bar{J}^1(z-1)\right]\\
&= \begin{cases}
j_1^+(z) + \\
\left[\tfrac{1}{2} - \rho(z)\right]\left[j_1^+(\bar{z}-1)\sum\limits_{n=z+1-\bar{z}}^{\infty}\Lambda_n - j_1^+(\bar{z}-1)\sum\limits_{n=z-1-\bar{z}}^{\infty}\Lambda_n\right] & \text{if } z \geq \bar{z}+1,\\
j_1^+(z) + \\
\left[\tfrac{1}{2} - \rho(z)\right]\left[j_1^+(\bar{z}-1)\sum\limits_{n=1}^{\infty}\Lambda_n - j_1^+(\bar{z}-1)\sum\limits_{n=0}^{\infty}\Lambda_n - j_1^+(\bar{z}-1)\right] & \text{if } z = \bar{z},\\
j_1^+(z) + \\
\left[\tfrac{1}{2} - \rho(z)\right]\left[\sum\limits_{i=1}^{\bar{z}-z-1} j_1^+(z-i) - \sum\limits_{i=1}^{\bar{z}-z+1} j_1^+(z-i)\right] & \text{if } z \leq \bar{z}-1,
\end{cases}\\
&= \begin{cases}
j_1^+(z) - j_1^+(\bar{z}-1)\Lambda_{z-\bar{z}-1}\frac{\frac{1}{2}-\rho(z)}{\frac{1}{2}+\rho(z)} & \text{if } z \geq \bar{z}+1,\\
j_1^+(z) - j_1^+(\bar{z}-1)\frac{\frac{1}{2}-\rho(z)}{\frac{1}{2}+\rho(z)} & \text{if } z = \bar{z},\\
j_1^+(z) - \left[\tfrac{1}{2} - \rho(z)\right]\left[j_1^+(z-1) + j_1^+(z)\right] & \text{if } z \leq \bar{z}-1,
\end{cases}\\
&= \begin{cases}
j_1^+(z) - j_1^+(\bar{z}-1)\left(\frac{\frac{1}{2}-\rho(z)}{\frac{1}{2}+\rho(z)}\right)\left(\frac{\frac{1}{2}-\rho(z-1)}{\frac{1}{2}+\rho(z-1)}\right)\cdots\left(\frac{\frac{1}{2}-\rho(\bar{z})}{\frac{1}{2}+\rho(\bar{z})}\right) & \text{if } z \geq \bar{z},\\
\left[\tfrac{1}{2} + \rho(z)\right]j_1^+(z) - \left[\tfrac{1}{2} - \rho(z)\right]j_1^+(z-1) & \text{if } z \leq \bar{z}-1.
\end{cases}
\end{aligned}
$$

For all $z \geq \bar{z}$, we have the following due to (L30-2):

$$
\begin{aligned}
j_1^+(z) &< j_1^+(z-1)\left(\frac{\frac{1}{2}-\rho(z)}{\frac{1}{2}+\rho(z)}\right)\\
&< j_1^+(z-2)\left(\frac{\frac{1}{2}-\rho(z)}{\frac{1}{2}+\rho(z)}\right)\left(\frac{\frac{1}{2}-\rho(z-1)}{\frac{1}{2}+\rho(z-1)}\right)\\
&\vdots\\
&< j_1^+(\bar{z}-1)\left(\frac{\frac{1}{2}-\rho(z)}{\frac{1}{2}+\rho(z)}\right)\left(\frac{\frac{1}{2}-\rho(z-1)}{\frac{1}{2}+\rho(z-1)}\right)\cdots\left(\frac{\frac{1}{2}-\rho(\bar{z})}{\frac{1}{2}+\rho(\bar{z})}\right).
\end{aligned}
$$

Thus, for all $z \geq \bar{z}$, the term $(*)$ is less than 0. On the other hand, for all $z \leq \bar{z} - 1$, $j_1^+(z) \geq j_1^+(z-1)\frac{\frac{1}{2}-\rho(z)}{\frac{1}{2}+\rho(z)}$ by (L30-2); thus the term $(*)$ is greater than or equal to 0. In either case, (B.23) holds.

Step 4. We claim that the negative bet (i.e., bluffing) is a dominated action for type-1 bettor. Thus, the type-1 bettor never needs to bluff. By the Bellman equation (3.18), it suffices to verify the following:

$$\underbrace{\bar{J}^1(z) - \left[j_1^-(z) + \bar{J}^1(z-1)\right]}_{(**)} > 0. \tag{B.24}$$

Let us first evaluate the term $(**)$ when $z \leq \bar{z}$:

$$\begin{aligned}
(**) &= \left[\bar{J}^1(z) - \bar{J}^1(z-1)\right] - j_1^-(z) \\
&= -j_1^+(z-1) - j_1^-(z) && [\text{by (3.15); } z \leq \bar{z}\,] \\
&= -(2-c)\rho(z-1) + \tfrac{c}{2} - (c-2)\rho(z) + \tfrac{c}{2} && [\text{by (3.14)}] \\
&= c - (2-c)[\rho(z-1) - \rho(z)] > 0,
\end{aligned}$$

where the inequality follows because by the lemma's hypothesis, $r < \bar{r}$, and hence by (L29-4),

$\rho(z-1)-\rho(z)<\frac{c}{2-c}$. We now evaluate the term $(**)$ when $z\geq\bar{z}+1$:

$$
\begin{aligned}
&(**)\\
&=\left[\bar{J}^1(z)-\bar{J}^1(z-1)\right]-j_1^-(z)\\
&=-j_1^-(z)-j_1^+(\bar{z}-1)\Lambda_{z-1-\bar{z}} && \text{[by (3.15); } z\geq\bar{z}+1\text{]}\\
&=(2-c)\rho(z)+\tfrac{c}{2}-j_1^+(\bar{z}-1)\Lambda_{z-1-\bar{z}} && \text{[by (3.14)]}\\
&\overset{(d)}{>}\left[(2-c)\rho(z-1)+\tfrac{c}{2}\right]\left(\frac{\frac{1}{2}-\rho(z-1)}{\frac{1}{2}+\rho(z-1)}\right)-j_1^+(\bar{z}-1)\Lambda_{z-1-\bar{z}}\\
&\vdots\\
&\overset{(d)}{>}\left[(2-c)\rho(\bar{z})+\tfrac{c}{2}\right]\underbrace{\left(\frac{\frac{1}{2}-\rho(z-1)}{\frac{1}{2}+\rho(z-1)}\right)\cdots\left(\frac{\frac{1}{2}-\rho(\bar{z})}{\frac{1}{2}+\rho(\bar{z})}\right)}_{\overset{(3.16)}{=}\Lambda_{z-1-\bar{z}}}-j_1^+(\bar{z}-1)\Lambda_{z-1-\bar{z}}\\
&=\left[(2-c)\rho(\bar{z})+\tfrac{c}{2}\right]\Lambda_{z-1-\bar{z}}-j_1^+(\bar{z}-1)\Lambda_{z-1-\bar{z}}\\
&=\left[(2-c)\rho(\bar{z})+\tfrac{c}{2}-(2-c)\rho(\bar{z}-1)+\tfrac{c}{2}\right]\Lambda_{z-1-\bar{z}} && \text{[by (3.14)]}\\
&=\left[c-(2-c)\big(\rho(\bar{z}-1)-\rho(\bar{z})\big)\right]\Lambda_{z-1-\bar{z}}\overset{(e)}{>}0,
\end{aligned}
$$

where (d) follows because by the lemma's hypothesis, $r<\bar{r}$, and thus

$$\frac{\frac{1}{2}-\rho(z-1)}{\frac{1}{2}+\rho(z-1)}<\frac{(2-c)\rho(z)+\frac{c}{2}}{(2-c)\rho(z-1)+\frac{c}{2}}$$

(see (L29-4)); and (e) follows because $\rho(\bar{z}-1)-\rho(\bar{z})<\frac{c}{2-c}$ (see (L29-4)). Combining our findings in the cases where $z\leq\bar{z}$ and $z\geq\bar{z}+1$, we conclude that the term $(**)$ is greater than 0. Therefore, based on all of the conclusions from Steps 1-4, we have the desired result.

■

B.7.6 Proofs of Auxiliary Lemmas

Proof of Lemma 29. We prove each part separately. To prove (L29-1), note that (B.15) can be viewed as an extension of the function $\rho(\cdot)$ from the domain $\mathbb{Z}$ to domain $\mathbb{R}$, because it is consistent with (B.8) on the integer domain $\mathbb{Z}$. By construction, $\rho(\cdot)$ is piecewise continuous

in the region $(-\infty, 0)$ and in the region $(0, \infty)$. To see that $\rho(\cdot)$ is twice differentiable in $\mathbb{R} \setminus \{0\}$, note that $\rho(\cdot)$ is twice differentiable in $(0, \infty)$. By (A1:1) and (A1:3), $F_1 \circ F_0^{-1}(\cdot)$ is a twice differentiable function in $(-\infty, 0)$. This implies that $\rho(x)$ is twice differentiable in $(0, \infty)$. To verify that $\rho(\cdot)$ is continuous at 0 and hence in $\mathbb{R}$, observe that $\rho(0+) = \rho(0)$ and that $\rho(0) = \frac{1}{r_0} = \frac{1}{2} - F_1\left(\frac{m_0+m_1}{2}\right) \overset{\text{Lem. } 16}{=} F_0\left(\frac{m_0+m_1}{2}\right) - \frac{1}{2}$. As a result,

$$\begin{aligned}&\rho(0-)\\ =&\tfrac{1}{2} - F_1 \circ F_0^{-1}\left(\tfrac{1}{2} + \rho(0+)\right) = \tfrac{1}{2} - F_1 \circ F_0^{-1}\left(\tfrac{1}{2} + \rho(0)\right) = \tfrac{1}{2} - F_1\left(\tfrac{m_0+m_1}{2}\right) = \rho(0).\end{aligned}$$

Lastly, to show that $\rho(x)$ is strictly decreasing in x, note that for all $x \in \mathbb{R} \setminus \{0\}$,

$$\rho'(x) = \begin{cases} -\frac{r}{(r_0+rx)^2} & \text{if } x > 0, \\ \frac{f_1\left[F_0^{-1}(\frac{1}{2}+\rho(-x))\right]}{f_0\left[F_0^{-1}(\frac{1}{2}+\rho(-x))\right]}\left(\frac{-r}{(r_0-rx)^2}\right) & \text{if } x < 0, \end{cases}$$

is strictly negative. The evaluation of $\rho'(x)$ when $x < 0$ is based on the inverse function theorem (IFT) (Rudin, 1976, Theorem 9.24). Invoking the mean value theorem (MVT) (Rudin, 1976, Theorem 5.9), we conclude that $\rho(x)$ strictly decreases in x.

To prove (L29-2), observe that

$$\rho(-\infty) = \tfrac{1}{2} - F_1 \circ F_0^{-1}\left(\tfrac{1}{2} + \rho(\infty)\right) = \tfrac{1}{2} - F_1 \circ F_0^{-1}\left(\tfrac{1}{2}\right) = \tfrac{1}{2} - F_1(m_0) = \tfrac{1}{2} - \alpha,$$

and $\rho(\infty) = \lim_{x \uparrow \infty} \frac{1}{r_0+rx} = 0$. Thus, (L29-2) holds by the (strict) monotonicity of $\rho(\cdot)$.

To prove (L29-3), note that since $\rho(z) = \frac{1}{r_0+rz}$ for $z \in \mathbb{Z}_+$ and $r \in (0, 4)$, $\liminf_{n\to\infty}\{n\rho(n)\} > \frac{1}{4}$. The rest follows from Lemma 37.

To prove (L29-4), observe that for all $z \in \mathbb{Z}$,

$$
\begin{aligned}
&\log\Big(\tfrac{\rho(z)}{\rho(z+1)}\Big) \\
&= -\big[\log\rho(x)\big]' = \tfrac{-\rho'(x)}{\rho(x)} && \text{[for some } x \in (z, z+1) \subset \mathbb{R}\setminus\{0\}\text{, by MVT]} \\
&= \begin{cases} \frac{r}{r_0+rx} & \text{if } x > 0, \\ \frac{f_1\left[F_0^{-1}(\frac{1}{2}+\rho(-x))\right]}{f_0\left[F_0^{-1}(\frac{1}{2}+\rho(-x))\right]}\Big(\frac{r}{(r_0-rx)^2}\Big)\frac{1}{\rho(x)} & \text{if } x < 0, \end{cases} && \text{[by IFT]} \\
&< \begin{cases} \frac{r}{r_0} & \text{if } x > 0, \\ \frac{f_1\left[F_0^{-1}(\frac{1}{2}+\rho(-x))\right]}{f_0\left[F_0^{-1}(\frac{1}{2}+\rho(-x))\right]}\Big(\frac{r}{r_0^2}\Big)r_0 & \text{if } x < 0, \end{cases} && [x<0 \Longrightarrow \rho(x) > \rho(0) = \tfrac{1}{r_0}] \\
&\le \max\Big\{1, \max_{s\in[m_0,m_1]}\frac{f_1(s)}{f_0(s)}\Big\}\frac{r}{r_0} = \frac{\zeta_0 r}{r_0}. && \text{[by (B.13); } F_0^{-1}\left(\tfrac{1}{2}+\rho(-x)\right) \in [m_0, m_1]]
\end{aligned}
$$

As a result, for all r such that $0 < r < \bar{r}_0$,

$$
\begin{aligned}
\tfrac{\rho(z+1)}{\rho(z)} = \exp\Big(\log\tfrac{\rho(z+1)}{\rho(z)}\Big) \ge \exp\Big(-\tfrac{\zeta_0 r}{r_0}\Big) &\ge 1 - \tfrac{\zeta_0 r}{r_0} && [e^x \ge 1+x\ \forall x \in \mathbb{R}] \\
&> 1 - \tfrac{\zeta_0}{r_0}\tfrac{cr_0}{(2-c)\zeta_0} = 1 - \tfrac{c}{2-c}. && [r < \bar{r}_0 = \tfrac{cr_0}{(2-c)\zeta_0}]
\end{aligned}
$$

To see (L29-4)(a), note that for all $z \in \mathbb{Z}$,

$$
\rho(z) - \rho(z+1) = \rho(z)\Big(1 - \tfrac{\rho(z+1)}{\rho(z)}\Big) < \rho(z)\tfrac{c}{2-c} \overset{(a)}{<} \tfrac{c}{2-c},
$$

where (a) holds because $\rho(z) < 1$. To see (L29-4)(b), observe that that since $\frac{\rho(z+1)}{\rho(z)} > 1 - \frac{c}{2-c} = \frac{2-2c}{2-c}$,

$$
\tfrac{(2-c)\rho(z+1)+\frac{c}{2}}{(2-c)\rho(z)+\frac{c}{2}} > \tfrac{(2-2c)\rho(z)+\frac{c}{2}}{(2-c)\rho(z)+\frac{c}{2}} = \tfrac{c[\frac{1}{2}-\rho(z)]+(2-c)\rho(z)}{c[\frac{1}{2}+\rho(z)]+(2-2c)\rho(z)} \overset{(b)}{>} \tfrac{\frac{1}{2}-\rho(z)}{\frac{1}{2}+\rho(z)},
$$

where (b) holds because $\frac{2-c}{2-2c} > 1 > \frac{\frac{1}{2}-\rho(z)}{\frac{1}{2}+\rho(z)}$. ■

Proof of Lemma 30. In light of (B.11), let

$$
\phi(z) := \tfrac{\rho(z-1)}{2\rho(z)} - \rho(z) - \rho(z-1), \tag{B.25}
$$

so that $\bar{z} = \inf\{z : \phi(z) > \frac{1}{2} - \frac{c}{2-c}\}$. We divide the proof into four steps.

Step 1. We claim that $\phi(z)$ strictly increases in z. That is to say, for all $z \in \mathbb{Z}$, $\phi(z) >$

$\phi(z-1)$. Let us first consider the case where $z \in \mathbb{Z}_+$:

$$\begin{aligned}
\phi(z) &= \tfrac{\rho(z-1)}{2\rho(z)} - \rho(z) - \rho(z-1) && \text{[by (B.25)]} \\
&= \tfrac{r_0+rz}{2(r_0+rz-r)} - \tfrac{1}{r_0+rz} - \tfrac{1}{r_0+rz-r} && [\rho(z) = \tfrac{1}{r_0+rz}\ \forall z \in \mathbb{Z}_+] \\
&= \tfrac{x}{2(x-r)} - \tfrac{1}{x} - \tfrac{1}{x-r} && [x := r_0 + rz \geq r_0 + r] \\
&= \tfrac{1}{2} + \left(\tfrac{r}{2} - 1\right) \tfrac{1}{x-r} - \tfrac{1}{x}, && [r < 2]
\end{aligned}$$

which strictly increases in x. Next, let us consider the case where $z \in \mathbb{N}_-$; i.e., z is a negative natural number (including zero). In light of (B.15), let us consider the continuous extension of $\phi(\cdot)$ to $(-\infty, 0]$, which is continuous in $(-\infty, 0]$ and differentiable in $(-\infty, 0)$. For all $z \in \mathbb{N}_-$,

$$\begin{aligned}
\phi(z) - \phi(z-1) &= \phi'(\tilde{x}) && \text{[for some } z-1 < \tilde{x} < z\text{, by MVT]} \\
&= \tfrac{\rho'(\tilde{x}-1)\rho(\tilde{x}) - \rho'(\tilde{x})\rho(\tilde{x}-1)}{2\rho^2(\tilde{x})} - \rho'(\tilde{x}) - \rho'(\tilde{x}-1) \\
&\geq \tfrac{\rho'(\tilde{x}-1)/\rho(\tilde{x}-1) - \rho'(\tilde{x})/\rho(\tilde{x})}{2\rho(\tilde{x})/\rho(\tilde{x}-1)} && [\rho'(x) < 0\ \forall x < 0] \\
&= \tfrac{\tilde{\rho}'(\tilde{x}) - \tilde{\rho}'(\tilde{x}-1)}{2\rho(\tilde{x})/\rho(\tilde{x}-1)} && [\tilde{\rho}(x) := \log(-\rho'(x)) \in \mathcal{C}^1 \text{ on } (-\infty, 0)\ ] \\
&= \tfrac{\tilde{\rho}'(\breve{x})}{2\rho(\tilde{x})/\rho(\tilde{x}-1)}, && \text{[for some } \tilde{x}-1 < \breve{x} < \tilde{x}\text{, by MVT]}
\end{aligned}$$

where MVT stands for the mean value theorem (Rudin, 1976, Theorem 5.9). Thus, to show that $\phi(z-1) - \phi(z) > 0$ for all $z \in \mathbb{N}_-$, it suffices to show that $\tilde{\rho}'(x) > 0$ for all $x < 0$. Let us compute $\tilde{\rho}(x) = \log(-\rho'(x))$:

$$\begin{aligned}
\tilde{\rho}(x) &= \log\left(-\tfrac{d}{dx}\left[\tfrac{1}{2} - F_1 \circ F_0^{-1}\left(\tfrac{1}{2} + \rho(-x)\right)\right]\right) && \text{[by (B.15)]} \\
&= \log\left(\tfrac{f_1 \circ F_0^{-1}\left(\frac{1}{2}+\rho(-x)\right)}{f_0 \circ F_0^{-1}\left(\frac{1}{2}+\rho(-x)\right)}\left(-\rho'(-x)\right)\right) && \text{[by IFT]} \\
&= \log\left(f_1 \circ g(x)\right) - \log\left(f_0 \circ g(x)\right) + \log\left(-\rho'(-x)\right), \quad [g(x) := F_0^{-1}(\tfrac{1}{2} + \rho(-x))]
\end{aligned}$$

where IFT stands for the inverse function theorem (Rudin, 1976, Theorem 9.24). Let us now

evaluate $\tilde{\rho}'(x)$:

$$\begin{aligned}
\tilde{\rho}'(x) &= \tfrac{(f_1'\circ g(x))g'(x)}{f_1\circ g(x)} - \tfrac{(f_0'\circ g(x))g'(x)}{f_0\circ g(x)} + \tfrac{\rho''(-x)}{-\rho'(-x)} \\
&= \left[\tfrac{f_1'\circ g(x)}{f_1\circ g(x)} - \tfrac{f_0'\circ g(x)}{f_0\circ g(x)}\right] g'(x) + \tfrac{\rho''(-x)}{-\rho'(-x)} \\
&= \underbrace{\left[\tfrac{f_1'\circ g(x)}{f_1\circ g(x)} - \tfrac{f_0'\circ g(x)}{f_0\circ g(x)}\right]\left(\tfrac{1}{f_0\circ g(x)}\right)}_{\geq \zeta_1;\ \text{see (B.14)}} \underbrace{\left[-\rho'(-x)\right]}_{>0} + \tfrac{\rho''(-x)}{-\rho'(-x)} \\
&\geq \zeta_1\left[-\rho'(-x)\right] + \tfrac{\rho''(-x)}{-\rho'(-x)} \\
&= \tfrac{\zeta_1 r}{(r_0 - rx)^2} + \tfrac{2}{(r_0 - rx)} && [\rho(-x) = \tfrac{1}{r_0 - rx}\ ] \\
&= \tfrac{2r_0 + \zeta_1 r - 2rx}{(r_0 - rx)^2} \overset{(a)}{>} 0,
\end{aligned}$$

where (a) holds because (i) $x < 0$ and (ii) $\zeta_1 r + 2r_0 > 0$, which is implied by $0 < r < \bar{r}_1$ and (B.14). Therefore, $\phi(z) - \phi(z-1) > 0$ for all $z \in \mathbb{Z}$.

Step 2. We claim that (L30-1) holds. By Step 1, $\phi(z)$ is strictly increasing in z. Thus, $\bar{z} < \infty$ if and only if $\phi(\infty) > \frac{1}{2} - \frac{c}{2-c}$. Similarly, $\bar{z} > -\infty$ if and only if $\phi(-\infty) < \frac{1}{2} - \frac{c}{2-c}$. Let us evaluate $\phi(\infty)$ and $\phi(-\infty)$:

$$\begin{aligned}
\phi(\infty) &= \lim_{z\uparrow\infty}\left(\tfrac{\rho(z-1)}{2\rho(z)} - \rho(z) - \rho(z-1)\right) \overset{(b)}{=} \tfrac{1}{2} - 0 - 0 = \tfrac{1}{2}, \\
\phi(-\infty) &= \lim_{z\downarrow-\infty}\left(\tfrac{\rho(z-1)}{2\rho(z)} - \rho(z) - \rho(z-1)\right) \overset{\text{(L29-2)}}{=} \tfrac{1}{2} - 2(\tfrac{1}{2} - \alpha) = -\tfrac{1}{2} + 2\alpha.
\end{aligned}$$

In the derivations above, (b) holds because $\rho(z) = \frac{1}{r_0 + rz}$ for every $z \in \mathbb{Z}_+$. Since $c \in (0,1)$, $\frac{1}{2} - \frac{c}{2-c} < \frac{1}{2} = \phi(\infty)$. This implies that $\bar{z} < \infty$. Moreover,

$$\begin{aligned}
\phi(-\infty) - \left(\tfrac{1}{2} - \tfrac{c}{2-c}\right) &= -\tfrac{1}{2} + 2\alpha - \tfrac{1}{2} + \tfrac{c}{2-c} && [\phi(-\infty) = -\frac{1}{2} + 2\alpha] \\
&= -2\rho(-\infty) + \tfrac{c}{2-c} && [\rho(-\infty) = \frac{1}{2} - \alpha] \\
&= \tfrac{2}{c-2}\left((2-c)\rho(-\infty) - \tfrac{c}{2}\right) = \tfrac{2j^+(-\infty)}{c-2}
\end{aligned}$$

Consequently, $\bar{z} > -\infty$ if and only if $j_1^+(-\infty) > 0$.

Step 3. We claim that (L30-2) holds. Note that for all $z \in \mathbb{Z}$,

$$
\begin{aligned}
& z \geq \bar{z} \\
& \iff \phi(z) \geq \tfrac{1}{2} - \tfrac{c}{2-c} && [\bar{z} = \inf\{z : \phi(z) > \tfrac{1}{2} - \tfrac{c}{2-c}\} \text{ and } \phi(z) \uparrow \text{in } z] \\
& \iff \tfrac{\rho(z-1)}{\rho(z)} - \rho(z) - \rho(z-1) > \tfrac{1}{2} - \tfrac{c}{2-c} && [\text{by the definition of } \phi(z)] \\
& \iff \tfrac{2-c}{2}\rho(z-1) - (2-c)\rho^2(z) - (2-c)\rho(z)\rho(z-1) - \tfrac{c}{4} > \tfrac{2-c}{2}\rho(z) - c\rho(z) - \tfrac{c}{4} \\
& \iff \tfrac{2-c}{2}\rho(z-1) - (2-c)\rho(z-1)\rho(z) - \tfrac{c}{4} + \tfrac{c}{2}\rho(z) > \tfrac{2-c}{2}\rho(z) - \tfrac{c}{4} + (2-c)\rho^2(z) - \tfrac{c}{2}\rho(z) \\
& \iff \left[(2-c)\rho(z-1) - \tfrac{c}{2}\right]\left[\tfrac{1}{2} - \rho(z)\right] > \left[(2-c)\rho(z) - \tfrac{c}{2}\right]\left[\tfrac{1}{2} + \rho(z)\right] \\
& \iff j_1^+(z-1)\left[\tfrac{1}{2} - \rho(z)\right] > j_1^+(z)\left[\tfrac{1}{2} + \rho(z)\right]
\end{aligned}
$$

Step 4. We claim that (L30-3) holds. Without loss of generality, assume that $\bar{z} > -\infty$ (otherwise, the statement trivially holds). By (L30-1), we know that $j^+(-\infty) > 0$. Let $\tilde{z} := \inf\{z : j_1^+(z) \leq 0\} \in \mathbb{Z}$. By definition, we have $j_1^+(\tilde{z}-1) > 0$. Moreover, by (L29-2), $\rho(\tilde{z}) \in (0, \frac{1}{2})$. Thus, $j_1^+(\tilde{z}-1)\left[\frac{1}{2} - \rho(\tilde{z})\right] > 0 \geq j_1^+(\tilde{z})\left[\frac{1}{2} + \rho(\tilde{z})\right]$. Furthermore, in light of (L30-2), $\bar{z} \leq \tilde{z}$. Therefore, $j_1^+(\bar{z}-1) \geq j_1^+(\tilde{z}-1) > 0$. ■

Proof of Lemma 31. By Lemma 32, Lemma 31 reduces to a special case of Lemma 38 where $\bar{z} = \bar{Z}$ and the residual probability sequence is $\{\rho(z) = \frac{1}{r_0 + rz}, z \in \mathbb{Z}_+\}$. ■

B.8 Performance Analysis of IP in Theorem 11

In this section, we provide the proof details regarding the supporting results for Theorem 11.

Proof of Lemma 5. Recall that $\Delta^{\pi_I}(T) := \max\left\{\Delta_0^{\pi_I,\xi_0^*}(T), \Delta_1^{\pi_I,\xi_1^*}(T)\right\}$, where $\Delta_i^{\pi_I,\xi_i^*}(T)$ is as in (3.6). Thus, it suffices to evaluate $\Delta_1^{\pi_I,\xi_1^*}(T)$ and $\Delta_0^{\pi_I,\xi_0^*}(T)$. We start by evaluating

$\Delta_1^{\pi_I,\xi_1^*}(T)$:

$$
\begin{aligned}
&\Delta_1^{\pi_I,\xi_1^*}(T) \\
&= \tfrac{cT}{2} - \sum_{t=1}^{T} \mathbb{E}_1^{\pi_I,\xi_1^*}\left[\mathbb{I}\{(X - s_t)d_t < 0\} - (1-c)\mathbb{I}\{(X-s_t)d_t > 0\}\right] && \text{[by (3.6)]} \\
&= \tfrac{cT}{2} - \sum_{t=1}^{T} \mathbb{E}_1^{0}\big[\mathbb{I}\{a_t = 0\} r_1(s_t) + \mathbb{I}\{a_t = +1\}(-j_1^+(s_t)) \\
&\quad + \mathbb{I}\{a_t = -1\}(-j_1^-(s_t))\big] && \text{[by (3.4) \& (3.12)]} \\
&= \sum_{t=1}^{T} \mathbb{E}_1^{0}\Big[\mathbb{I}\{Z_t \geq \bar{z}\}(4-2c) \underbrace{\left(F_1(\tilde{s}(Z_t)) - \tfrac{1}{2}\right)^2}_{=\rho^2(Z_t)} \\
&\qquad\qquad + \mathbb{I}\{Z_t < \bar{z}\} \underbrace{\left(\tfrac{c}{2} + j_1^+(\tilde{s}(Z_t))\right)}_{=(2-c)\rho(Z_t)}\Big] && \text{[by Thm. 10 \& (3.4)]} \\
&= \sum_{t=1}^{T} \mathbb{E}_1^{0}\left[\mathbb{I}\{Z_t \geq \bar{z}\}(4-2c)\rho^2(Z_t) + \mathbb{I}\{Z_t < \bar{z}\}(2-c)\rho(Z_t)\right] && \text{[by (3.9) \& (3.14)]} \\
&= \sum_{t=1}^{T} \mathbb{E}_1^{0} l(Z_t) = \sum_{t=1}^{T} \mathbb{E}_1^{\pi_I,\xi_1^*}[l(Z_t)].
\end{aligned}
$$

To evaluate $\Delta_0^{\pi_I,\xi_0^*}(T)$, let us leverage the symmetry relation between the cases where $i = 0$

and $i = 1$:

$$
\begin{aligned}
&\Delta_0^{\pi_I,\xi_0^*}(T) \\
&= \tfrac{cT}{2} - \sum_{t=1}^{T} \mathbb{E}_0^{\pi_I,\xi_0^*}\left[\mathbb{I}\{(X - s_t)d_t < 0\} - (1-c)\mathbb{I}\{(X - s_t)d_t > 0\}\right] && \text{[by (3.6)]} \\
&= \tfrac{cT}{2} - \sum_{t=1}^{T} \mathbb{E}_0^{0}\big[\mathbb{I}\{a_t = 0\} r_0(s_t) + \mathbb{I}\{a_t = +1\}(-j_0^+(s_t)) \\
&\qquad\qquad + \mathbb{I}\{a_t = -1\}(-j_0^-(s_t))\big] && \text{[by (3.4) \& (3.12)]} \\
&= \sum_{t=1}^{T} \mathbb{E}_0^{0}\Big[\mathbb{I}\{Z_t \le -\bar{z}\}(4-2c) \underbrace{\left(F_0(\tilde{s}(Z_t)) - \tfrac{1}{2}\right)^2}_{=\rho^2(-Z_t)} \\
&\qquad\qquad + \mathbb{I}\{Z_t > -\bar{z}\} \underbrace{\left(\tfrac{c}{2} + j_0^-(\tilde{s}(Z_t))\right)}_{=(2-c)\rho(-Z_t)}\Big] && \text{[by Thm. 10 \& (3.4)]} \\
&= \sum_{t=1}^{T} \mathbb{E}_0^{0}\left[\mathbb{I}\{Z_t \le -\bar{z}\}(4-2c)\rho^2(-Z_t) + \mathbb{I}\{Z_t > -\bar{z}\}(2-c)\rho(-Z_t)\right] && \text{[by (3.9) \& (3.14)]} \\
&= \sum_{t=1}^{T} \mathbb{E}_1^{0}\left[\mathbb{I}\{Z_t \ge \bar{z}\}(4-2c)\rho^2(Z_t) + \mathbb{I}\{Z_t < \bar{z}\}(2-c)\rho(Z_t)\right] && \text{[by (B.19)]} \\
&= \sum_{t=1}^{T} \mathbb{E}_1^{0} l(Z_t) = \sum_{t=1}^{T} \mathbb{E}_1^{\pi_I,\xi_1^*}[l(Z_t)].
\end{aligned}
$$

In the derivations above, we adopt the notational convention that when $\bar{z} = -\infty$, (3.20) reduces to $l(z) = (4-2c)\rho^2(z)$ for all $z \in \mathbb{Z}$. As a consequence of the above identities, we have the desired result. ■

Proof of Proposition 7. Let $r \in (0, \bar{r})$, where $\bar{r}$ is as in (B.12). In addition, let π_I be an Myopic Tracking policy with the residual probability sequence $\rho = \{\rho(z) = \frac{1}{r_0 + rz}, z \in \mathbb{Z}_+\}$, and extend ρ to $\mathbb{Z}$ according to (B.8). We first the following result, the proof of which is at the end of this section.

Lemma 32. *If $r \in (0, \bar{r})$, then the sequence $\rho = \{\rho(z) = \frac{1}{r_0 + rz}, z \in \mathbb{Z}_+\}$ is regular and slow vanishing.*

Because ρ is regular and slowly vanishing (by Lemma 32), we choose $\bar{Z} := \bar{z}$ in the context

of Proposition 15 and deduce that Statements (P15:1) and (P15:3) imply Statements (P7:1) and (P7:3), respectively.

The last step is to verify Statement (P7:2). By Lemmas 26 and 27, $\rho(z) \in (0, \frac{1}{2} - \alpha)$ for all $z \in \mathbb{Z}$. Thus, we deduce from (3.20) that $(4-2c)\rho^2(z) \leq l(z)$ for all $z \in \mathbb{Z}$. By Statement (P7:3), $\sum_{t=1}^{T} \mathbb{E}_1^0[\rho^2(Z_t)] \leq \frac{1}{4-2c} \sum_{t=1}^{T} \mathbb{E}_1^0[l(Z_t)] = O(\log T)$. Hence, $\sum_{t=1}^{T} \mathbb{E}_1^0[\rho(Z_t)] \leq \sqrt{T \sum_{t=1}^{T} \left(\mathbb{E}_1^0[\rho(Z_t)]\right)^2} \leq \sqrt{T \sum_{t=1}^{T} \mathbb{E}_1^0[\rho^2(Z_t)]} = O(\sqrt{T \log T})$. Moreover, $\sum_{t=1}^{T} \mathbb{E}_1^0[\rho(Z_t)] \geq \sum_{t=1}^{T} \mathbb{E}_1^0[\rho^2(Z_t)] \to \infty$ as $T \to \infty$. ■

Proof of Lemma 32. Observe that (i) $\lim_{z\to\infty} \{z\rho(z)\} = \lim_{z\to\infty} \left\{\frac{z}{r_0+rz}\right\} = \frac{1}{r} > \frac{1}{4}$ because $r < \bar{r} < 4$, and (ii) $\lim_{z\to\infty} \{\rho(z)\} = \lim_{z\to\infty} \left\{\frac{1}{r_0+rz}\right\} = 0$. Therefore, by Definition 4, ρ is slowly vanishing. Now, note that $\lim_{z\to\infty} \left[\left(\frac{\rho(z)}{\rho(z+1)} - 1\right)z\right] = \lim_{z\to\infty} \left[\left(\frac{r_0+rz+r}{r_0+rz} - 1\right)z\right] = 1$. As a result, by Definition 5, ρ is regular. ■

B.9 Analysis of the Random Blocking Model (Theorem 12)

This section provides the details for the proof of Theorem 12, which generalizes Theorems 8 and 9 to accommodate random blocking by myopic bettors. Let us restate Theorem 12 by breaking it into two separate results, the first generalizing Theorem 8 and the second generalizing Theorem 9.

The result below, which restates Statement (T12:1) in Theorem 12, generalizes Theorem 8 by incorporating myopic bettors' random blocking. We present its proof in Appendix B.9.2.

Theorem 15. *(low blocking probability) Suppose that the market maker uses a Bayesian policy π_B with pricing function $s^{\pi_B}(\cdot)$. Then there exists $\underline{q} = \underline{q}(\hat{\Xi}) \in (0,1)$ such that for all $q \leq \underline{q}$, $b_1 \in (0,1)$, and sufficiently small $c > 0$, we have the following:*

(T15:1) (non-convergence of spread line) For some $i \in \{0,1\}$, with strictly positive $\hat{\mathbb{P}}_i^{\pi_B, \hat{\xi}_i^*}$-probability, $\mathfrak{d}_t$ does not converge to zero.

(T15:2) (linearly growing regret) $\hat{\Delta}^{\pi_B}(T) = \Omega(T)$.

When $q = 0$, Theorem 15 reduces to Theorem 8. This means that all of the conclusions in Theorem 8 hold even if we perturb the random blocking probability by a small constant (independent of T).

The following result, which restates Statement (T12:2) in Theorem 12, generalizes Theorem 9 by allowing for random blocking by myopic bettors. Its proof is in Appendix B.9.3.

Theorem 16. *(high blocking probability) Suppose that the market maker uses a Bayesian policy* π_B *with a regular pricing function* $s^{\pi_B}(\cdot)$. *Then there exists* $\bar{q} = \bar{q}(\hat{\Xi}) \in (0,1)$ *such that for all* $q \geq \bar{q}$, $b_1 \in (0,1)$ *and* $i \in \{0,1\}$, *we have:*

(T16:1) (convergence of spread lines) $\mathfrak{d}_t$ *converges to zero almost surely under* $\hat{\mathbb{P}}_i^{\pi_B, \hat{\xi}_i^*}$.

(T16:2) (rate of convergence) $\hat{\mathbb{E}}_i^{\pi_B, \hat{\xi}_i^*}[\mathfrak{d}_t] = O\big(e^{-\lambda t}\big)$ *for some constant* $\lambda > 0$.

(T16:3) (bounded regret) $\hat{\Delta}^{\pi_B}(T) = O(1)$.

When $q = 1$, Theorem 16 reduces to Theorem 9, meaning that the conclusions in Theorem 9 continue to hold even if the random blocking probability is perturbed by a small constant (independent of T).

B.9.1 Main Proof Idea: the One-stage Analysis Under Random Blocking

Our proof roadmaps for Theorems 15 and 16 are similar to those for Theorems 8 and 9. What differentiates the generalized proofs from their original versions is extending the functions $D(\cdot,\cdot)$ and $R_i(\cdot,\cdot)$ introduced in Appendix B.3.3 to incorporate random blocking by myopic bettors. Formally, define

$$\hat{D}_i(b, \mathfrak{p}) :=$$

$$[(1-q)(1-\mathfrak{p}) + qF_i\left(s^\pi\left(b\right)\right)] \log\left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + \left[(1-q)\mathfrak{p} + q\bar{F}_i\left(s^\pi\left(b\right)\right)\right] \log\left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right)$$

as the expected increment of the market maker's log-likelihood process L_t after a single bet under H_i if (i) the current belief state is b, and (ii) when his bet is not blocked, the informed

bettor bets positively with probability $\mathfrak{p}$ and negatively with probability $1-\mathfrak{p}$. Here, the expectation is taken over the random blocking by myopic bettors, the randomized strategy of the informed bettor, and the random behavior of myopic bettors. The informed bettor misleads the market maker if $\hat{D}_1(b,\mathfrak{p}) < 0$ or $\hat{D}_0(b,\mathfrak{p}) > 0$. Now, define

$$\hat{R}_i(b,\mathfrak{p}) := (1-q)\left[(1-\mathfrak{p})j_i^-(s^\pi(b)) + \mathfrak{p}j_i^+(s^\pi(b))\right]$$

to be the informed bettor's expected profit from a single bet under H_i if (i) the current belief state is b, and (ii) when his bet is not blocked, the informed bettor bets positively with probability $\mathfrak{p}$ and negatively with probability $1-\mathfrak{p}$. Here, the expectation is taken over the random blocking by myopic bettors, the randomized strategy of the informed bettor, and the final realization of the event outcome X. The informed bettor makes a profit in expectation if $\hat{R}_i(b,\mathfrak{p}) > 0$.

Summary of key steps. In the proofs of Theorems 15 and 16, we utilize our one-stage analysis in the same way we utilize it in the proofs of Theorems 8 and 9. By incorporating the additional randomness from blocking, we immediately obtain the following extended version of Lemma 19.

Lemma 33. *Let $i \in \{0,1\}$. Suppose that the market maker uses a Bayesian policy π_B and the type-i informed bettor's policy ξ is given by the behavioral strategy $\{\mathfrak{p}(b)\}$. Then, we have the following:*

1. *If there exists $\delta > 0$ such that $\hat{\mathbb{E}}_i^{\pi_B,\xi}[L_{t+1} - L_t | b_t = b] = \hat{D}_i(b,\mathfrak{p}(b)) < -\delta$ for all $b \in (0,1)$ and $t \in \mathbb{Z}_+$, then (i) $\hat{\mathbb{E}}_i^{\pi_B,\xi}[b_t] = O(e^{-\lambda t})$ for some $\lambda > 0$, and (ii) $b_t \to 0$, $L_t \to -\infty$, $s_t \to s^\pi(0+)$ almost surely. If in addition, there exists $\bar{b} \in (0,1)$ such that $\hat{R}_i(b,\mathfrak{p}(b)) > \delta$ for all $b \in (0,\bar{b}]$, then $\hat{V}_i^{\pi_B,\xi}(T) = \Omega(T)$.*

2. *If there exists $\delta > 0$ such that $\hat{\mathbb{E}}_i^{\pi_B,\xi}[L_{t+1} - L_t | b_t = b] = \hat{D}_i(b,\mathfrak{p}(b)) > \delta$ for all $b \in (0,1)$ and $t \in \mathbb{Z}_+$, then (i) $\hat{\mathbb{E}}_i^{\pi_B,\xi}[1-b_t] = O(e^{-\lambda t})$ for some $\lambda > 0$, and (ii) $b_t \to 1$, $L_t \to \infty$, $s_t \to s^\pi(1-)$ almost surely. If in addition, there exists $\bar{b} \in (0,1)$ such that $\hat{R}_i(b,\mathfrak{p}(b)) > \delta$ for all $b \in [\bar{b},1)$, then $\hat{V}_i^{\pi_B,\xi}(T) = \Omega(T)$.*

Next, we provide below a summary of the results pertaining to the one-stage analysis under random blocking. Because of the new source of randomness, the one-stage analysis is a bit more involved. These results on the generalized one-stage analysis (i.e., Lemmas 34, 35, and 36 below) are what makes the proofs of Theorems 15 and 16 different from those of Theorems 8 and 9.

The following result applied to the scenario where $s^{\pi_B}(0+) \leq m_0$, $s^{\pi_B}(1-) \geq m_1$, and the blocking probability q is sufficiently small. Recall that $\Xi = (c, m_0, m_1, F_\epsilon)$ is the collection of problem input parameters, and $\hat{\Xi} := (m_0, m_1, F_\epsilon)$ is the collection of problem input parameters concerning distribution information only, i.e., those except the commission rate c.

Lemma 34. *(one-stage analysis for manipulation under random blocking) Suppose that the market maker uses a Bayesian policy π_B such that $s^{\pi_B}(0+) \leq m_0$ and $s^{\pi_B}(1-) \geq m_1$. Then, there exist $\bar{c}_0, \underline{q}_0, \mathfrak{p}_0 \in (0,1)$, which depend only on $\hat{\Xi}$, such that for all $c \leq \bar{c}_0$ and $q \leq \underline{q}_0$, there exist $\bar{b} = \bar{b}(\Xi, q) \in (0,1)$ and $\delta = \delta(\Xi, q) > 0$ satisfying the following:*

1. *(global manipulability) For all $b \in (0,1)$, $\hat{D}_1(b,0) < -\delta$ and $\hat{D}_0(b,1) > \delta$.*
2. *(local profitable manipulation; type-1) For all $b \in (0, \bar{b}]$, $\hat{D}_1(b, \mathfrak{p}_0) < -\delta$ and $\hat{R}_1(b, \mathfrak{p}_0) > \delta$.*
3. *(local profitable manipulation; type-0) For all $b \in [1-\bar{b}, 1)$, $\hat{D}_0(b, 1-\mathfrak{p}_0) > \delta$ and $\hat{R}_0(b, 1-\mathfrak{p}_0) > \delta$.*

Observing that Lemma 34 reduces to Lemma 20 when $q = 0$, we note that the conclusions in Lemma 20 hold even if we perturb the random blocking probability by a small constant (independent of T).

The following result applies to the case where $s^{\pi_B}(0+) > m_0$ or $s^{\pi_B}(1-) < m_1$, and the blocking probability q is sufficiently small.

Lemma 35. *(one-stage analysis for honest betting under random blocking) Suppose that the market maker uses a Bayesian policy π_B such that $s^{\pi_B}(0+) > m_0$ or $s^{\pi_B}(1-) < m_1$.*

Then, there exists $\bar{c}_1 = \bar{c}_1(\hat{\Xi}, \pi_B) \in (0,1)$ *such that for all* $c \leq \bar{c}_1$ *and* $q \in (0,1)$*, there exist* $\bar{b} = \bar{b}(\Xi, q, \pi_B) \in (0,1)$ *and* $\delta = \delta(\Xi, q, \pi_B) > 0$ *satisfying the following:*

1. *(correcting power) For all* $b \in (0,1)$*,* $\hat{D}_1(b,1) > \delta$ *and* $\hat{D}_0(b,0) < -\delta$*.*

2. *(local profitable honest betting) Either of the following is true:*

   - *(type-*1*) for all* $b \in [1-\bar{b}, 1)$*,* $\hat{R}_1(b,1) > \delta$*.*
   - *(type-*0*) for all* $b \in (0, \bar{b}]$*,* $\hat{R}_0(b,0) > \delta$*.*

Note that when $q = 0$, Lemma 35 reduces to the analysis in the proof of Proposition 10, where the informed bettor honestly bet all the time.

The following result applies to the case where the blocking probability q is sufficiently large.

Lemma 36. *(one-stage analysis for high blocking probability) There exists* $\bar{q} \in (0,1)$*, which depends only on* $\hat{\Xi}$*, such that for all* $\mathfrak{p} \in (0,1)$*,* $b \in (0,1)$ *and* $q \geq \bar{q}$*,* $\hat{D}_1(b, \mathfrak{p}) > 0$ *and* $\hat{D}_0(b, \mathfrak{p}) < 0$*.*

When $q = 0$, Lemma 36 reduces to the analysis in the proof of Theorem 9, where the informed bettor does not bet at all.

B.9.2 The Low Blocking Probability Case (Theorem 15)

Following the same roadmap as in the proof of Theorem 8 (Appendix B.3.1), we aim to identify profitable strategies for the informed bettor if the market maker uses BPs in presence of random blocking by myopic bettors. There are two cases regarding the values of $s^{\pi_B}(0+)$ and $s^{\pi_B}(1-)$, each corresponding to a profitable strategy for the informed bettor.

The result below generalizes Proposition 9 by accommodating random blocking by myopic bettors.

Proposition 12. *(bluffing under random blocking) Suppose that the market maker uses a Bayesian policy* π_B *with pricing function* $s^{\pi_B}(\cdot)$ *such that* $s^{\pi}(0+) \leq m_0$ *and* $s^{\pi_B}(1-) \geq m_1$*.*

Then there exists $\underline{q}_0 = \underline{q}_0(\hat{\Xi}) \in (0,1)$ *such that for every blocking probability* $q \leq \underline{q}_0$*, initial belief* $b_1 \in (0,1)$*, hypothesis* $i \in \{0,1\}$ *and sufficiently small commission rate* c*, the type-*i *informed bettor has a "bluffing" policy* ξ_b *satisfying the following:*

(P12:1) (belief and spread line dynamics) The posterior belief b_t *converges to* $(1-i)$ *and the spread line* s_t *converges to* m_{1-i} *almost surely under* $\hat{\mathbb{P}}_i^{\pi_B,\xi_b}$*;*

(P12:2) (linearly growing profit of the informed bettor) $\hat{V}_i^{\pi_B,\xi_b}(T) = \Omega(T)$*.*

Noting that Proposition 12 reduces to Proposition 9 when $q = 0$, we deduce that the conclusions in Proposition 9 remain to be valid even if the random blocking probability is perturbed by a small constant (independent of T).

Proof of Proposition 12. After making the appropriate notational changes to incorporate the new source of randomness due to probabilistic blocking (i.e., replacing $\mathbb{P}_i$, $\mathbb{E}_i$, ξ_i^*, V, Δ, D, and R_i with $\hat{\mathbb{P}}_i$, $\hat{\mathbb{E}}_i$, $\hat{\xi}_i^*$, $\hat{V}$, $\hat{\Delta}$, $\hat{D}_i$, and $\hat{R}_i$, respectively), this proposition follows from Lemmas 33 and 34 in the same way that Proposition 9 follows from Lemmas 19 and 20. ■

The result below generalizes Proposition 10 by accommodating random blocking by myopic bettors.

Proposition 13. *(honest betting under random blocking) Suppose that the market maker uses a Bayesian policy* π_B *with pricing function* $s^{\pi_B}(\cdot)$ *such that*

$$s^{\pi_B}(0+) > m_0 \text{ or } s^{\pi_B}(1-) < m_1.$$

Then for some hypothesis $i \in \{0,1\}$*, every blocking probability* $q \in (0,1)$*, initial belief* $b_1 \in (0,1)$*, and sufficiently small commission rate* c*, the type-*i *informed bettor has an "honest" policy* ξ_h *satisfying the following:*

(P13:1) (belief and spread line dynamics) With $\hat{\mathbb{P}}_i^{\pi_B,\xi_h}$*-probability 1, posterior belief* b_t *converges to the truth,* i*. The spread line* s_t *converges to a limit* s_∞*, but* $s_\infty \neq m_i$*;*

(P13:2) (linearly growing profit of the informed bettor) $\hat{V}_i^{\pi_B,\xi_h}(T) = \Omega(T)$.

Proposition 13 means that all the conclusions in Proposition 10 (which is a special case of Proposition 13 when $q = 0$) hold even if we perturb the random blocking probability arbitrarily within the range $(0, 1)$.

Proof of Proposition 13. This proposition follows from repeating the proof of Proposition 10 with the following changes to incorporate probabilistic blocking: (i) replacing the notations $\mathbb{P}_i$, $\mathbb{E}_i$, ξ_i^*, V, Δ, D, and R_i with $\hat{\mathbb{P}}_i$, $\hat{\mathbb{E}}_i$, $\hat{\xi}_i^*$, $\hat{V}$, $\hat{\Delta}$, $\hat{D}_i$, and $\hat{R}_i$, respectively; (ii) replacing Step 1 in the proof of Proposition 10 with Lemma 35 (this step corresponds to the one-stage analysis); and (iii) replacing Lemma 19 with Lemma 33 (this step corresponds to showing how the one-stage analysis leads to the final result). ■

Proof of Theorem 15. The logical deduction from Propositions 9 and 10 to Theorem 8 is the same as from Propositions 12 and 13 to Theorem 15. In other words, let $\underline{q} := \underline{q}_0$ in Proposition 12 and fix $q \in (0, \underline{q})$. The rest of the proof follows from repeating the arguments in that of Theorem 8, with the following changes to incorporate random blocking by myopic bettors: (i) replacing $\mathbb{P}_i$, $\mathbb{E}_i$, ξ_i^*, V, Δ, $\mathbb{I}\{a_t = 0\}$, and $\mathbb{I}\{a_t \neq 0\}$ with $\hat{\mathbb{P}}_i$, $\hat{\mathbb{E}}_i$, $\hat{\xi}_i^*$, $\hat{V}$, $\hat{\Delta}$, $\mathbb{I}\{a_t = 0 \text{ or } \chi_t = 1\}$, and $\mathbb{I}\{a_t \neq 0 \text{ and } \chi_t = 0\}$, respectively; and (ii) replacing Propositions 9 and 10 with Propositions 12 and 13, respectively. ■

B.9.3 The High Blocking Probability Case (Theorem 16)

Proof of Theorem 16. This theorem follows from repeating the proof of Theorem 9 with the following changes to incorporate probabilistic blocking: (i) replacing the notations $\mathbb{P}_i$, $\mathbb{E}_i$, ξ_i^*, V, Δ, D, and R_i with $\hat{\mathbb{P}}_i$, $\hat{\mathbb{E}}_i$, $\hat{\xi}_i^*$, $\hat{V}$, $\hat{\Delta}$, $\hat{D}_i$, and $\hat{R}_i$, respectively; and (ii) replacing Step 1 in the proof of Theorem 9 with Lemmas 33 and 36. ■

B.9.4 Proofs of Auxiliary Lemmas

Proof of Lemma 34. This proof is similar to the proof of Lemma 20. We complete the proof in four steps.

Step 1. We claim that there exists $\delta_1 = \delta_1(\hat{\Xi}) > 0$ and $\varepsilon_q = \varepsilon_q(\hat{\Xi}) > 0$ such that (i) $\hat{D}_1(b,0) < -\delta_1$ and (ii) $\hat{D}_0(b,1) > \delta_1$ for all $b \in (0,1)$ and $q \leq \varepsilon_q$. By Lemma 18 and Assumption (A1:3), there exist $\bar{\delta}, M > 0$, which depend only on $\hat{\Xi}$, such that $-M \leq \log\left(\frac{F_1(s)}{F_0(s)}\right) \leq -\bar{\delta}$ and $\bar{\delta} \leq \log\left(\frac{\bar{F}_1(s)}{\bar{F}_0(s)}\right) \leq M$ for all $s \in \mathcal{S}$. Thus,

$$\begin{aligned}\hat{D}_1(b,0) &= [(1-q) + qF_1(s^\pi(b))] \log\left(\tfrac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + q\bar{F}_1(s^\pi(b)) \log\left(\tfrac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right) \\ &\leq (1-q)(-\bar{\delta}) + qM = -\bar{\delta} + (M+\bar{\delta})q,\end{aligned}$$

and

$$\begin{aligned}\hat{D}_0(b,1) &= qF_0(s^\pi(b)) \log\left(\tfrac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + \left[(1-q) + q\bar{F}_0(s^\pi(b))\right] \log\left(\tfrac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right) \\ &\geq q(-M) + (1-q)(\bar{\delta}) = \bar{\delta} - (\bar{\delta}+M)q.\end{aligned}$$

To prove our claim in this step, we choose $\delta_1 = \frac{\bar{\delta}}{2}$ and $\varepsilon_q = \frac{\bar{\delta}}{4(\bar{\delta}+M)}$, both of which depend only on $\hat{\Xi}$.

Step 2. We claim that exists $\bar{c}_0, \hat{\varepsilon}_q, \mathfrak{p}_0 \in (0,1)$, which depend only on $\hat{\Xi}$, such that for all $c \leq \bar{c}_0$ and $q \leq \hat{\varepsilon}_q$,

- $\hat{D}_1(0+, \mathfrak{p}_0) < 0$ and $\hat{R}_1(0+, \mathfrak{p}_0) > 0$;
- $\hat{D}_0(1-, 1-\mathfrak{p}_0) > 0$ and $\hat{R}_0(1-, 1-\mathfrak{p}_0) > 0$.

That is, the type-1 (resp. type-0) informed bettor can make a profitable manipulation when the blocking probability is low, the commission rate is low, and the market maker's belief is close to 0 (resp. 1). To prove this claim, recall the following three quantities introduced in Step 2 of the proof of Lemma 20 (where random blocking was not present):

$$\kappa = \tfrac{-\log 2\alpha}{\log 2(1-\alpha)} > 1, \ \hat{\kappa} = \tfrac{(\bar{c}_0-2)(1-\alpha)+1}{(\bar{c}_0-2)(1-\alpha)+1-\bar{c}_0}, \text{ and } \bar{c}_0 = \tfrac{(\kappa-1)(1-2\alpha)}{2(\kappa-1)(1-\alpha)+1} \in (0,1).$$

In that proof, $\mathfrak{p}_0$ was constructed such that it depends only on $\alpha = F_1(m_0)$ and hence only on $\hat{\Xi}$, and moreover, it satisfies the following inequalities:

$$(1-\mathfrak{p}_0)\log\left(\frac{F_1(s^\pi(0+))}{F_0(s^\pi(0+))}\right)+\mathfrak{p}_0\log\left(\frac{\bar{F}_1(s^\pi(0+))}{\bar{F}_0(s^\pi(0+))}\right)<0,$$
$$\mathfrak{p}_0\log\left(\frac{F_1(s^\pi(1-))}{F_0(s^\pi(1-))}\right)+(1-\mathfrak{p}_0)\log\left(\frac{\bar{F}_1(s^\pi(1-))}{\bar{F}_0(s^\pi(1-))}\right)>0.$$

In light of this, we choose $\mathfrak{p}_0$ in the same way, and uniquely construct $\hat{\varepsilon}_1, \hat{\varepsilon}_2 > 0$ that satisfy the following:

$$(1-\hat{\varepsilon}_1)\left[(1-\mathfrak{p}_0)\log\left(\frac{F_1(s^\pi(0+))}{F_0(s^\pi(0+))}\right)+\mathfrak{p}_0\log\left(\frac{\bar{F}_1(s^\pi(0+))}{\bar{F}_0(s^\pi(0+))}\right)\right]+\hat{\varepsilon}_1 M=0,$$
$$(1-\hat{\varepsilon}_2)\left[\mathfrak{p}_0\log\left(\frac{F_1(s^\pi(1-))}{F_0(s^\pi(1-))}\right)+(1-\mathfrak{p}_0)\log\left(\frac{\bar{F}_1(s^\pi(1-))}{\bar{F}_0(s^\pi(1-))}\right)\right]-\hat{\varepsilon}_2 M=0.$$

One can verify that both $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_2$ depend only on $\hat{\Xi}$. Now, if $q \le \hat{\varepsilon}_q := \min\{\frac{\hat{\varepsilon}_1}{2}, \frac{\hat{\varepsilon}_2}{2}, \frac{1}{2}\}$, then

$$\begin{aligned}
\hat{D}_1(0+,\mathfrak{p}_0) &= \left[(1-q)(1-\mathfrak{p}_0)+qF_1\left(s^\pi\left(0+\right)\right)\right]\log\left(\frac{F_1(s^\pi(0+))}{F_0(s^\pi(0+))}\right)\\
&\quad+\left[(1-q)\mathfrak{p}_0+q\bar{F}_1\left(s^\pi\left(0+\right)\right)\right]\log\left(\frac{\bar{F}_1(s^\pi(0+))}{\bar{F}_0(s^\pi(0+))}\right)\\
&\le (1-q)\left[(1-\mathfrak{p}_0)\log\left(\frac{F_1(s^\pi(0+))}{F_0(s^\pi(0+))}\right)+\mathfrak{p}_0\log\left(\frac{\bar{F}_1(s^\pi(0+))}{\bar{F}_0(s^\pi(0+))}\right)\right]+qM\\
&< (1-\hat{\varepsilon}_1)\left[(1-\mathfrak{p}_0)\log\left(\frac{F_1(s^\pi(0+))}{F_0(s^\pi(0+))}\right)+\mathfrak{p}_0\log\left(\frac{\bar{F}_1(s^\pi(0+))}{\bar{F}_0(s^\pi(0+))}\right)\right]+\hat{\varepsilon}_1 M=0,
\end{aligned}$$

and

$$\begin{aligned}
\hat{D}_0(b,1-\mathfrak{p}_0) &= \left[(1-q)\mathfrak{p}_0+qF_0\left(s^\pi\left(1-\right)\right)\right]\log\left(\frac{F_1(s^\pi(1-))}{F_0(s^\pi(1-))}\right)\\
&\quad+\left[(1-q)(1-\mathfrak{p}_0)+q\bar{F}_0\left(s^\pi\left(1-\right)\right)\right]\log\left(\frac{\bar{F}_1(s^\pi(1-))}{\bar{F}_0(s^\pi(1-))}\right)\\
&\ge (1-q)\left[\mathfrak{p}_0\log\left(\frac{F_1(s^\pi(1-))}{F_0(s^\pi(1-))}\right)+(1-\mathfrak{p}_0)\log\left(\frac{\bar{F}_1(s^\pi(1-))}{\bar{F}_0(s^\pi(1-))}\right)\right]-qM\\
&> (1-\hat{\varepsilon}_2)\left[\mathfrak{p}_0\log\left(\frac{F_1(s^\pi(1-))}{F_0(s^\pi(1-))}\right)+(1-\mathfrak{p}_0)\log\left(\frac{\bar{F}_1(s^\pi(1-))}{\bar{F}_0(s^\pi(1-))}\right)\right]-\hat{\varepsilon}_2 M=0.
\end{aligned}$$

Moreover, for all $q \le \hat{\varepsilon}_q$ and $c \le \bar{c}_0$,

$$\hat{R}_1(0+,\mathfrak{p}_0)=(1-q)[(1-\mathfrak{p}_0)j_1^-(s^\pi\left(0+\right))+\mathfrak{p}_0 j_1^+(s^\pi\left(0+\right))]\overset{(a)}{>}0,$$
$$\hat{R}_0(1-,1-\mathfrak{p}_0)=(1-q)[\mathfrak{p}_0 j_0^-(s^\pi\left(1-\right))+(1-\mathfrak{p}_0)j_0^+(s^\pi\left(1-\right))]\overset{(b)}{>}0.$$

In the derivations above, (a) and (b) follow from Step 2 in the proof in Lemma 20.

Step 3. Based on Step 2, there exist $\bar{b}, \delta_2, \delta_3, \delta_4, \delta_5 > 0$, all of which depend only on q and Ξ, such that

- *(local profitable manipulation; type-1)* $\hat{D}_1(b, \mathfrak{p}_0) < -\delta_2$ and $\hat{R}_1(b, \mathfrak{p}_0) > \delta_3$ for all $b \in (0, \bar{b}]$;
- *(local profitable manipulation; type-0)* $\hat{D}_1(b, 1-\mathfrak{p}_0) > \delta_4$ and $\hat{R}_0(b, 1-\mathfrak{p}_0) > \delta_5$ for all $b \in (1-\bar{b}, 1]$.

The existence is guaranteed by the local continuity of $\hat{D}_1(b, \mathfrak{p}_0)$ and $\hat{R}_1(b, \mathfrak{p}_0)$ with respect to b at point $0+$; as well as that of $\hat{D}_0(b, 1-\mathfrak{p}_0)$ and $\hat{R}_0(b, 1-\mathfrak{p}_0)$ with respect to b at point $1-$.

Step 4. In light of Steps 1 and 3, we complete the proof by taking $\underline{q}_0 = \min\{\varepsilon_q, \hat{\varepsilon}_q\}$ and $\delta := \min\{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5\}$. Note that $\underline{q}_0$ depends only on $\hat{\Xi}$, and δ depends only on (Ξ, q).

■

Proof of Lemma 35. Choose $\bar{\delta} > 0$ as in Lemma 18 so that $\log\left(\frac{F_0(s)}{F_1(s)}\right) \geq \bar{\delta}$ and $\log\left(\frac{\bar{F}_1(s)}{\bar{F}_0(s)}\right) \geq \bar{\delta}$ for all $s \in \mathcal{S}$. Let $q \in (0,1)$. Note that for all $b \in (0,1)$,

$$\begin{aligned}
&\hat{D}_1(b,1) \\
&= qF_1\left(s^\pi(b)\right) \log\left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + \left[(1-q) + q\bar{F}_1\left(s^\pi(b)\right)\right] \log\left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right) \\
&= (1-q)\log\left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right) + q\left[F_1\left(s^\pi(b)\right)\log\left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + \bar{F}_1\left(s^\pi(b)\right)\log\left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right)\right] \\
&\geq (1-q)\bar{\delta} > 0.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\hat{D}_0(b,0) \\
&= \left[(1-q) + qF_0\left(s^\pi(b)\right)\right] \log\left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + q\bar{F}_0\left(s^\pi(b)\right) \log\left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right) \\
&= (1-q)\log\left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + q\left[F_0\left(s^\pi(b)\right)\log\left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + \bar{F}_0\left(s^\pi(b)\right)\log\left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right)\right] \\
&\leq -(1-q)\bar{\delta} < 0.
\end{aligned}$$

As in the proof of Proposition 10, choose

$$\bar{c}_1 := \min\left\{\max\left\{\tfrac{2F_0(s^{\pi_B}(0+))-1}{2F_0(s^{\pi_B}(0+))}, \tfrac{1-2F_1(s^{\pi_B}(1-))}{2\bar{F}_1(s^{\pi_B}(1-))}\right\}, \tfrac{1}{2}\right\} \in (0,1),$$

and let $c \in (0, \bar{c}_1]$. If $s^{\pi_B}(0+) > m_0$, then $\hat{R}_0(0+,0) = (1-q)j_0^-(s^{\pi_B}(0+)) > 0$. If $s^{\pi_B}(1-) < m_1$, then $\hat{R}_1(1-,1) = (1-q)j_1^+(s^{\pi_B}(1-)) > 0$. Since $\hat{R}_0(b,\mathfrak{p})$ is continuous in b at $0+$ and $\hat{R}_1(b,\mathfrak{p})$ is continuous in b at $1-$, there exist $\bar{b}, \hat{\delta} > 0$, which depend only on Ξ, q, π_B, such that either of the following is true:

- (type-1) for all $b \in [1-\bar{b}, 1)$, $\hat{R}_1(b,1) > \hat{\delta}$ (this happens if $s^{\pi_B}(1-) < m_1$).
- (type-0) for all $b \in (0, \bar{b}]$, $\hat{R}_0(b,0) > \hat{\delta}$ (this happens if $s^{\pi_B}(0+) > m_0$).

The proof is finished by choosing $\delta := \min\{\frac{(1-q)\bar{\delta}}{2}, \hat{\delta}\}$. ■

Proof of Lemma 36. By Lemma 18 and Assumption (A1:3), there exist $\varepsilon, \bar{\delta}, M > 0$, which depend only on $\hat{\Xi}$, such that $-M \le \log\left(\frac{F_1(s)}{F_0(s)}\right) \le -\bar{\delta}$, $\bar{\delta} \le \log\left(\frac{\bar{F}_1(s)}{\bar{F}_0(s)}\right) \le M$, $F_0(s) - F_1(s) \ge \bar{\delta}$, $\varepsilon \le F_0(s)$, and $F_1(s) \le 1-\varepsilon$ for all $s \in \mathcal{S}$. Let $d_{KL}(x,y) := x\log\left(\frac{x}{y}\right) + (1-x)\log\left(\frac{1-x}{1-y}\right)$ be the Kullback-Leibler divergence between two Bernoulli random variables with success probabilities x and y. This function is continuous and strictly positive on the closed polytope $\mathcal{P} := \{(x,y) : y - x \ge \bar{\delta}, \varepsilon \le x \le 1-\varepsilon, y \le 1-\varepsilon\}$ and hence $\inf_{(x,y)\in\mathcal{P}}\{d_{KL}(x,y)\} > 0$. Observe that

$$\begin{aligned}
&\hat{D}_1(b,\mathfrak{p}) \\
&= \left[(1-q)(1-\mathfrak{p}) + qF_1\left(s^\pi(b)\right)\right]\log\left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + \left[(1-q)\mathfrak{p} + q\bar{F}_1\left(s^\pi(b)\right)\right]\log\left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right) \\
&\ge q\left[F_1\left(s^\pi(b)\right)\log\left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + \bar{F}_1\left(s^\pi(b)\right)\log\left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right)\right] - (1-q)M \\
&= q\left[d_{KL}\left(F_1\left(s^\pi(b)\right), F_0\left(s^\pi(b)\right)\right)\right] - (1-q)M \\
&= q\Big[\inf_{(x,y)\in\mathcal{P}}\{d_{KL}(x,y)\}\Big] - (1-q)M \\
&= q\Big[\inf_{(x,y)\in\mathcal{P}}\{d_{KL}(x,y)\} + M\Big] - M.
\end{aligned}$$

Similarly,

$$
\begin{aligned}
&\hat{D}_0(b,\mathfrak{p}) \\
&= [(1-q)(1-\mathfrak{p}) + qF_0\left(s^\pi\left(b\right)\right)] \log\left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + \left[(1-q)\mathfrak{p} + q\bar{F}_0\left(s^\pi\left(b\right)\right)\right] \log\left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right) \\
&\leq q\left[F_0\left(s^\pi\left(b\right)\right) \log\left(\frac{F_1(s^\pi(b))}{F_0(s^\pi(b))}\right) + \bar{F}_0\left(s^\pi\left(b\right)\right) \log\left(\frac{\bar{F}_1(s^\pi(b))}{\bar{F}_0(s^\pi(b))}\right)\right] + (1-q)M \\
&= -q\left[F_0\left(s^\pi\left(b\right)\right) \log\left(\frac{F_0(s^\pi(b))}{F_1(s^\pi(b))}\right) + \bar{F}_0\left(s^\pi\left(b\right)\right) \log\left(\frac{\bar{F}_0(s^\pi(b))}{\bar{F}_1(s^\pi(b))}\right)\right] + (1-q)M \\
&= -q\left[d_{KL}(\bar{F}_0\left(s^\pi\left(b\right)\right), \bar{F}_1\left(s^\pi\left(b\right)\right))\right] + (1-q)M \\
&\leq -q\Big[\inf_{(x,y)\in\mathcal{P}}\{d_{KL}(x,y)\} + M\Big] + M.
\end{aligned}
$$

We complete the proof by letting $\bar{q} := \frac{M}{M+\inf_{(x,y)\in\mathcal{P}}\{d_{KL}(x,y)\}} \in (0,1)$. ■

B.10 Analysis of the Budget-constrained Model (Theorem 13)

Proof of Theorem 13. Let π_B be the market maker's Bayesian policy, and $b_1 \in (0,1)$ be her initial belief. We complete the proof in three steps.

Step 1. We claim that for a sufficiently small commission rate $c > 0$, $\breve{\Delta}^{\pi_B}(T;K) = \Omega(T \wedge K)$. To prove this claim, observe that in the proofs of Propositions 9 and 10, we construct strategies for the informed bettor such that the informed bettor bets every time and the bets do not depend on the time horizon. Thus, using the arguments in the proofs of Proposition 9 and 10, we deduce that for some hypothesis $i \in \{0,1\}$ and sufficiently small commission rate c, the type-i informed bettor has a feasible strategy $\breve{\xi}_i$ such that $\breve{V}^{\pi_B,\breve{\xi}_i}(T;K) = \Omega(T \wedge K)$. Because the informed bettor makes profits at the expense of the market maker, we deduce from the arguments in the proof of Theorem 8 that $\breve{\Delta}^{\pi_B}(T;K) = \Omega(T \wedge K)$.

Step 2. We claim that $\breve{\Delta}^{\pi_B}(T;K) = O(T)$. The intuition for this is that the market maker's regret is at most a constant per bet. Formally, note that

$$
\breve{\Delta}^{\pi_B}(T;K) = \max\{\Delta_0^{\pi_B,\breve{\xi}_0^*}(T), \Delta_1^{\pi_B,\breve{\xi}_1^*}(T)\}
$$

and for $i \in \{0,1\}$,

$$\begin{aligned}\Delta_i^{\pi_B,\breve{\xi}_i^*}(T) &= \tfrac{cT}{2} - \sum_{t=1}^{T} \breve{\mathbb{E}}_i^{\pi_B,\breve{\xi}_i^*}\left[\mathbb{I}\{(X-s_t)\,d_t<0\} - (1-c)\mathbb{I}\{(X-s_t)\,d_t>0\}\right] \\ &= \sum_{t=1}^{T} \breve{\mathbb{E}}_i^{\pi_B,\breve{\xi}_i^*}\left[\tfrac{c}{2} - \mathbb{I}\{(X-s_t)\,d_t<0\} + (1-c)\mathbb{I}\{(X-s_t)\,d_t>0\}\right] \\ &\le \sum_{t=1}^{T} \breve{\mathbb{E}}_i^{\pi_B,\breve{\xi}_i^*}\left[\tfrac{c}{2} + (1-c)\right] = (1-\tfrac{c}{2})T.\end{aligned}$$

Step 3. We claim that if the pricing function $s^{\pi_B}(\cdot)$ is regular, then $\breve{\Delta}^{\pi_B}(T;K) = O(K)$. Without loss of generality, suppose that K is sublinear in T, that is, $\limsup_{T\to\infty}\{\frac{K}{T}\} = 0$. Choose $i \in \{0,1\}$, and observe that:

$$\begin{aligned}&\Delta_i^{\pi_B,\breve{\xi}_i^*}(T) \\ &= \sum_{t=1}^{T} \breve{\mathbb{E}}_i^{\pi_B,\breve{\xi}_i^*}\left[\tfrac{c}{2} - \mathbb{I}\{(X-s_t)\,d_t<0\} + (1-c)\mathbb{I}\{(X-s_t)\,d_t>0\}\right] \\ &= \sum_{t=1}^{T} \breve{\mathbb{E}}_i^{\pi_B,\breve{\xi}_i^*}\left[\tfrac{c}{2} - \mathbb{I}\{(X-s_t)\,d_t<0\} + (1-c)\mathbb{I}\{(X-s_t)\,d_t>0\}\right] \\ &\quad \left(\mathbb{I}\Big\{a_t = 0 \text{ or } \sum_{\ell=1}^{t}|a_\ell| > K\Big\} + \mathbb{I}\Big\{a_t \neq 0 \text{ and } \sum_{\ell=1}^{t}|a_\ell| \le K\Big\}\right) \\ &= \sum_{t=1}^{T} \breve{\mathbb{E}}_i^{\pi_B,\breve{\xi}_i^*}\left[\tfrac{c}{2} - r_i(s_t)\right]\mathbb{I}\Big\{a_t = 0 \text{ or } \sum_{\ell=1}^{t}|a_\ell| > K\Big\} \\ &+ \sum_{t=1}^{T} \breve{\mathbb{E}}_i^{\pi_B,\breve{\xi}_i^*}\left[\tfrac{c}{2} - \mathbb{I}\{(X-s_t)\,d_t<0\} + (1-c)\mathbb{I}\{(X-s_t)\,d_t>0\}\right]\mathbb{I}\{a_t \neq 0\} \\ &\le \sum_{t=1}^{T} \breve{\mathbb{E}}_i^{\pi_B,\breve{\xi}_i^*}\left[\tfrac{c}{2} - r_i(s_t)\right] + (1-\tfrac{c}{2})K.\end{aligned}$$

Because K is sublinear in T, L_t diverges in a linear speed and s_t converges to m_i in a linear speed. Hence, the first term above is independent of T and at most linear in K. As a result, $\breve{\Delta}_i^{\pi_B,\breve{\xi}_i^*}(T) = O(K)$. ■

B.11 On the Lower Bound of Regret (Theorem 14)

We prove Theorem 14 in a more general setting when the informed bettor participates in the market with a threshold strategy.

B.11.1 Description of the Setting

The $\bar{Z}$-threshold response strategy. To shed light on how the residual probability sequence should be selected, let us first consider the following family of threshold strategies for the informed bettor.

Definition 6. *($\bar{Z}$-threshold strategy) Given $\bar{Z} \in \mathbb{Z} \cup \{-\infty\}$, the type-i informed bettor's $\bar{Z}$-threshold strategy $\xi_i^{\bar{Z}}$ is such that $\xi_1^{\bar{Z}}(z) = \mathbb{I}\{z < \bar{Z}\}$ and $\xi_0^{\bar{Z}}(z) = -\mathbb{I}\{z > -\bar{Z}\}$.*

Note that the $\bar{Z}$-threshold strategy might not necessarily be the informed bettor's best response strategy in general. In fact, although the definition of IP focuses on Markov policies with state variable Z_t, the general characterization of informed bettor's best response strategy for an arbitrary IP is quite complex. However, the analysis of generic $\bar{Z}$-threshold strategies helps us understand how to make a suitable choice for the function $\rho(\cdot)$. In particular, the informed bettor's best response strategy is indeed of threshold type when $\rho(z) = \frac{1}{r_0 + rz}$ with an appropriate value of r (see Theorem 10), where the threshold $\bar{z}$ maximizes the informed bettor's profit. Since we evaluate the profit performance of IP under generic choices for the threshold $\bar{Z}$ and the function $\rho(\cdot)$, our analysis in this section generalizes Theorem 10 and 11. In this sense, a well-performing IP should at least have reasonably good performance against the informed bettor's $\bar{Z}$-threshold strategies.

To account for the generic choice of $\bar{Z}$ instead of $\bar{z}$, let us introduce

$$\tilde{\mathbb{P}}_i^z(\cdot) := \mathbb{P}_i^{\pi_I, \xi_i^{\bar{Z}}}(\cdot \,|\, Z_1 = z)\,,$$

$\tilde{\mathbb{E}}_i^z[\cdot] := \mathbb{E}_i^{\pi_I, \xi_i^{\bar{Z}}}[\cdot \,|\, Z_1 = z]$, and $\tilde{l}(z) := (2-c)\rho(z)\mathbb{I}\{z < \bar{Z}\} + (4-2c)\rho^2(z)\mathbb{I}\{z \geq \bar{Z}\}$ for the shorthand notations in this section. We also denote by $\tilde{\Delta}^{\pi_I}(T)$ the market maker's regret

under IP against the informed bettor's $\bar{Z}$-threshold strategies.

B.11.2 Key Intermediate Results

Our main results in this section characterize the dynamics of $\{Z_t\}$ under the market maker's IP and the informed bettor's $\bar{Z}$-threshold strategy. For simplicity, we consider the case where $i = 1$ without loss of generality. The analysis for the case where $i = 0$ follows from the symmetry relations between $\tilde{\mathbb{P}}_0^z$ and $\tilde{\mathbb{P}}_1^z$ akin to (B.19). The next result characterizes the convergence behavior of Z_t when $\rho(\cdot)$ is fast vanishing.

Proposition 14. *Let $\rho = \{\rho(z), z \in \mathbb{Z}_+\}$ be a fast vanishing residual probability sequence, $m > \bar{Z} - 2$, and $z \in \mathbb{Z}$. Then, $\{m\}$ is recurrent under $\tilde{\mathbb{P}}_1^z$.*

Proposition 14 means that when $\rho(z) < \frac{1}{4z}$ in the limit, Z_t does not diverge to infinity. Note that this is an undesirable property because it implies that the spread line s_t does not converge to the correct median. The main reason behind Proposition 14 is that if $\rho(z)$ becomes too small (or $\tilde{s}(z)$ gets close to the median) as $z \to \infty$, the process Z_t behaves like a simple symmetric random walk on $\mathbb{Z}$, which is recurrent. From a technical point of view, this result is closely related to the (standard) characterization of recurrence of a birth and death chain with a reflecting boundary point (see, e.g., Durrett, 2019).

In contrast to Proposition 14, Proposition 15 below characterizes the dynamics of the market when $\rho(\cdot)$ is slowly vanishing.

Proposition 15. *Let $\rho = \{\rho(z), z \in \mathbb{Z}_+\}$ be a slowly vanishing residual probability sequence. Then, we have the following:*

(P15:1) For all sufficiently large $M > 0$, $\sum_t \tilde{\mathbb{E}}_1^0[\mathbb{I}\{Z_t \leq M\}]$ converges.

(P15:2) $Z_t \to \infty$ almost surely under $\tilde{\mathbb{P}}_1^0$. As a result, $\mathfrak{d}_t$ converges to zero almost surely under $\tilde{\mathbb{P}}_1^0$.

(P15:3) If ρ is regular, then $\sum_t \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)]$ diverges at a rate satisfying $\sum_{t=1}^T \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)] = O\big(\sum_{t=1}^T \rho(t)\big)$.

(P15:4) If ρ is regular, then $\tilde{\Delta}^{\pi_I}(T)$ diverges in T at a rate satisfying $\tilde{\Delta}^{\pi_I}(T) = O\big(\sum_{t=1}^T \rho(t)\big)$.

Proposition 15 means that when $\rho(z)$ vanishes more slowly than $\frac{1}{4z}$ as $z \to \infty$, the following happens: (i) the spread line converges to the correct median, and (ii) under certain regularity conditions, the market maker's regret (loss) against the $\bar{Z}$-threshold strategy grows in the order of $O\big(\sum_{t=1}^T \rho(t)\big)$.

The above proposition extends the convergence and regret analysis in Proposition 7 and Theorem 11 in the following way: Statements (P15:1) and (P15:3) generalize Statements (P7:1) and (P7:3) in Proposition 7, respectively. As a consequence, Statements (P15:2) and (P15:4) generalize Statements (T11:1) and (T11:3) in Theorem 11, respectively. In all of the generalizations, we study a larger family of $\rho(\cdot)$ rather than the particular choice of $\rho(z) = \frac{1}{r_0 + rz}$ with $r \in (0, \bar{r})$, and we study a generic $\bar{Z}$-threshold strategy rather than the particular choice of ξ_1^*.

B.11.3 Proof of Theorem 14

Consider an IP for the market maker with some residual probability sequence $\rho = \{\rho(z) : \mathbb{Z}_+ \to \frac{1}{2} - \alpha\}$. Let c be sufficiently large as in Lemma 22 so that $\xi_i^* = \xi_\emptyset$. By construction and using the arguments in the proof of Lemma 5, we deduce that $\Delta^{\pi_I}(T) = \max\{\Delta_0^{\pi_I,\xi_\emptyset}(T), \Delta_1^{\pi_I,\xi_\emptyset}(T)\} = \max\{\tilde{\Delta}_0^{\pi_I,\xi_\emptyset}(T), \tilde{\Delta}_1^{\pi_I,\xi_\emptyset}(T)\} = \tilde{\Delta}^{\pi_I}(T) = \sum_{t=1}^T \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)]$, where the last equality follows from (B.30). Based on this, the rest of the proof is analyzes the term $\sum_{t=1}^T \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)]$. If ρ is fast vanishing (and not necessarily regular), then the market state Z_t is recurrent; see Proposition 14. In particular, the state $\{0\}$ is recurrent. As a result, $\lim_{T\to\infty} \sum_{t=1}^T \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)] \geq \lim_{T\to\infty} \sum_{t=1}^T \tilde{\mathbb{E}}_1^0[\mathbb{I}\{Z_t = 0\}]\tilde{l}(0) = \infty$. If ρ is slowly vanishing and regular, then $\sum_{t=1}^T \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)]$ diverges in T at a rate of $\sum_{t=1}^T \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)] = O\big(\sum_{t=1}^T \rho(t)\big)$;

see Proposition 15. ■

B.12 Convergence and Regret Analysis for IP in Propositions 14 and 15

In this section, we provide the details for the proofs of Proposition 14 and 15.

B.12.1 Preliminaries for the Convergence Analysis

We now introduce some auxiliary lemmas, the proofs of which are deferred to Appendix B.12.4.

Lemma 37. *Let* $\{x_n,\, n \in \mathbb{Z}_+\}$ *be a sequence satisfying* $x_n \in (0, \frac{1}{2})$ *for all* $n \in \mathbb{Z}_+$*, and let* $y_n := \prod_{m=1}^{n} \frac{\frac{1}{2}-x_m}{\frac{1}{2}+x_m}$ *for all* $n \in \mathbb{Z}_+$*. Then, we have the following:*

- *If* $\liminf_{n\to\infty}\{nx_n\} > \frac{1}{4}$*, then* $\sum_n y_n$ *converges.*
- *If* $\limsup_{n\to\infty}\{nx_n\} < \frac{1}{4}$*, then* $\sum_n y_n$ *diverges.*

The following definition provides a general condition for a residual probability sequence to be either fast vanishing or slowly vanishing.

Definition 7. *The residual probability sequence* $\{\rho(z), z \in \mathbb{Z}_+\}$ *is strictly upper bounded if* $\sup\{\rho(z) : z \in \mathbb{Z}_+\} < \frac{1}{2} - \alpha$.

Definition 7 is equivalent to saying that, as Z_t diverges to infinity (i.e., there is very strong evidence in support of one hypothesis), the spread line does not converge to the median under the *opposite* hypothesis. In particular, if ρ is either fast vanishing or slowly vanishing, then $\lim_{z\to\infty}\{\rho(z)\} = 0$, and hence, ρ is strictly upper bounded.

The next result uses the machinery in Lemma 4 to characterize the quantity

$$\sum_{t=1}^{T} \tilde{\mathbb{E}}_1^z \mathbb{I}\{Z_t \le M\}$$

for sufficiently large M, which is interpreted as the expected staying time of Z_t away from ∞. It is a generalization of Lemma 31 under the informed bettor's generic $\bar{Z}$-threshold strategy and the market maker's generic residual probability sequence $\{\rho(z), z \in \mathbb{Z}_+\}$.

Lemma 38. *Suppose that the residual probability sequence $\{\rho(z) \in (0, \frac{1}{2} - \alpha), z \in \mathbb{Z}_+\}$ is strictly upper bounded. For all $\bar{Z} \in \mathbb{Z} \cup \{-\infty\}$ and $M \in \mathbb{Z}$ satisfying $M > \bar{Z} - 2$, there exists an increasing function $\tilde{u} : \{z \in \mathbb{Z} : z > \bar{Z} - 2\} \to \mathbb{R}$ such that*

$$\sum_{t=1}^{T} \tilde{\mathbb{E}}_1^z \mathbb{I}\{Z_t \leq M\} = \tilde{\mathbb{E}}_1^z \tilde{u}(Z_{T+1}) - \tilde{u}(z) \quad \text{for all } z \in \mathbb{Z} \text{ satisfying } z > \bar{Z} - 2 \text{ and } T \in \mathbb{Z}_+. \tag{B.26}$$

The closed-form expression for $\tilde{u}(\cdot)$ is as follows:

$$\tilde{u}(z) = \begin{cases} \left(1 + \sum\limits_{n=z+1}^{M} \prod\limits_{m=n}^{M} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)}\right) \tilde{\beta} + \sum\limits_{n=z+1}^{M} \sum\limits_{k=n}^{M} \frac{1}{\frac{1}{2}+\rho(k)} \prod\limits_{m=n}^{k} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} & \text{if } \bar{Z} - 2 < z \leq M, \\ 0 & \text{if } z = M + 1, \\ \left(1 + \sum\limits_{n=M+2}^{z-1} \prod\limits_{m=M+2}^{n} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\right) \beta & \text{if } z \geq M + 2, \end{cases} \tag{B.27}$$

where $\beta > 0$ and $\tilde{\beta} < 0$ are finite constants given by

$$\begin{cases} \tilde{\beta} = -\prod_{m=\bar{Z}}^{M} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} - \sum_{k=\bar{Z}}^{M} \frac{1}{\frac{1}{2}-\rho(k)} \prod_{m=k}^{M} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}, \\ \beta = -\frac{\frac{1}{2}-\rho(M+1)}{\frac{1}{2}+\rho(M+1)} \tilde{\beta}. \end{cases} \tag{B.28}$$

B.12.2 Preliminaries for the Regret Analysis

As mentioned earlier, we generalize the function $l(\cdot)$ to allow for a generic threshold $\bar{Z}$ as follows:

$$\tilde{l}(z) = (2 - c)\rho(z)\mathbb{I}\{z < \bar{Z}\} + (4 - 2c)\rho^2(z)\mathbb{I}\{z \geq \bar{Z}\}. \tag{B.29}$$

The difference between $l(\cdot)$ and $\tilde{l}(\cdot)$ is that the former is defined for the informed bettor's optimal threshold $\bar{z}$ while the latter is defined for a generic threshold $\bar{Z}$. We introduce this

new notation so that a generalized version of (3.19) holds, i.e.,

$$\tilde{\Delta}^{\pi_I}(T) = \sum_{t=1}^{T} \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)]. \tag{B.30}$$

Next, we introduce some auxiliary lemmas, whose proofs are in Appendix B.12.5.

Lemma 39. *Let $\{x_n, n \in \mathbb{Z}_+\}$ be a sequence satisfying $x_n \in (0, \frac{1}{2})$ for all $n \in \mathbb{Z}_+$, and let $a_n := \sum_{k=1}^{n} \frac{x_k^2}{\frac{1}{2}-x_k} \prod_{m=k}^{n} \frac{\frac{1}{2}-x_m}{\frac{1}{2}+x_m}$ for all $n \in \mathbb{Z}_+$. Then, $\sum_n a_n$ diverges. If in addition, $\liminf_{n\to\infty}\{nx_n\} > \frac{1}{4}$, $\lim_{n\to\infty}\big\{\big(\frac{x_n}{x_{n+1}} - 1\big)n\big\} \in [0, \infty]$ exists, and $\lim_{n\to\infty}\{x_n\} \in [0, \frac{1}{2}]$ exists, then $a_n = O(x_n)$ as $n \to \infty$.*

The result below uses the machinery in Lemma 4 to characterize the quantity

$$\sum_{t=1}^{T} \tilde{\mathbb{E}}_1^z \tilde{l}(Z_t)$$

when $\bar{Z} > -\infty$, which is closely related to the market maker's regret, $\tilde{\Delta}^{\pi_I}(T)$.

Lemma 40. *Let the residual probability sequence $\{\rho(z), z \in \mathbb{Z}_+\}$ be slowly vanishing and regular, and $\bar{Z} > -\infty$. Then, there exists a function $v : \{\bar{Z}-1, \bar{Z}, \ldots\} \to \mathbb{R}$ that satisfies the following:*

1. *For all $z \in \{\bar{Z}-1, \bar{Z}, \ldots\}$ and $T \in \mathbb{Z}_+$,*

$$\frac{1}{4-2c} \sum_{t=1}^{T} \tilde{\mathbb{E}}_1^z \tilde{l}(Z_t) = \tilde{\mathbb{E}}_1^z v(Z_{T+1}) - v(z). \tag{B.31}$$

2. *$v(z)$ is increasing in z.*

3. *$v(z) \uparrow \infty$ as $z \uparrow \infty$ with a rate satisfying $v(z) = O\big(\sum_{k=1}^{z} \rho(k)\big)$.*

The closed-form expression of $v(\cdot)$ is given by

$$v(z) = \begin{cases} 0 & \text{if } z = \bar{Z}-1, \\ \left(1 + \sum_{n=\bar{Z}}^{z-1} \prod_{m=\bar{Z}}^{n} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\right) \frac{\rho(\bar{Z}-1)}{2} + \sum_{n=\bar{Z}}^{z-1} \sum_{k=\bar{Z}}^{n} \frac{\rho^2(k)}{\frac{1}{2}-\rho(k)} \prod_{m=k}^{n} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} & \text{if } z \geq \bar{Z}. \end{cases} \tag{B.32}$$

The following result uses the construction in Lemma 4 to characterize the quantity

$$\sum_{t=1}^{T} \tilde{\mathbb{E}}_1^z \tilde{l}(Z_t)$$

when $\bar{Z} = -\infty$, in order to analyze $\tilde{\Delta}^{\pi_I}(T)$.

Lemma 41. *Let the residual probability sequence $\{\rho(z), z \in \mathbb{Z}_+\}$ be slowly vanishing and regular, and $\bar{Z} = -\infty$. Then, there exists function $\tilde{v} : \mathbb{Z} \to \mathbb{R}_+$ that satisfies the following:*

1. *For all $z \in \mathbb{Z}$ and $T \in \mathbb{Z}_+$:*

$$\frac{1}{4-2c} \sum_{t=1}^{T} \tilde{\mathbb{E}}_1^z \tilde{l}(Z_t) = \tilde{\mathbb{E}}_1^z \tilde{v}(Z_{T+1}) - \tilde{v}(z). \tag{B.33}$$

2. *$\tilde{v}(z)$ is increasing in z.*

3. *$\tilde{v}(z) \uparrow \infty$ as $z \uparrow \infty$ at a rate satisfying $\tilde{v}(z) = O\big(\sum_{k=1}^{z} \rho(k)\big)$.*

The closed-form expression of $\tilde{v}(\cdot)$ is given by

$$\tilde{v}(z) = \begin{cases} \left(1 + \sum\limits_{n=z+1}^{-1} \prod\limits_{m=n}^{-1} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)}\right) \tilde{B} + \sum\limits_{n=z+1}^{-1} \sum\limits_{k=n}^{-1} \frac{\rho^2(k)}{\frac{1}{2}+\rho(k)} \prod\limits_{m=n}^{k} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} & \text{if } z \leq -1, \\ 0 & \text{if } z = 0, \\ \left(1 + \sum\limits_{n=1}^{z-1} \prod\limits_{m=1}^{n} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\right) B + \sum\limits_{n=1}^{z-1} \sum\limits_{k=1}^{n} \frac{\rho^2(k)}{\frac{1}{2}-\rho(k)} \prod\limits_{m=k}^{n} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} & \text{if } z \geq 1. \end{cases} \tag{B.34}$$

where B and $\tilde{B}$ are finite constants.

B.12.3 Key Steps of the Convergence and Regret Analysis

Proof of Proposition 14. Fix a residual probability sequence ρ that is fast vanishing. First, observe that if $\bar{Z} > -\infty$, then Z_t is a standard birth and death chain with a reflecting boundary. Then, $\{\bar{Z}-1\}$ is recurrent (Durrett, 2019, Exercise 5.3.4). Because $\{\bar{Z}-1, \bar{Z}, \ldots\}$ is an irreducible class, any state in that class is recurrent. Second, if $\bar{Z} = -\infty$, then we may treat $\{Z_t\}$ as a concatenation of two sub-processes: the first one is $\{Z_t\}$ restricted to $\mathbb{Z}_+ \cup \{0\}$, and the second is $\{Z_t\}$ restricted to $\mathbb{Z}_- \cup \{0\}$. We show that either sub-process visits the

state $\{0\}$ infinitely often. That is, $\{0\}$ is recurrent. The first sub-process is a standard birth and death chain with a reflecting boundary $\{0\}$. Let $\tau_0 := \inf\{t > 1 : Z_t = 0\}$ be the hitting time of of state $\{0\}$. Then, $\tilde{\mathbb{P}}_1^0(\tau_0 < \infty \,|\, Z_2 > 0) = 1$ (Durrett, 2019). The second sub-process is a "reflected" version of a standard birth and death chain. Since the residual probability sequence $\{\rho(z), z \in \mathbb{Z}_+\}$ is fast vanishing, it is strictly upper bounded (see Definition 7). Now consider the extended residual probability sequence $\{\rho(z), z \in \mathbb{Z}\}$ in the sense of (B.7). By Lemma 27, $\{\rho(z), z \in \mathbb{Z}_-\}$ is bounded away from 0. Choose $\varepsilon > 0$ such that $\rho(z) \geq \varepsilon > 0$ for all $z \leq \mathbb{Z}_-$. Then, for all $z \in \mathbb{Z}_-$, $\frac{1}{2} + \rho(z) \geq \frac{1}{2} + \varepsilon > \frac{1}{2}$. Hence, $\tilde{\mathbb{P}}_1^0(\tau_0 < \infty \,|\, Z_2 < 0) = 1$. Combining our findings, we have $\tilde{\mathbb{P}}_1^0(\tau_0 < \infty) = 1$; i.e., $\{0\}$ is recurrent. Because Z_t is irreducible when $\bar{Z} = -\infty$, all states in $\{z \in \mathbb{Z} : z > \bar{Z} - 2\}$ are recurrent. ■

Proof of Statements (P15:1) and (P15:2) in Proposition 15. Fix a residual probability sequence ρ that is slowly vanishing. Without loss of generality, let us assume that the initial value z is greater than $\bar{Z} - 2$, and Z_t is a Markov chain restrained in $\{z \in \mathbb{Z} : z > \bar{Z} - 2\}$; otherwise, Z_t increases in a deterministic manner under $\tilde{\mathbb{P}}_1^z$ until Z_t hits the region $\{z \in \mathbb{Z} : z > \bar{Z} - 2\}$.

To prove Statement (P15:1), we choose $M > \bar{Z} - 2$ and verify that $\sum_{t=1}^{T} \tilde{\mathbb{E}}_1^z \mathbb{I}\{Z_t \leq M\}$ is bounded in T. Because ρ is slowly vanishing, $\liminf_{z\to\infty}\{z\rho(z)\} > \frac{1}{4}$ and $\lim_{z\to\infty}\{\rho(z)\} = 0$ (see Definition 4). Moreover, ρ is also strictly upper bounded (see Definition 7). Invoking Lemma 38, we have $\sum_{t=1}^{T} \tilde{\mathbb{E}}_1^z \mathbb{I}\{Z_t \leq M\} = \tilde{\mathbb{E}}_1^z \tilde{u}(Z_{T+1}) - \tilde{u}(z)$, where $\tilde{u}(\cdot)$ is as in (B.27). Thus, it suffices to show that $\tilde{u}(\cdot)$ is bounded from above. Recall from Lemma 38 that $\tilde{u}(\cdot)$ is increasing, which implies that it is sufficient to prove that $\lim_{z\to\infty}\{\tilde{u}(z)\} < \infty$. By (B.27), it is equivalent to $\sum_n \prod_{m=1}^{n} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}$ being convergent. This follows from Lemma 37 as ρ is slowly vanishing.

Statement (P15:2) follows from Statement (P15:1). By the Borel-Cantelli lemma, we deduce that $\tilde{\mathbb{P}}_1^0\{Z_t \leq M \text{ infinitely often}\} = 0$. As a result, $Z_t \to \infty$ almost surely under $\tilde{\mathbb{P}}_1^0$. Since ρ is vanishing, this implies that $\rho(Z_t) \to 0$ and $\mathfrak{d}_t = |s_t - m_1| = |F_1^{-1}(\frac{1}{2} - \rho(Z_t)) -$

$F_1^{-1}(\frac{1}{2})| \to 0$ almost surely under $\tilde{\mathbb{P}}_1^0$. ■

Proof of Statements (P15:3) and (P15:4) in Proposition 15. Fix a residual probability sequence ρ that is both slowly vanishing and regular. For this proof, it suffices to verify Statement (P15:3) because Statement (P15:4) follows from Statement (P15:3) and (B.30). We complete the proof in two steps.

Step 1. We claim that $\sum_{t=1}^{T} \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)] = O\big(\sum_{t=1}^{T} \rho(t)\big)$. Observe that, for all $T \geq \bar{Z} \vee 1$,

$$\begin{aligned}
\frac{1}{4-2c}\sum_{t=1}^{T} \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)] &\overset{(a)}{=} \begin{cases} \tilde{\mathbb{E}}_1^0 \tilde{v}(Z_{T+1}) - \tilde{v}(0) & \text{if } \bar{Z} = -\infty, \\ \tilde{\mathbb{E}}_1^0 v(Z_{T+1}) - v(0) & \text{if } -\infty < \bar{Z} \leq 1, \\ \tilde{\mathbb{E}}_1^{\bar{Z}-1} v(Z_{T+2-\bar{Z}}) - v(\bar{Z}-1) + \sum_{z=0}^{\bar{Z}-2} \frac{\rho(z)}{2} & \text{if } 2 \leq \bar{Z}, \end{cases} \\
&\overset{(b)}{\leq} \begin{cases} \tilde{v}(T) - \tilde{v}(0) & \text{if } \bar{Z} = -\infty, \\ v(T) - v(0) & \text{if } -\infty < \bar{Z} \leq 1, \\ v(Z_{T+1-\bar{Z}}) - v(\bar{Z}-1) + \sum_{z=0}^{\bar{Z}-2} \frac{\rho(z)}{2} & \text{if } 2 \leq \bar{Z}, \end{cases} \\
&\overset{(c)}{=} O\bigg(\sum_{t=1}^{T} \rho(t)\bigg).
\end{aligned}$$

In the derivations above, part (a) follows from invoking (B.31) and (B.33) for the cases where $\bar{Z} = -\infty$ and $\bar{Z} > -\infty$, respectively. Moreover, the third piece of (a) holds because when

$\bar{Z} \geq 2$, Z_t increments by one with certainty (i.e., $Z_t = t-1$) until Z_t hits $\bar{Z}-1$; consequently,

$$
\begin{aligned}
&\sum_{t=1}^{T} \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)] \\
&= \sum_{t=1}^{\bar{Z}-1} \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)] + \sum_{t=\bar{Z}}^{T+1} \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)] \\
&= \sum_{t=1}^{\bar{Z}-1} \tilde{l}(t-1) + \sum_{t=\bar{Z}}^{T} \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)] \\
&= \sum_{t=1}^{\bar{Z}-1} \tilde{l}(t-1) + \sum_{t=1}^{T-\bar{Z}+1} \tilde{\mathbb{E}}_1^{\bar{Z}-1}[\tilde{l}(Z_t)] \qquad \text{[by the Markov property of } Z_t\text{]} \\
&\overset{\text{(B.29)}}{=} \sum_{z=0}^{\bar{Z}-2} (2-c)\rho(z) + \sum_{t=1}^{T-\bar{Z}+1} \tilde{\mathbb{E}}_1^{\bar{Z}-1}[\tilde{l}(Z_t)] \\
&\overset{\text{(B.31)}}{=} \sum_{z=0}^{\bar{Z}-2} (2-c)\rho(z) + (4-2c)\big[\tilde{\mathbb{E}}_1^{\bar{Z}-1} v(Z_{T+2-\bar{Z}}) - v(\bar{Z}-1)\big].
\end{aligned}
$$

Part (b) holds because (i) for all $z \in \mathbb{Z}$ and $t \in \mathbb{Z}_+$, $Z_{t+1} \leq t + z$ almost surely under $\tilde{\mathbb{P}}_1^z$; and (ii) by Lemmas 40 and 41, $v(\cdot)$ and $\tilde{v}(\cdot)$ are increasing functions. Part (c) holds since $v(T)$ and $\tilde{v}(T)$ are $O\big(\sum_{t=1}^{T} \rho(t)\big)$.

Step 2. We claim that $\sum_{t=1}^{T} \tilde{\mathbb{E}}_1^0[\tilde{l}(Z_t)] \to \infty$ as $T \to \infty$. To prove this claim, it suffices to show that $\tilde{\mathbb{E}}_1^0[v(Z_T)] \to \infty$ and $\tilde{\mathbb{E}}_1^0[\tilde{v}(Z_T)] \to \infty$ as $T \to \infty$; see part (a) in Step 1. Because ρ is slowly vanishing, $Z_t \to \infty$ almost surely under $\tilde{\mathbb{P}}_1^0$; see Statement (P15:2). By Lemma 40, $v(Z_T)$ is a nonnegative random variable such that $v(Z_T) \to \infty$ almost surely as $T \to \infty$ under $\tilde{\mathbb{P}}_1^0$. By Markov's inequality, $\tilde{\mathbb{E}}_1^0[v(Z_T)] \geq x\tilde{\mathbb{P}}_1^0(v(Z_T) \geq x)$ for all $x > 0$. Taking T to ∞ and then x to ∞, we deduce that $\tilde{\mathbb{E}}_1^0[v(Z_T)] \to \infty$ as $T \to \infty$. Repeating the same arguments for $\tilde{\mathbb{E}}_1^0[\tilde{v}(Z_T)]$ and invoking Lemma 41, we obtain the desired result. ■

B.12.4 Proofs of Auxiliary Lemmas for Convergence Analysis

Proof of Lemma 37. Observe that $\frac{y_n}{y_{n+1}} = \frac{\frac{1}{2}+x_{n+1}}{\frac{1}{2}-x_{n+1}} = 1 + \frac{2x_{n+1}}{\frac{1}{2}-x_{n+1}}$. Thus, we have the following:

- If $\liminf_{n\to\infty}\{nx_n\} > \frac{1}{4}$, then

$$\liminf_{n\to\infty}\left\{\left(\frac{y_n}{y_{n+1}}-1\right)n\right\} = \liminf_{n\to\infty}\left\{\frac{2x_{n+1}}{\frac{1}{2}-x_{n+1}}n\right\} \geq \liminf_{n\to\infty}\{4x_{n+1}n\} > 1.$$

- If $\limsup_{n\to\infty}\{nx_n\} < \frac{1}{4}$, then

$$\limsup_{n\to\infty}\left\{\left(\frac{y_n}{y_{n+1}}-1\right)n\right\} = \limsup_{n\to\infty}\left\{\frac{2x_{n+1}}{\frac{1}{2}-x_{n+1}}n\right\} = \limsup_{n\to\infty}\{4x_{n+1}n\} < 1.$$

Invoking Raabe's test of convergence of series with positive terms (Bromwich, 1908, p. 33), we complete the proof. ■

Proof of Lemma 38. With the initial point $z \in \mathbb{Z}$ satisfying $z > \bar{Z} - 2$, the Markov chain Z_t is contained in the region $\{z \in \mathbb{Z} : z > \bar{Z} - 2\}$. In particular, if $\bar{Z}$ is finite, the state $\bar{Z} - 1$ is a reflecting boundary. Invoking Lemma 4 with $S = \{z \in \mathbb{Z} : z > \bar{Z} - 2\}$, it is sufficient to solve the difference equation $\tilde{\mathbb{E}}_1^z \tilde{u}(Z_2) - \tilde{u}(z) = \mathbb{I}\{z \leq M\}$ for all $z \in S$. The rest of the proof confirms that (B.27) presents such a solution with the desired monotonicity property. We complete the remainder of the proof in three steps.

Step 1. We verify that both β and $\tilde{\beta}$ are finite, and hence the function $\tilde{u}(\cdot)$ in (B.27) is finitely valued. Note that it suffices to check that $\tilde{\beta}$ is finite. As ρ is strictly upper bounded (see Definition 7), we deduce from Lemma 27 that there exists $\varepsilon > 0$ satisfying $\rho(z) \geq \varepsilon > 0$ for all $z \leq M$. Thus,

$$\begin{aligned}|\tilde{\beta}| = \prod_{m=\bar{Z}}^{M} \tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} + \sum_{k=\bar{Z}}^{M} \tfrac{1}{\frac{1}{2}-\rho(k)} \prod_{m=k}^{M} \tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} &= \underbrace{\prod_{m=\bar{Z}}^{M} \tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}}_{\leq 1} + \sum_{k=\bar{Z}}^{M} \underbrace{\tfrac{1}{\frac{1}{2}+\rho(k)}}_{\leq \frac{1}{\frac{1}{2}+\varepsilon}} \underbrace{\prod_{m=k+1}^{M} \tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}}_{\leq \left(\frac{\frac{1}{2}-\varepsilon}{\frac{1}{2}+\varepsilon}\right)^{M-k}} \\ &\overset{(a)}{\leq} 1 + \sum_{k=-\infty}^{M} \tfrac{1}{\frac{1}{2}+\varepsilon} \left(\tfrac{\frac{1}{2}-\varepsilon}{\frac{1}{2}+\varepsilon}\right)^{M-k} < \infty,\end{aligned}$$

where (a) follows because $\rho(z) \geq \varepsilon$ for all $z \leq M$, which implies that $\prod_{m=k+1}^{M} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} \leq \left(\frac{\frac{1}{2}-\varepsilon}{\frac{1}{2}+\varepsilon}\right)^{M-k}$ and $\frac{1}{\frac{1}{2}+\rho(k)} \leq \frac{1}{\frac{1}{2}+\varepsilon}$ for all $k \leq M$.

Step 2. We verify that $\tilde{u}(\cdot)$ is increasing. That is, $\tilde{u}(z-1) - \tilde{u}(z) < 0$ for all $z \in \mathbb{Z}$

satisfying $z \geq \bar{Z}$. In particular, when $\bar{Z} > -\infty$, $\tilde{u}(\bar{Z}-1) - \tilde{u}(\bar{Z}) = -1$. Since $\tilde{\beta} < 0$ and $\beta > 0$, we can directly verify from the expression in (B.27) that $\tilde{u}(M) = \tilde{\beta} < 0 = \tilde{u}(M+1) < \beta = \tilde{u}(M+2) \leq \tilde{u}(M+3) \leq \tilde{u}(M+4) \leq \cdots$ Therefore, it suffices to show that $\tilde{u}(z-1) - \tilde{u}(z) < 0$ for all z satisfying $\bar{Z} - 1 < z \leq M$. Note that

$$
\begin{aligned}
&\tilde{u}(z-1) - \tilde{u}(z) \\
&= \tilde{\beta} \prod_{m=z}^{M} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} + \sum_{k=z}^{M} \tfrac{1}{\frac{1}{2}+\rho(k)} \prod_{m=z}^{k} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} \\
&\overset{\text{(B.28)}}{=} \left(- \prod_{m=\bar{Z}}^{M} \tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} - \sum_{k=\bar{Z}}^{M} \tfrac{1}{\frac{1}{2}-\rho(k)} \prod_{m=k}^{M} \tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} \right) \prod_{m=z}^{M} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} + \sum_{k=z}^{M} \tfrac{1}{\frac{1}{2}+\rho(k)} \prod_{m=z}^{k} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} \\
&\overset{\bar{Z} \leq z}{\leq} \left(- \prod_{m=\bar{Z}}^{M} \tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} - \sum_{k=z}^{M} \tfrac{1}{\frac{1}{2}-\rho(k)} \prod_{m=k}^{M} \tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} \right) \prod_{m=z}^{M} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} + \sum_{k=z}^{M} \tfrac{1}{\frac{1}{2}+\rho(k)} \prod_{m=z}^{k} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} \\
&= - \prod_{m=\bar{Z}}^{z-1} \tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} - \sum_{k=z}^{M} \tfrac{1}{\frac{1}{2}-\rho(k)} \prod_{m=z}^{k-1} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} + \sum_{k=z}^{M} \tfrac{1}{\frac{1}{2}+\rho(k)} \prod_{m=z}^{k} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} \\
&= - \prod_{m=\bar{Z}}^{z-1} \tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} < 0.
\end{aligned}
$$

The inequality above is tight when $\bar{Z} > -\infty$ and $z = \bar{Z}$.That is, when $\bar{Z} > -\infty$, $\tilde{u}(\bar{Z}-1) - \tilde{u}(\bar{Z}) = -1$.

Step 3. As explained above, by Lemma 4, it suffices to show that for all $z > \mathbb{Z} - 2$ satisfying $z \geq \bar{Z}$, $\tilde{\mathbb{E}}_1^z \tilde{u}(Z_2) - \tilde{u}(z) = \mathbb{I}\{z \leq M\}$. Let us first look at the left-hand side of the equation:

$$
\tilde{\mathbb{E}}_1^z \tilde{u}(Z_2) - \tilde{u}(z) = \begin{cases} \left[\frac{1}{2} + \rho(z)\right] \tilde{u}(z+1) + \left[\frac{1}{2} - \rho(z)\right] \tilde{u}(z-1) - \tilde{u}(z) & \text{if } z \geq \bar{Z}, \\ \tilde{u}(z+1) - \tilde{u}(z) & \text{if } z < \bar{Z}. \end{cases}
$$

Let us then look at the right-hand side of the equation:

$$
\mathbb{I}\{z \leq M\} = \begin{cases} 1 & \text{if } z \leq M, \\ 0 & \text{if } z > M. \end{cases}
$$

Note that we only consider the values of M that exceed $\bar{Z} - 2$. We study four cases for the

value of z: $z \geq M+2$, $z = M+1$, $\bar{Z} - 1 < z \leq M$, and $z = \bar{Z} - 1$ (the last case is valid only when $\bar{Z} > -\infty$).

1. When $z \geq M+2$, we show that $\left[\frac{1}{2} + \rho(z)\right] \tilde{u}(z+1) + \left[\frac{1}{2} - \rho(z)\right] \tilde{u}(z-1) - \tilde{u}(z) = 0$. By Lemma 28, $\tilde{u}(\cdot)$ restricted to the domain $[M+1, \infty)$ solves the difference equation (B.9) with $\hat{z} = M+2$, $\delta = \beta$, $p(z) = \frac{1}{2} + \rho(z)$, and $x(z) = 0$ for all $z \geq M+1$. That is, $\tilde{u}(\cdot)$ satisfies the following:

$$\begin{cases} \tilde{u}(M+1) = 0,\ \tilde{u}(M+2) = \beta \\ \left[\frac{1}{2} + \rho(z)\right] \tilde{u}(z+1) + \left[\frac{1}{2} - \rho(z)\right] \tilde{u}(z-1) - \tilde{u}(z) = 0 \ \text{ for all } z \geq M+2. \end{cases}$$

   Thus, for all $z \geq M+2$, $\left[\frac{1}{2} + \rho(z)\right] \tilde{u}(z+1) + \left[\frac{1}{2} - \rho(z)\right] \tilde{u}(z-1) - \tilde{u}(z) = 0$.

2. When $z = M+1$, the definitions of β and $\tilde{\beta}$ imply that $\left[\frac{1}{2} + \rho(M+1)\right] \tilde{u}(M+2) + \left[\frac{1}{2} - \rho(M+1)\right] \tilde{u}(M) - \tilde{u}(M+1) = \left[\frac{1}{2} + \rho(M+1)\right] \beta + \left[\frac{1}{2} - \rho(M+1)\right] \tilde{\beta} = 0$.

3. When $\bar{Z} - 1 < z \leq M$, we show that $\left[\frac{1}{2} + \rho(z)\right] \tilde{u}(z+1) + \left[\frac{1}{2} - \rho(z)\right] \tilde{u}(z-1) - \tilde{u}(z) = 1$. To that end, we introduce an auxiliary sequence $\{y(z), z \in \mathbb{N}\}$ such that

$$y(z) = \begin{cases} 0 & \text{if } z = 0, \\ \left(1 + \sum_{n=1}^{z-1} \prod_{m=1}^{n} \frac{\frac{1}{2}+\rho(M+1-m)}{\frac{1}{2}-\rho(M+1-m)}\right) \tilde{\beta} + \sum_{n=1}^{z-1} \sum_{k=1}^{n} \frac{1}{\frac{1}{2}+\rho(M+1-k)} \prod_{m=n}^{k} \frac{\frac{1}{2}+\rho(M+1-m)}{\frac{1}{2}-\rho(M+1-m)} & \text{if } z \geq 1. \end{cases}$$

   The above construction of $y(\cdot)$ has two desirable properties. First, $y(\cdot)$ is a "reflected" version of $\tilde{u}(\cdot)$. Specifically, for all $z \in \mathbb{Z}$ satisfying $\bar{Z} - 1 < z \leq M$,

$$\begin{aligned} & y(M+1-z) \\ &= \left(1 + \sum_{n=1}^{M-z} \prod_{m=1}^{n} \frac{\frac{1}{2}+\rho(M+1-m)}{\frac{1}{2}-\rho(M+1-m)}\right) \tilde{\beta} + \sum_{n=1}^{M-z} \sum_{k=1}^{n} \frac{1}{\frac{1}{2}+\rho(M+1-k)} \prod_{m=n}^{k} \frac{\frac{1}{2}+\rho(M+1-m)}{\frac{1}{2}-\rho(M+1-m)} \\ &\overset{(a)}{=} \left(1 + \sum_{\check{n}=z+1}^{M} \prod_{\check{m}=\check{n}}^{M} \frac{\frac{1}{2}+\rho(\check{m})}{\frac{1}{2}-\rho(\check{m})}\right) \tilde{\beta} + \sum_{\check{n}=z+1}^{M} \sum_{\check{k}=\check{n}}^{M} \frac{1}{\frac{1}{2}+\rho(\check{k})} \prod_{\check{m}=\check{n}}^{\check{k}} \frac{\frac{1}{2}+\rho(\check{m})}{\frac{1}{2}-\rho(\check{m})} = \tilde{u}(z), \end{aligned}$$

where (a) holds by the following change of variables: $\check{n} = M+1-n$, $\check{m} = M+1-m$, and $\check{k} = M+1-k$. Second, we deduce from Lemma 28 that $y(\cdot)$ solves the difference equation (B.9) with $\hat{z} = 1, \delta = \tilde{\beta}$, $p(z) = \frac{1}{2} - \rho(M+1-z)$. That is, $y(0) = 0$, $y(1) = \tilde{\beta}$, and

$$\left[\tfrac{1}{2} - \rho(M+1-z)\right] y(z+1) + \left[\tfrac{1}{2} + \rho(M+1-z)\right] y(z-1) - y(z) = 1 \text{ for all } z \geq 1.$$

Substituting $\tilde{u}(z) = y(M+1-z)$, we obtain the following for $\tilde{u}(\cdot)$:

$$\begin{cases} \tilde{u}(M+1) = 0,\ \tilde{u}(M) = \tilde{\beta}, \\ \left[\tfrac{1}{2} - \rho(z)\right] \tilde{u}(z-1) + \left[\tfrac{1}{2} + \rho(z)\right] \tilde{u}(z+1) - \tilde{u}(z) = 1 \ \text{ for all } \bar{Z} - 1 < z \leq M. \end{cases}$$

Thus, for all z satisfying $\bar{Z} - 1 < z \leq M$, $\left[\frac{1}{2} + \rho(z)\right] \tilde{u}(z+1) + \left[\frac{1}{2} - \rho(z)\right] \tilde{u}(z-1) - \tilde{u}(z) = 1$.

4. When $z = \bar{Z} - 1$, we show that $\tilde{u}(\bar{Z}) - \tilde{u}(\bar{Z}-1) = 1$ (this case is valid only if $\bar{Z} > -\infty$). Note that

$$\begin{aligned} \tilde{u}(\bar{Z}-1) - \tilde{u}(\bar{Z}) &= \tilde{\beta} \prod_{m=\bar{Z}}^{M} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} + \sum_{k=\bar{Z}}^{M} \tfrac{1}{\frac{1}{2}+\rho(k)} \prod_{m=\bar{Z}}^{k} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} \\ &\overset{(B.28)}{=} -1 - \sum_{k=\bar{Z}}^{M} \tfrac{1}{\frac{1}{2}-\rho(k)} \prod_{m=\bar{Z}}^{k-1} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} + \sum_{k=\bar{Z}}^{M} \tfrac{1}{\frac{1}{2}+\rho(k)} \prod_{m=\bar{Z}}^{k} \tfrac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} = -1. \end{aligned}$$

Combining our findings in the four cases above, we conclude that $\tilde{\mathbb{E}}_1^z \tilde{u}(Z_2) - \tilde{u}(z) = \mathbb{I}\{z \leq M\}$ for all $z \in \mathbb{Z}$ satisfying $z > \bar{Z} - 2$. ■

B.12.5 Proofs of Auxiliary Lemmas for Regret Analysis

Proof of Lemma 39. We first derive a recursive relation for $\{a_n\}$ as follows:

$$\begin{aligned}
&a_{n+1}\\
=&\sum_{k=1}^{n+1}\frac{x_k^2}{\frac{1}{2}-x_k}\prod_{m=k}^{n+1}\frac{\frac{1}{2}-x_m}{\frac{1}{2}+x_m}\\
=&\frac{\frac{1}{2}-x_{n+1}}{\frac{1}{2}+x_{n+1}}\underbrace{\sum_{k=1}^{n}\frac{x_k^2}{\frac{1}{2}-x_k}\prod_{m=k}^{n}\frac{\frac{1}{2}-x_m}{\frac{1}{2}+x_m}}_{=a_n}+\frac{x_{n+1}^2}{\frac{1}{2}+x_{n+1}}\\
=&\frac{\frac{1}{2}-x_{n+1}}{\frac{1}{2}+x_{n+1}}a_n+\frac{x_{n+1}^2}{\frac{1}{2}+x_{n+1}}. \qquad \text{(B.35)}
\end{aligned}$$

We break the rest of the proof into two steps.

Step 1. We claim that $\sum_n a_n$ diverges. To prove this claim, we deduce from (B.35) the following for sufficiently large n:

$$\begin{aligned}
\left(\frac{a_n}{a_{n+1}}-1\right)n&=\left(\frac{\frac{1}{2}+x_{n+1}}{\frac{1}{2}-x_{n+1}}-\frac{x_{n+1}^2}{\frac{1}{2}-x_{n+1}}\frac{1}{a_{n+1}}-1\right)n\\
&=\left(\frac{2x_{n+1}}{\frac{1}{2}-x_{n+1}}-\frac{x_{n+1}^2}{\frac{1}{2}-x_{n+1}}\frac{1}{a_{n+1}}\right)n\\
&=\left[-2+\frac{1}{a_{n+1}}-\left(\frac{\frac{1}{2}-x_{n+1}}{a_{n+1}}+\frac{\frac{1}{4a_{n+1}}-1}{\frac{1}{2}-x_{n+1}}\right)\right]n && \text{[rearranging terms]}\\
&\le\left[-2+\frac{1}{a_{n+1}}-2\sqrt{\frac{1}{a_{n+1}}\left(\frac{1}{4a_{n+1}}-1\right)}\right]n && [a+b\ge 2\sqrt{ab}\ \forall a,b>0\ ]\\
&=\frac{4n}{\frac{1}{a_{n+1}}-2+2\sqrt{\frac{1}{4a_{n+1}^2}-\frac{1}{a_{n+1}}}}\le\frac{4n}{\frac{1}{a_{n+1}}-2}=\frac{4a_{n+1}n}{1-2a_{n+1}}.
\end{aligned}$$

Suppose towards a contradiction that $\sum_n a_n$ converges. This implies that $\frac{4a_{n+1}n}{1-2a_{n+1}}\to 0$ as $n\to\infty$ because $\sum_n \frac{1}{n}$ is a divergent series. However, due to the derivations above, this means $\left(\frac{a_n}{a_{n+1}}-1\right)n\to 0$ as $n\to\infty$. Applying Raabe's test of convergence of series with positive terms (Bromwich, 1908, p. 33) to $\{a_n\}$, we note that $\sum_n a_n$ diverges, leading to a contradiction as desired.

Step 2. We claim that if $\lim_{n\to\infty}\{x_n\}>0$, then $a_n=O(x_n)$. Observe that when $\lim_{n\to\infty}\{x_n\}>0$, there exists $\varepsilon_0\in(0,\frac{1}{2})$ and $N_0\in\mathbb{N}$ such that $x_n\ge\varepsilon_0>0$ for all $n\ge N_0$.

Thus, to prove our claim in this step, it suffices to show that $\{a_n\}$ is bounded from above. For $n \geq N_0$,

$$\begin{aligned} a_{n+1} = \tfrac{\frac{1}{2}-x_{n+1}}{\frac{1}{2}+x_{n+1}} a_n + \tfrac{x_{n+1}^2}{\frac{1}{2}+x_{n+1}} &\leq \tfrac{\frac{1}{2}-\varepsilon_0}{\frac{1}{2}+\varepsilon_0} a_n + \tfrac{1}{2} \\ &\leq \tfrac{\frac{1}{2}-\varepsilon_0}{\frac{1}{2}+\varepsilon_0} \left(\tfrac{\frac{1}{2}-\varepsilon_0}{\frac{1}{2}+\varepsilon_0} a_{n-1} + \tfrac{1}{2} \right) + \tfrac{1}{2} \\ &\vdots \\ &\leq \limsup_{k\to\infty} \left\{ \left(\tfrac{\frac{1}{2}-\varepsilon_0}{\frac{1}{2}+\varepsilon_0} \right)^k a_{N_0} + \tfrac{1}{2} \left(1 + \cdots + \left(\tfrac{\frac{1}{2}-\varepsilon_0}{\frac{1}{2}+\varepsilon_0} \right)^{k-1} \right) \right\}, \end{aligned}$$

which is a finite constant independent of n.

Step 3. We claim that if (i) $\liminf_{n\to\infty}\{nx_n\} > \frac{1}{4}$, (ii) $\lim_{n\to\infty}\left\{\left(\frac{x_n}{x_{n+1}} - 1\right)n\right\}$ exists, and (iii) $\lim_{n\to\infty}\{x_n\} = 0$, then $a_n = O(x_n)$. Let us define $c_n := \frac{a_n}{x_n}$. To show that $a_n = O(x_n)$, it suffices to prove that $\{c_n\}$ is bounded from above. Invoking the recursive relation between a_{n+1} and a_n in (B.35), we obtain a recursive relation between c_{n+1} and c_n. That is, $c_{n+1} = \frac{\frac{1}{2}-x_{n+1}}{\frac{1}{2}+x_{n+1}} \frac{x_n}{x_{n+1}} c_n + \frac{x_{n+1}}{\frac{1}{2}+x_{n+1}}$. We seek finite constants $M_1, M_0, N_1 \in \mathbb{N}$ such that for all $n \geq N_1$, (i) $c_n \geq M_0$ implies that $c_{n+1} \leq c_n$, and (ii) $c_n \leq M_0$ implies that $c_{n+1} \leq M_1$. If such a tuple (M_0, M_1, N_1) exists, the sequence $\{c_n\}$ is bounded by the value $\max\{c_1, \ldots, c_{N_1+1}, M_0, M_1\}$. To find (M_0, M_1, N_1), let us evaluate the following quantity for every $M > 0$:

$$\begin{aligned} (*) &= \tfrac{1}{x_n} \left(\tfrac{\frac{1}{2}-x_{n+1}}{\frac{1}{2}+x_{n+1}} \tfrac{x_n}{x_{n+1}} M + \tfrac{x_{n+1}}{\frac{1}{2}+x_{n+1}} - M \right) \\ &= n \left(\tfrac{x_n}{x_{n+1}} - 1 \right) \tfrac{M}{nx_n} - \tfrac{4}{1+2x_{n+1}} M + \tfrac{x_{n+1}}{x_n} \tfrac{1}{\frac{1}{2}+x_{n+1}} \\ &< n \left(\tfrac{x_n}{x_{n+1}} - 1 \right) \tfrac{M}{nx_n} - \tfrac{4}{1+2x_{n+1}} M + 2 \left(\tfrac{x_{n+1}}{x_n} \right). \end{aligned}$$

We claim that $(*)$ is negative for sufficiently large n and M. To show this claim, recall that $\sum_n x_n$ diverges. Because $\lim_{n\to\infty}\left\{\left(\frac{x_n}{x_{n+1}} - 1\right)n\right\}$ exists, we apply Raabe's test to the sequence $\{x_n\}$ and conclude that $\lim_{n\to\infty}\left\{\left(\frac{x_n}{x_{n+1}} - 1\right)n\right\} \leq 1$. That is, there exists a constant $A \leq 1$ such that $\frac{x_n}{x_{n+1}} = 1 + \frac{A}{n} + o\left(\frac{1}{n}\right)$. This implies that $\lim_{n\to\infty}\left\{\frac{x_n}{x_{n+1}}\right\} = \lim_{n\to\infty}\left\{\frac{x_{n+1}}{x_n}\right\} = 1$. Moreover, $\lim_{n\to\infty}\{x_n\} = 0$. Combining all the pieces, we know that

there exist $\varepsilon_1 > 0$ and $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$, (i) $n\big(\frac{x_n}{x_{n+1}} - 1\big) \leq 1 + \frac{2\varepsilon_1}{1+\varepsilon_1}$, (ii) $nx_n \geq \frac{1}{4} + \varepsilon_1$, (iii) $x_{n+1} \leq \frac{\varepsilon_1}{2(1+4\varepsilon_1)}$, and (iv) $\frac{1}{2} \leq \frac{x_{n+1}}{x_n} \leq 2$. Thus, for all $n \geq N_1$,

$$\begin{aligned}
(*) &< \left(1 + \tfrac{2\varepsilon_1}{1+\varepsilon_1}\right) \tfrac{M}{\frac{1}{4}+\varepsilon_1} - \tfrac{4}{1+\frac{\varepsilon_1}{1+4\varepsilon_1}} M + 4 \\
&= 4M \left[\left(1 + \tfrac{2\varepsilon_1}{1+\varepsilon_1}\right) \tfrac{1}{1+4\varepsilon_1} - 1 + \tfrac{\frac{\varepsilon_1}{1+4\varepsilon_1}}{1+\frac{\varepsilon_1}{1+4\varepsilon_1}} \right] + 4 \\
&< 4M \left[1 - \tfrac{2\varepsilon_1}{1+4\varepsilon_1} - 1 + \tfrac{\varepsilon_1}{1+4\varepsilon_1} \right] + 4 = -\tfrac{4\varepsilon_1}{1+\varepsilon_1} M + 4. \qquad \text{(B.36)}
\end{aligned}$$

Choose M_0 such that $-\frac{4\varepsilon_1}{1+\varepsilon_1} M + 4 < 0$ for all $M \geq M_0$. This is possible because $\frac{4\varepsilon_1}{1+\varepsilon_1} > 0$. By construction, for all n satisfying $n \geq N_1$ and $c_n \geq M_0$,

$$\begin{aligned}
c_{n+1} - c_n &= \tfrac{\frac{1}{2}-x_{n+1}}{\frac{1}{2}+x_{n+1}} \tfrac{x_n}{x_{n+1}} c_n + \tfrac{x_{n+1}}{\frac{1}{2}+x_{n+1}} - c_n \\
&= x_n \tfrac{1}{x_n} \left(\tfrac{\frac{1}{2}-x_{n+1}}{\frac{1}{2}+x_{n+1}} \tfrac{x_n}{x_{n+1}} c_n + \tfrac{x_{n+1}}{\frac{1}{2}+x_{n+1}} - c_n \right) \\
&< x_n \left(-\tfrac{4\varepsilon_1}{1+\varepsilon_1} c_n + 4 \right) < 0. \qquad \text{[choosing } M = c_n \geq M_0 \text{ in (B.36)]}
\end{aligned}$$

Lastly, it is useful to observe that for all n such that $n \geq N_1$ and $c_n \leq M_0$,

$$c_{n+1} = \underbrace{\tfrac{\frac{1}{2}-x_{n+1}}{\frac{1}{2}+x_{n+1}}}_{\leq 1} \underbrace{\tfrac{x_n}{x_{n+1}}}_{\leq 2} c_n + \underbrace{\tfrac{x_{n+1}}{\frac{1}{2}+x_{n+1}}}_{\leq 1} \leq 2M_0 + 1.$$

Thus, it suffices to choose $M_1 = 2M_0 + 1$. We have found (N_1, M_0, M_1) as desired. ■

Proof of Lemma 40. With the initial point $z \in \{\bar{Z} - 1, \bar{Z}, \ldots\}$, the Markov chain Z_t is contained in the region $\{\bar{Z} - 1, \bar{Z}, \ldots\}$. In particular, if $\bar{Z}$ is finite, the state $\bar{Z} - 1$ is a reflecting boundary. Invoking Lemma 4 with $S = \{\bar{Z} - 1, \bar{Z}, \ldots\}$, it suffices to solve the difference equation $\tilde{\mathbb{E}}_1^z v(Z_2) - v(z) = \frac{\tilde{l}(z)}{4-2c}$ for all $z \in S$. The remainder of the proof verifies that (B.32) provides such a solution with the desired monotonicity and growth properties. We complete the rest of the proof in three steps.

Step 1. We claim that when $v(\cdot)$ is as in (B.32), $\tilde{\mathbb{E}}_1^z v(Z_2) - v(z) = \frac{\tilde{l}(z)}{4-2c}$ for all $z \in \{\bar{Z}-1, \bar{Z}, \ldots\}$. Observe that $v(\cdot)$ solves the difference equation (B.9) with $\hat{z} = \bar{Z}$, $\delta = \frac{\rho(\bar{Z}-1)}{2}$,

$x(z) = \rho^2(z)$, and $p(z) = \frac{1}{2} + \rho(z)$. That is,

$$\begin{cases} v(\bar{Z}-1) = 0,\, v(\bar{Z}) = \frac{\rho(\bar{Z}-1)}{2}, \\ \left[\frac{1}{2}+\rho(z)\right] v(z+1) + \left[\frac{1}{2}-\rho(z)\right] v(z-1) - v(z) = \rho^2(z) \;\text{ for all } z \geq \bar{Z}. \end{cases}$$

As a result, for all $z \in \{\bar{Z}-1, \bar{Z}, \ldots\}$,

$$\begin{aligned} \tilde{\mathbb{E}}_1^z[v(Z_2)] - v(z) &= \begin{cases} v(z+1) - v(z) & \text{if } z = \bar{Z}-1, \\ \left[\frac{1}{2}+\rho(z)\right] v(z+1) + \left[\frac{1}{2}-\rho(z)\right] v(z-1) - v(z) & \text{if } z \geq \bar{Z}, \end{cases} \\ &= \begin{cases} \frac{\rho(\bar{Z}-1)}{2} & \text{if } z = \bar{Z}-1, \\ \rho^2(z) & \text{if } z \geq \bar{Z}, \end{cases} \\ &\overset{\text{(B.29)}}{=} \frac{\tilde{l}(z)}{4-2c}. \end{aligned}$$

Step 2. We claim that with $v(z)$ increases in z. This follows from the fact that $v(z)$ defined in (B.32) is a partial sum of nonnegative terms.

Step 3. We claim that $v(z) \to \infty$ as $z \to \infty$ with a rate such that

$$\limsup_{z\to\infty} \left\{ \frac{v(z)}{\sum_{k=1}^z \rho(k)} \right\} < \infty.$$

To prove this claim, we analyze $v(z)$ by introducing two auxiliary sequences. First, let $a_k := \sum_{n=k}^{\infty} \prod_{m=k}^{n} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}$ for $k \in \{\bar{Z}-1, \bar{Z}, \ldots\}$. Because ρ is slowly vanishing (see Definition 4), we invoke Lemma 37 with $x_m = \rho(m)$, and deduce that $a_1 < \infty$. Moreover, since a_k satisfies the inductive relation $a_k = \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}(1 + a_{k+1})$, we also deduce that $a_k < \infty$ for all $k \in \mathbb{Z}$. We introduce our second auxiliary sequence as follows: for $z \geq \bar{Z} \vee 2$, let

$$b_z := \sum_{n=\bar{Z}}^{z-1} \sum_{k=\bar{Z}}^{n} \frac{\rho^2(k)}{\frac{1}{2}-\rho(k)} \prod_{m=k}^{n} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} \overset{(a)}{=} \sum_{\check{n}=1}^{z-\bar{Z}} \sum_{\check{k}=1}^{\check{n}} \frac{\rho^2(\check{k}+\bar{Z}-1)}{\frac{1}{2}-\rho(\check{k}+\bar{Z}-1)} \prod_{\check{m}=\check{k}}^{\check{n}} \frac{\frac{1}{2}-\rho(\check{m}+\bar{Z}-1)}{\frac{1}{2}+\rho(\check{m}+\bar{Z}-1)},$$

where (a) holds by the following change of variables: $\check{k} := k - \bar{Z} + 1, \check{n} := n - \bar{Z} + 1$ and $m' := m - \bar{Z} + 1$. We claim that $\{b_z\}$ is a sequence that diverges at a rate satisfying $b_z = O(\rho(\bar{Z}) + \ldots + \rho(z-1))$. Invoking the first part of Lemma 39 with $x_m = \rho(m + \bar{Z} - 1)$, we know that $b_z \uparrow \infty$ as $z \uparrow \infty$. Moreover, since ρ is regular and slowly vanishing, the

following three conditions hold:

1. Since ρ is slowly vanishing,

$$\begin{aligned}&\liminf_{m\to\infty}\{m\rho(m+\bar{Z}-1)\}\\ &\geq\Big(\liminf_{m\to\infty}\{\tfrac{m}{m+\bar{Z}-1}\}\Big)\Big(\liminf_{m\to\infty}\{(m+\bar{Z}-1)\rho(m+\bar{Z}-1)\}\Big)>\frac{1}{4}.\end{aligned}$$

2. Since ρ is regular,

$$\begin{aligned}&\lim_{m\to\infty}\big\{\big(\frac{\rho(m+\bar{Z}-1)}{\rho(m+\bar{Z})}-1\big)m\big\}\\ &=\bigg(\lim_{m\to\infty}\big\{\big(\frac{\rho(m+\bar{Z}-1)}{\rho(m+\bar{Z})}-1\big)(m-\bar{Z}+1)\big\}\bigg)\cdot\bigg(\lim_{m\to\infty}\frac{m}{m-\bar{Z}+1}\bigg)\end{aligned}$$

   exists.

3. Since ρ is slowly vanishing,

$$\lim_{m\to\infty}\{\rho(m+\bar{Z}-1)\}=\lim_{m\to\infty}\{\rho(m)\}=0.$$

Therefore, we deduce from the second part of Lemma 39 that $b_z=O(\rho(\bar{Z})+\ldots+\rho(z-1))$. With the introduction of $\{a_k\}$ and $\{b_z\}$, let us evaluate $v(z)$ below for every $z\geq\bar{Z}\vee 2$:

$$\begin{aligned}v(z)=&\bigg(1+\sum_{n=\bar{Z}}^{z-1}\prod_{m=\bar{Z}}^{n}\tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\bigg)\tfrac{\rho(\bar{Z}-1)}{2}+\sum_{n=\bar{Z}}^{z-1}\sum_{k=\bar{Z}}^{n}\tfrac{\rho^2(k)}{\frac{1}{2}-\rho(k)}\prod_{m=k}^{n}\tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\\ \leq&\bigg(1+\sum_{n=\bar{Z}}^{\infty}\prod_{m=\bar{Z}}^{n}\tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\bigg)\tfrac{\rho(\bar{Z}-1)}{2}+\sum_{n=\bar{Z}}^{z-1}\sum_{k=\bar{Z}}^{n}\tfrac{\rho^2(k)}{\frac{1}{2}-\rho(k)}\prod_{m=k}^{n}\tfrac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\\ =&(1+a_{\bar{Z}})\tfrac{\rho(\bar{Z}-1)}{2}+b_z.\end{aligned}$$

Now, note that $\bar{Z}$ and $a_{\bar{Z}}$ are finite constants independent of z, and $\{b_z\}$ is a sequence that diverges at a rate satisfying $b_z=O(\rho(\bar{Z})+\ldots+\rho(z-1))$. This completes the proof. ■

Proof of Lemma 41. Because ρ is slowly vanishing, it is also strictly upper bounded. Thus, by Lemma 27, there exists $\varepsilon_0>0$ such that $\rho(z)\geq\varepsilon_0$ for all $z\in\mathbb{N}_-$. Let

$$B:=\tfrac{\rho^2(0)}{\frac{1}{2}+\rho(0)}+\tfrac{\frac{1}{2}-\rho(0)}{\frac{1}{2}+\rho(0)}\tfrac{\frac{1}{2}+\varepsilon_0}{8\varepsilon_0}\quad\text{and}\quad\tilde{B}:=-\tfrac{\frac{1}{2}+\varepsilon_0}{8\varepsilon_0}.\tag{B.37}$$

To complete the proof, we invoke Lemma 4 with $S = \mathbb{Z}$, and deduce that it suffices solve the difference equation $\tilde{\mathbb{E}}_1^z \tilde{v}(Z_2) - \tilde{v}(z) = \frac{\tilde{l}(z)}{4-2c}$ for $z \in \mathbb{Z}$. Due to the dynamics of $\{Z_t\}$ and the definition of $\tilde{l}(\cdot)$ in (B.29), this simplifies to $\left[\frac{1}{2} - \rho(z)\right] \tilde{v}(z-1) + \left[\frac{1}{2} + \rho(z)\right] \tilde{v}(z+1) - \tilde{v}(z) = \rho^2(z)$ for $z \in \mathbb{Z}$. The rest of the proof confirms that (B.34) presents such a solution with the desired monotonicity and growth properties. We complete the remainder of the proof in three steps.

Step 1. We claim that when $\tilde{v}(\cdot)$ is as in (B.34), $\left[\frac{1}{2} - \rho(z)\right] \tilde{v}(z-1) + \left[\frac{1}{2} + \rho(z)\right] \tilde{v}(z+1) - \tilde{v}(z) = \rho^2(z)$ for all $z \in \mathbb{Z}$. We can view this function as a concatenation of two one-sided functions: one defined on $\mathbb{N}$ and the other one defined on $\mathbb{Z}_- \cup \{0\}$. For the first piece (defined on $\mathbb{N}$), we invoke Lemma 28 and observe that $\tilde{v}(\cdot)$ restricted to $\mathbb{N}$ satisfies the difference equation (B.9) with $\hat{z} = 1$, $\delta = B$, $x(z) = \rho^2(z)$, and $p(z) = \frac{1}{2} + \rho(z)$. That is,

$$\begin{cases} \tilde{v}(0) = 0,\ \tilde{v}(1) = B, \\ \left[\frac{1}{2} + \rho(z)\right] v(z+1) + \left[\frac{1}{2} - \rho(z)\right] v(z-1) - v(z) = \rho^2(z) \ \text{ for all } z \geq 1. \end{cases}$$

For the second piece (defined on $\mathbb{Z}_- \cup \{0\}$), let us introduce an auxiliary sequence $\{y(z), z \in \mathbb{N}\}$ as follows:

$$y(z) := \begin{cases} 0 & \text{if } z = 0, \\ \left(1 + \sum_{n=1}^{z-1} \prod_{m=1}^{n} \frac{\frac{1}{2}+\rho(-m)}{\frac{1}{2}-\rho(-m)}\right) \tilde{B} + \sum_{n=1}^{z-1} \sum_{k=1}^{n} \frac{\rho^2(-k)}{\frac{1}{2}+\rho(-k)} \prod_{m=n}^{k} \frac{\frac{1}{2}+\rho(-m)}{\frac{1}{2}-\rho(-m)} & \text{if } z \geq 1. \end{cases}$$

First, we verify that $y(\cdot)$ is a "reflected" version of $\tilde{v}(\cdot)$ restricted to $\mathbb{Z}_- \cup \{0\}$. That is, for all $z \in \mathbb{Z}_-$,

$$\begin{aligned} \tilde{v}(z) &= \left(1 + \sum_{n=z+1}^{-1} \prod_{m=n}^{-1} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)}\right) \tilde{B} + \sum_{n=z+1}^{-1} \sum_{k=n}^{-1} \frac{\rho^2(k)}{\frac{1}{2}+\rho(k)} \prod_{m=n}^{k} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} \\ &\overset{(a)}{=} \left(1 + \sum_{n=1}^{\check{z}-1} \prod_{\check{m}=1}^{\check{n}} \frac{\frac{1}{2}+\rho(-\check{m})}{\frac{1}{2}-\rho(-\check{m})}\right) \tilde{B} + \sum_{\check{n}=1}^{\check{z}-1} \sum_{\check{k}=1}^{\check{n}} \frac{\rho^2(-\check{k})}{\frac{1}{2}+\rho(-\check{k})} \prod_{\check{m}=\check{n}}^{\check{k}} \frac{\frac{1}{2}+\rho(-\check{m})}{\frac{1}{2}-\rho(-\check{m})} \\ &= y(-z) \end{aligned}$$

where (a) holds by the following change of variables: $\check{m} := -m, \check{n} := -n, \check{z} := -z$. Second,

note that $y(\cdot)$ solves the difference equation (B.9) with $\hat{z} = 1$, $\delta = \tilde{B}$, $x(z) = \rho^2(-z)$, and $p(z) = \frac{1}{2} - \rho(-z)$; i.e.,

$$\begin{cases} y(0) = 0,\, y(1) = \tilde{B}, \\ \left[\frac{1}{2} - \rho(-z)\right] y(z+1) + \left[\frac{1}{2} + \rho(-z)\right] y(z-1) - y(z) = \rho^2(-z) \ \text{ for all } z \geq 1. \end{cases}$$

Because $\tilde{v}(z) = y(-z)$ for all $z \in \mathbb{Z}_-$, we have

$$\begin{cases} \tilde{v}(0) = 0,\, \tilde{v}(-1) = \tilde{B}, \\ \left[\frac{1}{2} - \rho(z)\right] \tilde{v}(z-1) + \left[\frac{1}{2} + \rho(z)\right] \tilde{v}(z+1) - \tilde{v}(z) = \rho^2(z) \ \text{ for all } z \leq -1. \end{cases}$$

Lastly, it suffices to verify that B and $\tilde{B}$ are such that $\left[\frac{1}{2} - \rho(0)\right] \tilde{v}(-1) + \left[\frac{1}{2} + \rho(0)\right] \tilde{v}(1) - \tilde{v}(0) = \rho^2(0)$. In fact,

$$\begin{aligned} &\left[\tfrac{1}{2} - \rho(0)\right] \tilde{v}(-1) + \left[\tfrac{1}{2} + \rho(0)\right] \tilde{v}(1) - \tilde{v}(0) \\ &\qquad = \left[\tfrac{1}{2} - \rho(0)\right] \tilde{B} + \left[\tfrac{1}{2} + \rho(0)\right] B \\ &\qquad = -\left[\tfrac{1}{2} - \rho(0)\right] \tfrac{\frac{1}{2}+\varepsilon_0}{8\varepsilon_0} + \left[\tfrac{1}{2} + \rho(0)\right] \left[\tfrac{\rho^2(0)}{\frac{1}{2}+\rho(0)} + \tfrac{\frac{1}{2}-\rho(0)}{\frac{1}{2}+\rho(0)} \tfrac{\frac{1}{2}+\varepsilon_0}{8\varepsilon_0}\right] \quad \text{[by (B.37)]} \\ &\qquad = \rho^2(0). \end{aligned}$$

This completes the proof of our claim in Step 1.

Step 2. We claim that $\tilde{v}(z)$ increases in z. To prove this claim, we directly verify that $\tilde{v}(\cdot)$ increases on $\mathbb{N}$, as $\tilde{v}(z)$ is a partial sum of nonnegative terms for $z \in \mathbb{N}$ according to

(B.34). For all $z \in \mathbb{Z}_-$, note that

$$\begin{aligned}
&\frac{\tilde{v}(z) - \tilde{v}(z+1)}{\prod_{m=z+1}^{-1} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)}} \\
&= \frac{1}{\prod_{m=z+1}^{-1} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)}} \left[\left(\prod_{m=z+1}^{-1} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} \right) \tilde{B} + \sum_{k=z+1}^{-1} \frac{\rho^2(k)}{\frac{1}{2}+\rho(k)} \prod_{m=z+1}^{k} \frac{\frac{1}{2}+\rho(m)}{\frac{1}{2}-\rho(m)} \right] \\
&= \tilde{B} + \sum_{k=z+1}^{-1} \frac{\rho^2(k)}{\frac{1}{2}+\rho(k)} \prod_{m=k+1}^{-1} \frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} \\
&\overset{(b)}{\leq} \tilde{B} + \sum_{k=-\infty}^{-1} \frac{\rho^2(-\infty)}{\frac{1}{2}+\rho(-\infty)} \prod_{m=k+1}^{-1} \frac{\frac{1}{2}-\varepsilon_0}{\frac{1}{2}+\varepsilon_0} \\
&= \tilde{B} + \tfrac{\frac{1}{2}+\varepsilon_0}{8\varepsilon_0} \overset{(c)}{=} 0,
\end{aligned}$$

where: (b) follows because $\rho(z) \geq \varepsilon_0$ for all $z \in \mathbb{N}_-$ and the function $x \mapsto \frac{x^2}{\frac{1}{2}+x}$ increases in x when $x \geq 0$, and (c) follows because $\tilde{B} = -\frac{\frac{1}{2}+\varepsilon_0}{8\varepsilon_0}$ by definition; see (B.37).

Step 3. We claim that $\tilde{v}(z) \to \infty$ as $z \to \infty$ with a rate such that

$$\limsup_{z \to \infty} \left\{ \frac{\tilde{v}(z)}{\sum_{k=1}^{z} \rho(k)} \right\} < \infty.$$

To prove this claim, we first note that because ρ is regular and slowly vanishing, the following three conditions hold:

1. $\liminf_{m \to \infty} \{m\rho(m)\} > \frac{1}{4}$, because ρ is slowly vanishing.
2. $\lim_{m \to \infty} \left\{ \left(\frac{\rho(m)}{\rho(m+1)} - 1 \right) m \right\} \in [0, \infty]$ exists, because ρ is regular.
3. $\lim_{m \to \infty} \{\rho(m)\} = 0$, because ρ is slowly vanishing.

Hence, invoking Lemma 37 and Lemma 39 with $x_n = \rho(n)$, we have the following for $z \geq \mathbb{Z}_+$,

$$\begin{aligned}
&\sum_{n=1}^{z-1}\sum_{k=1}^{n}\frac{\rho^2(k)}{\frac{1}{2}-\rho(k)}\prod_{m=k}^{n}\frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)} \leq \tilde{v}(z)\\
&= \left(1+\sum_{n=1}^{z-1}\prod_{m=1}^{n}\frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\right)B + \sum_{n=1}^{z-1}\sum_{k=1}^{n}\frac{\rho^2(k)}{\frac{1}{2}-\rho(k)}\prod_{m=k}^{n}\frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\\
&\leq \left(1+\sum_{n=1}^{\infty}\prod_{m=1}^{n}\frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\right)B + M\sum_{n=1}^{z-1}\rho(n)\\
&\leq \left(1+\sum_{n=1}^{\infty}\prod_{m=1}^{n}\frac{\frac{1}{2}-\rho(m)}{\frac{1}{2}+\rho(m)}\right)B + M\sum_{n=1}^{z}\rho(n)\\
&\leq \check{C} + M\big(\rho(1)+\cdots+\rho(z)\big)
\end{aligned}$$

where $\check{C}, M$ are finite constants independent of z. This completes the proof. ■

REFERENCES

Agrawal, S., Avadhanula, V., Goyal, V., and Zeevi, A. (2017). Thompson sampling for the MNL-bandit. In *COLT*, pages 76–78.

Agrawal, S., Avadhanula, V., Goyal, V., and Zeevi, A. (2019). MNL-bandit: A dynamic learning approach to assortment selection. *Operations Research*, 67(5):1453–1485.

Agrawal, S., Delage, E., Peters, M., Wang, Z., and Ye, Y. (2011). A unified framework for dynamic prediction market design. *Operations Research*, 59(3):550–568.

Ailon, N. (2012). An active learning algorithm for ranking from pairwise preferences with an almost optimal query complexity. *J. Mach. Learn. Res.*, 13(Jan):137–164.

Ailon, N., Charikar, M., and Newman, A. (2005). Aggregating inconsistent information: Ranking and clustering. In *STOC*, pages 684–693.

Ailon, N., Charikar, M., and Newman, A. (2008). Aggregating inconsistent information: ranking and clustering. *Journal of the ACM*, 55(5):23.

Ali, A. and Meila, M. (2012). Experiments with Kemeny ranking: What works when? *Math. Social Sci.*, 64(1):28–40.

Allen, F. and Gale, D. (1992). Stock-price manipulation. *The Review of Financial Studies*, 5(3):503–529.

Alon, N. (2006). Ranking tournaments. *SIAM J. Discrete Math.*, 20(1):137–142.

Araman, V. and Caldentey, R. (2016). Crowdvoting the timing of new product introduction. *Available at SSRN: https://ssrn.com/abstract=2723515*.

Araman, V. F. and Caldentey, R. (2009). Dynamic pricing for nonperishable products with demand learning. *Operations Research*, 57(5):1169–1188.

Audibert, J.-Y. and Bubeck, S. (2010). Best arm identification in multi-armed bandits. In *COLT*, pages 41–53.

Back, K. (1992). Insider trading in continuous time. *The Review of Financial Studies*, 5(3):387–409.

Back, K. and Baruch, S. (2004). Information in securities markets: Kyle meets Glosten and Milgrom. *Econometrica*, 72(2):433–465.

Ban, A. (2018). Strategy-proof incentives for predictions. In *International Conference on Web and Internet Economics*, pages 51–65. Springer.

Ban, G.-Y. and Keskin, N. B. (2018). Personalized dynamic pricing with machine learning: High dimensional features and heterogeneous elasticity. Available at SSRN: https://ssrn.com/abstract=2972985.

Betabrand (2018). `https://www.betabrand.com`. Accessed: 2018-11-17.

Bossaerts, P., Fine, L., and Ledyard, J. (2002). Inducing liquidity in thin financial markets through combined-value trading mechanisms. *European Economic Review*, 46(9):1671–1695.

Brabham, C. D. (2010). Moving the crowd at Threadless. *Inf. Commun. Soc.*, 13:1122–1145.

Braverman, M. and Mossel, E. (2008). Noisy sorting without resampling. In *SODA*, pages 268–276.

Braverman, M. and Mossel, E. (2009). Sorting from noisy information. *arXiv preprint arXiv:0910.1191*.

Bromwich, T. J. I. (1908). *An Introduction to the Theory of Infinite Series*. Macmillan and Company, Ltd.

Bubeck, S., Munos, R., and Stoltz, G. (2011). Pure exploration in finitely-armed and continuous-armed bandits. *Theor. Comput. Sci.*, 412(19):1832–1852.

Caldentey, R. and Stacchetti, E. (2010). Insider trading with a random deadline. *Econometrica*, 78(1):245–283.

Camerer, C. F. (1989). Does the basketball market believe in the "hot hand"? *The American Economic Review*, 79(5):1257–1261.

Caragiannis, I., Procaccia, A. D., and Shah, N. (2013). When do noisy votes reveal the truth? In *EC*, pages 143–160.

Caro, F. and Gallien, J. (2007). Dynamic assortment with demand learning for seasonal consumer goods. *Manag. Sci.*, 53(2):276–292.

Charbit, P., Thomassé, S., and Yeo, A. (2007). The minimum feedback arc set problem is NP-hard for tournaments. *Comb. Probab. Comput.*, 16(1):1–4.

Charon, I. and Hudry, O. (2010). An updated survey on the linear ordering problem for weighted or unweighted tournaments. *Ann. Oper. Res.*, 175(1):107–158.

Chen, B., Chao, X., and Ahn, H.-S. (2015). Coordinating pricing and inventory replenishment with nonparametric demand learning. Available at SSRN: https://ssrn.com/abstract=2694633.

Chen, X., Li, Y., and Mao, J. (2018). A nearly instance optimal algorithm for top-k ranking under the multinomial logit model. In *SODA*, pages 2504–2522.

Chen, X. and Wang, Y. (2017). A note on tight lower bound for MNL-bandit assortment selection models. *arXiv preprint arXiv:1709.06109*.

Chen, X. and Wang, Z. (2016). Bayesian dynamic learning and pricing with strategic customers. Available at SSRN: https://ssrn.com/abstract=2715730.

Chen, Y. and Pennock, D. M. (2007). A utility framework for bounded-loss market makers. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence*, pages 49–56. AUAI Press.

Chen, Y. and Vaughan, J. W. (2010). A new understanding of prediction markets via no-regret learning. In *Proceedings of the 2010 ACM Conference on Economics and Computation (EC '10)*, pages 189–198. ACM.

Chernoff, H. (1959). Sequential design of experiments. *Ann. of Math. Stat.*, 30(3):755–770.

Chernoff, H. (1972). *Sequential Analysis and Optimal Design*. SIAM.

Chung, F. and Lu, L. (2006). Concentration inequalities and martingale inequalities: A survey. *Internet Mathematics*, 3(1):79–127.

Ciocan, D. F. and Farias, V. F. (2014). Fast demand learning for display advertising revenue management. Working paper, MIT.

CNN (2018). Supreme court lets states legalize sports gambling. `https://www.cnn.com/2018/05/14/politics/sports-betting-ncaa-supreme-court/index.html`. Accessed: Oct 30, 2018.

Conitzer, V., Davenport, A., and Kalagnanam, J. (2006). Improved bounds for computing Kemeny rankings. In *AAAI*, pages 620–626.

Dantzig, G. (1963). *Linear Programming and Extensions*. Princeton Univ. Press, Princeton, NJ.

Davenport, A. and Kalagnanam, J. (2004). A computational study of the Kemeny rule for preference aggregation. In *AAAI*, pages 697–702.

den Boer, A. and Keskin, N. B. (2017). Dynamic pricing with demand learning and reference effects. Available at SSRN: https://ssrn.com/abstract=3092745.

den Boer, A. and Keskin, N. B. (2019). Discontinuous demand functions: Estimation and pricing. Forthcoming in *Management Science*. Available at SSRN: https://ssrn.com/abstract=3003984.

Désir, A., Goyal, V., Jagabathula, S., and Segev, D. (2018). Mallows-smoothed distribution over rankings approach for modeling choice. *Available at SSRN 3172997*.

Devanur, N. R., Peres, Y., and Sivan, B. (2014). Perfect Bayesian equilibria in repeated sales. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2014)*, pages 983–1002.

Draglia, V., Tartakovsky, A. G., and Veeravalli, V. V. (1999). Multihypothesis sequential probability ratio tests. I. Asymptotic optimality. *IEEE Trans. Inform. Theory*, 45(7):2448–2461.

Durrett, R. (2019). *Probability: theory and examples*, volume 49. Cambridge university press.

Eden, S. (2013). Meet the world's top NBA gambler. `http://www.espn.com/blog/playbook/dollars/post/_/id/2935/meet-the-worlds-top-nba-gambler`. Accessed: Oct 30, 2018.

Falahatgar, M., Orlitsky, A., Pichapati, V., and Suresh, A. T. (2017). Maximum selection and ranking under noisy comparisons. In *ICML*, pages 1088–1096.

Ferreira, K., Simchi-Levi, D., and Wang, H. (2017). Online network revenue management using Thompson sampling. Available at SSRN: https://ssrn.com/abstract=2588730.

Fomin, F. V., Lokshtanov, D., Raman, V., and Saurabh, S. (2010). Fast local search algorithm for weighted feedback arc set in tournaments. In *AAAI*, pages 65–70.

Freeman, R. and Pennock, D. M. (2018). An axiomatic view of the parimutuel consensus wagering mechanism. In *Proceedings of the 17th International Conference on Au-*

tonomous Agents and MultiAgent Systems, pages 1936–1938. International Foundation for Autonomous Agents and Multiagent Systems.

Freeman, R., Pennock, D. M., and Wortman Vaughan, J. (2017). The double clinching auction for wagering. In *Proceedings of the 2017 ACM Conference on Economics and Computation (EC '17)*, pages 43–60. ACM.

Gabillon, V., Ghavamzadeh, M., and Lazaric, A. (2012). Best arm identification: A unified approach to fixed budget and fixed confidence. In *NIPS*, pages 3212–3220.

Gandar, J. M., Dare, W. H., Brown, C. R., and Zuber, R. A. (1998). Informed traders and price variations in the betting market for professional basketball games. *The Journal of Finance*, 53(1):385–401.

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *COLT*, pages 998–1027.

Glosten, L. R. and Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100.

Golec, J. and Tamarkin, M. (1991). The degree of inefficiency in the football betting market: Statistical tests. *Journal of Financial Economics*, 30(2):311 – 323.

Gray, P. K. and Gray, S. F. (1997). Testing market efficiency: Evidence from the NFL sports betting market. *The Journal of Finance*, 52(4):1725–1737.

Grötschel, M., Jünger, M., and Reinelt, G. (1984). A cutting plane algorithm for the linear ordering problem. *Oper. Res.*, 32(6):1195–1220.

Hanson, R. (2003). Combinatorial information market design. *Information Systems Frontiers*, 5(1):107–119.

Hanson, R. (2007). Logarithmic market scoring rules for modular combinatorial information aggregation. *Journal of Prediction Markets*, 1(1):3–15.

Harrison, J. M., Keskin, N. B., and Zeevi, A. (2012). Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science*, 58(3):570–586.

Hausch, D. B. and Ziemba, W. T., editors (2008). *Handbook of Sports and Lottery Markets*. Handbooks in Finance. North Holland, 1st edition.

Heckel, R., Shah, N. B., Ramchandran, K., Wainwright, M. J., et al. (2019). Active ranking from pairwise comparisons and when parametric assumptions do not help. *Ann. of Stat.*, 47(6):3099–3126.

Huang, Y., Singh, V. P., and Srinivasan, K. (2014). Crowdsourcing new product ideas under consumer learning. *Manag. Sci.*, 60.

Huang, Z., Liu, J., and Wang, X. (2018). Learning optimal reserve price against non-myopic bidders. *arXiv preprint arXiv:1804.11060*.

Huddart, S., Hughes, J. S., and Levine, C. B. (2001). Public disclosure and dissimulation of insider trades. *Econometrica*, 69(3):665–681.

Jamieson, K. G., Katariya, S., Deshpande, A., and Nowak, R. D. (2015). Sparse dueling bandits. In *AISTATS*, pages 416–424.

Jiang, X., Lim, L.-H., Yao, Y., and Ye, Y. (2011). Statistical ranking and combinatorial Hodge theory. *Math. Program.*, 127(1):203–244.

Jun, K.-S., Li, L., Ma, Y., and Zhu, J. (2018). Adversarial attacks on stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 3640–3649.

Kanoria, Y. and Nazerzadeh, H. (2019). Dynamic reserve prices for repeated auctions: Learning from bids. Available at SSRN: https://ssrn.com/abstract=2444495.

Kaufmann, E., Cappé, O., and Garivier, A. (2016). On the complexity of best arm identification in multi-armed bandit models. *J. Mach. Learn. Res.*, 17(1):1–42.

Kenyon-Mathieu, C. and Schudy, W. (2007). How to rank with few errors. In *STOC*, pages 95–103. ACM.

Keskin, N. B. and Birge, J. R. (2019). Dynamic selling mechanisms for product differentiation and learning. *Operations Research*, 67(4):1069–1089.

Keskin, N. B. and Zeevi, A. (2014). Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5):1142–1167.

Keskin, N. B. and Zeevi, A. (2016). Chasing demand: Learning and earning in a changing environment. *Mathematics of Operations Research*, 42(2):277–307.

Keskin, N. B. and Zeevi, A. (2018). On incomplete learning and certainty-equivalence control. *Operations Research*, 66(4):1136–1167.

King, A. and Lakhani, K. R. (2013). Using open innovation to identify the best ideas. *MIT Sloan Man. Rev.*, 55(1):41.

Krylov, N. V. (2002). *Introduction to the Theory of Random Processes*, volume 43. American Mathematical Soc.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335.

Lacey, N. J. (1990). An estimation of market efficiency in the nfl point spread betting market. *Applied Economics*, 22(1):117–129.

LEGO (2018). `https://https://ideas.lego.com`. Accessed: 2018-11-17.

Levin, J. and Nalebuff, B. (1995). An introduction to vote-counting schemes. *J. Econ. Perspect.*, 9(1):3–26.

Levina, T., Levin, Y., McGill, J., and Nediak, M. (2009). Dynamic pricing with online learning and strategic consumers: an application of the aggregating algorithm. *Operations Research*, 57(2):327–341.

Levitt, S. (2004). Why are gambling markets organised so differently from financial markets? *The Economic Journal*, 114:223–246.

Li, X., Liu, J., and Ying, Z. (2014). Generalized sequential probability ratio test for separate families of hypotheses. *Sequential Analysis*, 33(4):539–563.

Lin, J.-C. and Howe, J. S. (1990). Insider trading in the OTC market. *The Journal of Finance*, 45(4):1273–1284.

Lykouris, T., Mirrokni, V., and Paes Leme, R. (2018). Stochastic bandits robust to adversarial corruptions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 114–122.

Marinesi, S. and Girotra, K. (2012). Information acquisition through customer voting systems. *Available at SSRN: https://ssrn.com/abstract=2191940*.

McLennan, A. (1984). Price dispersion and incomplete learning in the long run. *Journal of Economic Dynamics and Control*, 7(3):331–347.

Mitchell, J. E. and Borchers, B. (1996). Solving real-world linear ordering problems using a primal-dual interior point cutting plane method. *Ann. Oper. Res.*, 62(1):253–276.

Moskowitz, T. J. (2015). Asset pricing and sports betting. Available at SSRN: https://ssrn.com/abstract=2635517.

Naghshvar, M., Javidi, T., et al. (2013). Active sequential hypothesis testing. *Ann. of Stat.*, 41(6):2703–2738.

NGISC (1999). National gambling impact study commission final report. https://govinfo.library.unt.edu/ngisc/reports/fullrpt.html. Accessed: Oct 30, 2018.

NYTimes (2019). New jersey embraces sports gambling, and a billion-dollar business is born. https://www.nytimes.com/2019/01/29/sports/sports-gambling-new-jersey.html. Accessed: Feb 18, 2019.

OE (2017). Economic impact of legalized sports betting. Technical report, Oxford Economics, Wayne, PA. Accessed: Oct 30, 2018.

Øksendal, B. (2003). *Stochastic Differential Equations*. Springer.

Ostrovsky, M. (2012). Information aggregation in dynamic markets with strategic traders. *Econometrica*, 80(6):2595–2647.

Paul, R. J. and Weinbach, A. P. (2012). Does sportsbook.com set pointspreads to maximize profits? *The Journal of Prediction Markets*, 1(3):209–218.

Pee, L. G. (2016). Customer co-creation in B2C e-commerce: Does it lead to better new products? *Elec. Commerce Res.*, 16(2):217–243.

PREFLIB (2019). http://www.preflib.org. Accessed: 2019-11-21.

Raykar, C. V., Yu, S., Zhao, H. L., Valadez, H. G., Florin, C., Bogoni, L., and Moy, L. (2010). Learning from crowds. *J. Mach. Learn. Res.*, 11:1297–1322.

Routledge, B. R. (1999). Adaptive learning in financial markets. *The Review of Financial Studies*, 12(5):1165–1202.

Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill.

Rusmevichientong, P., Shen, Z.-J. M., and Shmoys, D. B. (2010). Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Oper. Res.*, 58(6):1666–1680.

Russo, D. (2016). Simple Bayesian algorithms for best arm identification. In *COLT*, pages 1417–1418.

Sauer, R. D. (2005). The state of research on markets for sports betting and suggested future directions. *Journal of Economics and Finance*, 29(3):416–426.

Sauré, D. and Zeevi, A. (2013a). Optimal dynamic assortment planning with demand learning. *Manuf. Serv. Oper. Manag.*, 15(3):387–404.

Sauré, D. and Zeevi, A. (2013b). Optimal dynamic assortment planning with demand learning. *Manuf. Serv. Oper. Manag.*, 15(3):387–404.

Schalekamp, F. and Zuylen, A. v. (2009). Rank aggregation: Together we're strong. In *ALENEX*, pages 38–51. SIAM.

Schneider, J. and Hall, J. (2011). Why most product launches fail. *Harvard Bus. Rev.*, (April):21–23.

Schwartz, D. G. (2018). Nevada sports betting totals: 1984-2017. `https://gaming.unlv.edu/reports/NV_sportsbetting.pdf`. Accessed: Oct 30, 2018.

Shah, N. B. and Wainwright, M. J. (2017). Simple, robust and optimal ranking from pairwise comparisons. *J. Mach. Learn. Res.*, 18(1):7246–7283.

Shin, D. and Zeevi, A. (2017). Dynamic pricing and learning with online product reviews. Working Paper, Columbia University.

Slivkins, A. (2019). Introduction to multi-armed bandits. *Foundations and Trends in Machine Learning*, 12(1-2):1–286.

Statista (2018). Sports betting - statistics & facts. `https://www.statista.com/topics/1740/sports-betting`. Accessed: Oct 30, 2018.

Stern, H. (1991). On the probability of winning a football game. *The American Statistician*, 45(3):179–183.

Stern, H. S. (1994). A brownian motion model for the progress of sports scores. *Journal of the American Statistical Association*, 89(427):1128–1134.

Szörényi, B., Busa-Fekete, R., Paul, A., and Hüllermeier, E. (2015). Online rank elicitation for Plackett-Luce: A dueling bandits approach. In *NeurIPS*, pages 604–612.

Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations Research*, 26(2):305–321.

Ulu, C., Honhon, D., and Alptekinoglu, A. (2012). Learning consumer tastes through dynamic assortments. *Oper. Res.*, 60(4):833–849.

Van Zuylen, A. and Williamson, D. P. (2009). Deterministic pivoting algorithms for constrained ranking and clustering problems. *Math. Oper. Res.*, 34(3):594–620.

Vinayak, R. K. and Hassibi, B. (2016). Crowdsourced clustering: Querying edges vs triangles. In *Advances in Neural Information Processing Systems*, pages 1316–1324.

Wald, A. (1973). *Sequential Analysis*. Courier Corporation.

Wauthier, F. L., Jordan, M. I., and Jojic, N. (2013). Efficient ranking from pairwise comparisons. In *ICML*, pages III–109–117.

Wolfers, J. (2006). Point shaving: Corruption in NCAA basketball. *The American Economic Review*, 96(2):279–283.

Wolfers, J. and Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126.

Wolfers, J. and Zitzewitz, E. (2006). Five open questions about prediction markets. Technical report, National Bureau of Economic Research.

Young, H. P. (1988). Condorcet's theory of voting. *Am. Political Sci. Rev.*, 82(4):1231–1244.

Zuber, R. A., Gandar, J. M., and Bowers, B. D. (1985). Beating the spread: Testing the efficiency of the gambling market for national football league games. *Journal of Political Economy*, 93(4):800–806.