Policy Learning with Competing Agents

Roshni Sahoo
rsahoo@stanford.edu

Stefan Wager
swager@stanford.edu

Stanford University

Abstract

Decision makers often aim to learn a treatment assignment policy under a capacity constraint on the number of agents that they can treat. When agents can respond strategically to such policies, competition arises, complicating the estimation of the effect of the policy. In this paper, we study capacity-constrained treatment assignment in the presence of such interference. We consider a dynamic model where heterogeneous agents myopically best respond to the previous treatment assignment policy. When the number of agents is large but finite, we show that the threshold for receiving treatment under a given policy converges to the policy's mean-field equilibrium threshold. Based on this result, we develop a consistent estimator for the policy effect and demonstrate in simulations that it can be used for learning optimal capacity-constrained policies in the presence of strategic behavior.

1 Introduction

In policy learning, a decision maker aims to map observed individual characteristics to treatment assignments [Bhattacharya and Dupas, 2012, Kitagawa and Tetenov, 2018, Manski, 2004]. For example, a school must decide which applicants to admit and an employer must decide which candidates should be extended offers. The observed data is typically assumed to be exogenous to the treatment assignment policy. However, when human agents being considered for the treatment have knowledge of the assignment policy, the observed data is not exogenous because agents may change their behavior in response to the policy.

A growing body of work focuses on policy learning in the presence of strategic human behavior [Björkegren et al., 2020, Frankel and Kartik, 2019a, Munro, 2020]. In these works, an agent's treatment assignment only depends on their own strategic behavior and is unaffected by the behavior of others in the population. This setup implicitly assumes that the decision maker does not have a capacity constraint on the number of agents they can treat.

However, in applications such as college admissions and job hiring, strategic behavior arises while the decision maker also has a capacity constraint on the number of agents they can treat. For example, students may enroll in test preparation services and take advanced courses to improve their chances of getting accepted to college [Bound et al., 2009], while a college can only accept a small fraction of the applicant pool. Similarly, job candidates may join intensive bootcamps to improve their career prospects [Thayer and Ko, 2017], while an employer has a fixed number of positions to fill. To enforce the capacity constraint, a decision maker uses a selection criteria, such as a machine learning model, to score agents and assigns treatments to agents who score above a threshold, given by a quantile of the score distribution [Bhattacharya and Dupas, 2012]. Competition arises because an agent's treatment assignment depends on how their score ranks relative to that of other agents.

In this work, we study the problem of capacity-constrained treatment assignment in the presence of strategic behavior. We frame the problem in a dynamic setting. At each time step t, agents report their covariates to the decision maker and the decision maker assigns treatments. Suppose a decision maker deploys a fixed selection criteria for all time. At time step $t+1$, agents react to the policy from time step t, which depends on the fixed selection criteria and the threshold for receiving treatment at time step t. To enforce the capacity constraint, the decision maker sets the threshold for receiving treatment at time step

Draft version: October 2022. Code available at https://github.com/roshni714/policy-learning-competing-agents.

$t+1$ to the appropriate quantile of the score distribution observed at time step $t+1$. So, the threshold for receiving treatment depends on agents' strategic behavior. At an equilibrium induced by a fixed selection criteria, the threshold for receiving treatment is fixed over time. The goal of the decision maker, and the main goal of this work, is to find a selection criteria that obtains low equilibrium policy loss, which is the policy loss obtained at the equilibrium induced by the selection criteria.

The goal of learning a policy that minimizes the equilibrium policy loss is motivated by prior works that estimate policy effects or treatment effects at equilibrium [Heckman et al., 1998, Munro et al., 2021, Wager and Xu, 2021]. Heckman et al. [1998] estimates the effect of a tuition subsidy program on college enrollment by accounting for the program's impact on the equilibrium college skill price. Munro et al. [2021] estimates the effect of a binary intervention in a marketplace setting by accounting for the impact of the intervention on the resulting supply-demand equilibrium. Wager and Xu [2021] estimates the effect of supply-side payments on a platform's utility in equilibrium.

We outline a dynamic model for capacity-constrained treatment assignment in the presence of generic strategic behavior and specify the form of strategic behavior we consider in Section 2. Key elements of our model include that agents are myopic, so the covariates they report to the decision maker at time step $t+1$ depend only on the state of the system in time step t. Also, drawing on the aggregative games literature [Acemoglu and Jensen, 2010, 2015, Corchón, 1994], we assume that agents respond to an aggregate of other agents' actions. In particular, at time step $t+1$, agents will react to the threshold for receiving treatment from time step t, which is an aggregate of agents' strategic behavior from time step t. Finally, based on Frankel and Kartik [2019a,b], we assume that agents are heterogenous in their raw covariates (covariates prior to modification) and in their ability to deviate from their raw covariates in their reported covariates.

In Section 3, we give conditions on our model that guarantee existence and uniqueness of equilibria in the mean-field regime, the limiting regime where at each time step, an infinite number of agents are considered for the treatment. Furthermore, we show that under additional conditions, the mean-field equilibrium arises via fixed-point iteration. In Section 4, we translate these results to the finite regime, where a finite number of agents, sampled i.i.d. at each time step, are considered for treatment. We show that as the number of agents grows large, the behavior of the system converges to the equilibrium of the mean-field model in a stochastic version of fixed-point iteration.

In Section 5, we propose a method for learning the selection criteria that minimizes the equilibrium policy loss. Based on Wager and Xu [2021], we take the approach of optimizing selection criteria via gradient descent. To this end, we develop a consistent estimator for the policy effect, the gradient of the equilibrium policy loss. To estimate the policy effect without disturbing the equilibrium, we follow the approach of Munro et al. [2021], Wager and Xu [2021]. We recover components of the policy effect by applying symmetric, mean-zero perturbations to the selection criteria and the threshold for receiving treatment for each unit and running regressions from the perturbations to outcomes of interest. In Section 6, through simulations, we validate that our policy effect estimator can be used to learn optimal capacity-constrained policies in the presence of strategic behavior.

1.1 Related Work

The problem of learning optimal treatment assignment policies has received attention in econometrics, statistics, and computer science [Athey et al., 2018, Bhattacharya and Dupas, 2012, Kallus and Zhou, 2021, Kitagawa and Tetenov, 2018, Manski, 2004]. Treatments can be discrete-valued (typically, binary) or continuous-valued, and the policy may be subject to budget or capacity constraints. Most related to our work, Bhattacharya and Dupas [2012] study the problem of optimal capacity-constrained treatment assignment, where the decision maker can only allocate treatments to $1-q$ proportion of the population, where $q \in (0,1)$. They show that the welfare-maximizing assignment policy is a threshold rule on the agents' scores, where agents who score above q-th quantile of the score distribution are allocated treatment. Our work differs from Bhattacharya and Dupas [2012] because we do not assume that the distribution of potential outcomes is exogenous to the treatment assignment policy.

Björkegren et al. [2020], Frankel and Kartik [2019a], Munro [2020] study policy learning in the presence of strategic behavior. Björkegren et al. [2020] proposes a structural model for manipulation, estimates the parameters of this model with data from a field experiment, and computes the optimal policy under the estimated model. Frankel and Kartik [2019a] demonstrates that optimal policies that account for strategic behavior will underweight manipulable data. Munro [2020] studies the optimal unconstrained assignment of binary-valued treatments in the presence of strategic behavior, without parametric assumptions on agent behavior. The main difference between our work and these previous works is that we account for the equilibrium effects of strategic behavior that arise via competition.

The area of strategic classification in computer science is also related to our work [Brückner et al., 2012, Dalvi et al., 2004, Dong et al., 2018, Hardt et al., 2016, Jagadeesan et al., 2021, Levanon and Rosenfeld, 2022]. These works propose models for the interaction between the classifier and the strategic agent and methods for training classifiers that are robust to gaming. In addition, other works in this area investigate how decision makers can design classifiers that incentivize agents to invest effort in improving, instead of gaming [Ahmadi et al., 2022, Kleinberg and Raghavan, 2020]. Nevertheless, the setting of strategic classification implicitly assumes that an agent's classification does not depend on the behavior of others in the population, limiting the applicability of these methods to our setting of policy learning with capacity constraints.

To the best of our knowledge, Liu et al. [2021] is the only existing work that studies capacity-constrained allocation in the presence of strategic behavior. Liu et al. [2021] introduces the problem of *strategic ranking*, where agents' rewards depend on their ranks after investing effort in modifying their covariates. They consider a setting where agents are heterogenous in their raw covariates but homogenous in their ability to modify their covariates. Under these assumptions, the authors find that agents' post-effort ranking preserves their original ranking by raw covariates and analyze the implications this has on decision maker, agent, and societal utility. Our work differs from Liu et al. [2021] because following Frankel and Kartik [2019a,b], we assume agents are heterogenous in both their raw covariates and ability to modify their reported covariates. When agents can be heterogenous across both dimensions, ranks are not necessarily preserved after the agents have modified their covariates. In our model, the selection criteria modulates how the equilibrium post-effort ranks change from the pre-effort ranks; in other words, strategic behavior changes who receives treatment, and thus fundamentally alters the nature of the resulting policy learning problem.

The problem of estimating the effect of an intervention in a marketplace setting is also relevant to our work. Marketplace interventions can impact the resulting supply-demand equilibrium, introducing interference and complicating estimation of the intervention's effect [Blake and Coey, 2014, Heckman et al., 1998]. We find that our setting yields analogous challenges to estimating the effect of a marketplace intervention because when agents are strategic and the decision maker is capacity-constrained, the selection criteria impacts the equilibrium threshold for receiving treatment. To estimate an intervention's effect without disturbing the market equilibrium, Munro et al. [2021], Wager and Xu [2021] propose a local experimentation scheme, motivated by mean-field modeling. Methodologically, we adapt their mean-field modeling and estimation strategies to estimate the effect of a policy in our setting.

Finally, we note that our dynamic model draws on game theory concepts, such as the myopic best response and dynamic aggregative games. Our assumption that agents are myopic, or will take decisions based on information from short time horizons, is a standard heuristic used in many previous works [Cournot, 1982, Kandori et al., 1993, Monderer and Shapley, 1996]. In addition, our assumption that agents account for the behavior of other agents through an aggregate quantity of their actions is a paradigm borrowed from aggregative games [Acemoglu and Jensen, 2010, 2015, Corchón, 1994]. Most related to our work, Acemoglu and Jensen [2015] considers a dynamic setting where the market aggregate at time step t is an aggregate function of all the agents' best responses from time step t, and an agent's best response at time step $t+1$ is selected from a constraint set determined by the market aggregate from time step t. Analogously, in our work, the "market aggregate" is the threshold for receiving treatment. The threshold for receiving treatment is a particular quantile of the agents' scores, so we can view it as a function of agents' reported covariates (agents' best responses). Furthermore, the covariates that agents report in time step $t+1$ depend on the value of the market aggregate, or the threshold for receiving treatment, in time step t.

2 Model

In this section, we first define a dynamic model for capacity-constrained treatment assignment in the presence of strategic behavior and define the decision maker's equilibrium policy loss in terms of this model. We then propose a model for agent behavior in terms of myopic utility maximization and provide conditions under which the resulting best response functions vary smoothly in problem parameters.

2.1 Dynamic Model

Our dynamic model is similar to the dynamic aggregative game model presented in Acemoglu and Jensen [2015] in that an aggregate quantity, the threshold for receiving treatment, depends on agents' actions, and the value of the aggregate quantity from a previous time step informs agent behavior in the subsequent time step.

Let $q \in (0,1)$. At each time step $t \in \{1,2,3\dots\}$, the decision maker assigns treatments to $1-q$ proportion of a target population based on observed covariates. The decision maker's selection criteria is a linear model $\boldsymbol{\beta} \in \mathcal{B}$ where $\mathcal{B} = \mathbb{S}^{d-1}$. The selection criteria is applied to observed covariates $\mathbf{x}$. The decision maker fixes the linear model $\boldsymbol{\beta}$ for all t, but they adjust the threshold for receiving treatment at each time step to ensure that the capacity constraint is satisfied. The decision maker's policy has the form of a threshold rule

$$\pi(\mathbf{x}; \boldsymbol{\beta}, s) = \mathbb{I}(\boldsymbol{\beta}^T\mathbf{x} \geq s), \tag{2.1}$$

where $\boldsymbol{\beta} \in \mathcal{B}$ and $s \in \mathbb{R}$. Suppose that each agent has a private type $\nu \sim F$. Let the policy at time step t be $\pi(\mathbf{x}; \boldsymbol{\beta}, s^t)$. At time step $t+1$, an agent with type ν will report covariates $\mathbf{x}(\boldsymbol{\beta}, s^t, \nu)$ to the decision maker, reacting strategically to the policy deployed in time step t; see Section 2.3 for a detailed specification for $\mathbf{x}(\boldsymbol{\beta}, s, \nu)$. Following Bhattacharya and Dupas [2012], we have that s^{t+1}, the threshold for receiving treatment at time step $t+1$, is equal to the q-th quantile of the marginal distribution of $\boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta}, s^t, \nu)$.

2.2 Population Equilibria and Policy Loss

The decision maker observes a loss $\ell(\pi, \nu)$ if they assign a treatment $\pi \in \{0,1\}$ to an agent with type ν. Note that an agent's type may not be directly observable, but the decision maker can still measure $\ell(\pi, \nu)$ for each agent. As an example, in college admissions, the decision maker may aim to admit students with high academic ability. Assuming that first-year GPA is a reasonable proxy for academic ability, the decision maker can set $\ell(1, \nu)$ to be the negation of an admitted agent's first-year GPA. Since the decision maker cannot assess the academic ability of students they did not admit, the decision maker incurs a loss $\ell(0, \nu) = 0$ on these students.

Given some specification of $\ell(\pi, \nu)$ that the decision maker can observe, we can define the population policy loss in the presence of strategic behavior and a capacity constraint. Let s denote the previous threshold for receiving treatment, which is the threshold that agents respond to, and let r denote the realized threshold for receiving treatment, which is the threshold that enforces the capacity constraint. The decision maker's population policy loss is given by

$$L(\boldsymbol{\beta}, s, r) = \mathbb{E}\left[\ell(\pi(\mathbf{x}(\boldsymbol{\beta}, s, \nu); \boldsymbol{\beta}, r), \nu)\right]. \tag{2.2}$$

To enforce the capacity constraint, r must be set to the q-th quantile of the marginal distribution over $\boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta}, s, \nu)$. In our dynamic model, at time step $t+1$, the decision maker's population policy loss is given by $L(\boldsymbol{\beta}, s^t, s^{t+1})$.

At an equilibrium induced by a fixed selection criteria $\boldsymbol{\beta}$, the threshold for receiving treatment is fixed over time. In other words, the previous and realized thresholds for receiving treatment are equal. Let $s(\boldsymbol{\beta})$ be the equilibrium threshold induced by the fixed selection criteria $\boldsymbol{\beta}$. If $s^t = s(\boldsymbol{\beta})$, then we have that $s^{t+1}, s^{t+2}\dots$ is a constant sequence where each term is $s(\boldsymbol{\beta})$. In the following definition, we express the decision maker's policy loss at equilibrium.

Definition 1 (Equilibrium Policy Loss)**.** Given a fixed selection criteria $\boldsymbol{\beta} \in \mathcal{B}$. Let $s(\boldsymbol{\beta})$ be an equilibrium threshold, i.e., $s(\boldsymbol{\beta})$ is equal to the q-th quantile of the marginal distribution over $\boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), \nu)$. The decision maker's population policy loss at equilibrium is given by

$$L_{\text{eq}}(\boldsymbol{\beta}) = L(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})) = \mathbb{E}\left[\ell(\pi(\mathbf{x}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), \nu); \boldsymbol{\beta}, s(\boldsymbol{\beta})), \nu)\right].$$

Under conditions where the equilibrium is guaranteed to exist and is unique, the decision maker aims to find $\boldsymbol{\beta}$ such that $L_{\text{eq}}(\boldsymbol{\beta})$ is minimized. Such an objective is motivated by the observation that it may not be feasible for the decision maker to change their selection criteria at each time step. Instead, the decision maker aims to select $\boldsymbol{\beta}$ that performs well with respect to the equilibrium behavior of the system.

2.3 Agent Behavior

Next, we specify a flexible model for agent behavior and establish when agent behavior exhibits useful properties, such as continuity and contraction. In our model, agents are heterogenous in their raw covariates and ability to modify the covariates they report to the decision maker, and they myopically choose their reported covariates based on a previous policy. Following Frankel and Kartik [2019a,b], we suppose that each agent has a private type $\nu = (\boldsymbol{\eta}, \boldsymbol{\gamma})$ sampled from a joint distribution F. An agent's raw covariates, or covariates prior to modification, are denoted by $\boldsymbol{\eta} \in \mathcal{X}$, where $\mathcal{X}$ is a convex, compact subset of $\mathbb{R}^d$. The agent's ability to change their raw covariates is given by $\boldsymbol{\gamma} \in \mathcal{G}$, where $\mathcal{G}$ is bounded. The support of F is contained in $\mathcal{X} \times \mathcal{G}$. Note that F has bounded support.

An agent, with knowledge of the selection criteria $\boldsymbol{\beta} \in \mathcal{B}$ and the previous threshold for receiving treatment $s \in \mathbb{R}$ aims to deviate from their raw covariates in hopes of getting the treatment. Let the function $c_\nu(\mathbf{y}) : \mathbb{R}^d \to \mathbb{R}$ capture the cost of an agent with type ν deviating $\mathbf{y}$ from their raw covariates $\boldsymbol{\eta}$. In addition, we suppose the agent has imperfect control over the realized value of their modified covariates. For example, an agent can influence their performance on an exam by changing the number of hours they study but cannot perfectly control their exam score. To capture this uncertainty, the agent's modified covariates are subject to noise $\boldsymbol{\epsilon} \sim N(0, \sigma^2 I_d)$. As a result, the agent's utility function takes the following form

$$u(\mathbf{x}; \boldsymbol{\beta}, s, \nu) = -c_\nu(\mathbf{x} - \boldsymbol{\eta}; \boldsymbol{\gamma}) + \pi(\mathbf{x} + \boldsymbol{\epsilon}; \boldsymbol{\beta}, s). \tag{2.3}$$

The left term is the cost to the agent of deviating from their raw covariates. The right term is the reward from receiving the treatment. Taking the expectation over the noise yields the following expected utility function. Let G be the CDF of the distribution $N(0, \sigma^2)$.

$$\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)\right] = -c_\nu(\mathbf{x} - \boldsymbol{\eta}; \boldsymbol{\gamma}) + 1 - G(s - \boldsymbol{\beta}^T \mathbf{x}). \tag{2.4}$$

We show an example expected utility function.

Example 2 (Expected Utility Function with Quadratic Cost)**.** This expected utility function has a quadratic cost of deviating from the raw covariates. Let $\boldsymbol{\gamma} \in \mathcal{G} \subset (\mathbb{R}^+)^d$.

$$\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)\right] = -(\mathbf{x} - \boldsymbol{\eta})^T \mathrm{Diag}(\boldsymbol{\gamma})(\mathbf{x} - \boldsymbol{\eta}) + 1 - G(s - \boldsymbol{\beta}^T \mathbf{x}). \tag{2.5}$$

We note that the cost function $c_\nu(\mathbf{x} - \boldsymbol{\eta}; \boldsymbol{\gamma}) = (\mathbf{x} - \boldsymbol{\eta})^T \mathrm{Diag}(\boldsymbol{\gamma})(\mathbf{x} - \boldsymbol{\eta})$ is $2 \cdot \lambda_{\min}(\mathrm{Diag}(\boldsymbol{\gamma}))$-strongly convex.

The best response mapping for an agent is obtained by finding the covariates $\mathbf{x} \in \mathcal{X}$ that maximize the expected utility function, as follows

$$\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) = \underset{\mathbf{x} \in \mathcal{X}}{\mathrm{argmax}}\, \mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)\right]. \tag{2.6}$$

The covariates that an agent reports to the decision maker is the agent's best response subject to noise $\boldsymbol{\epsilon} \sim N(0, \sigma^2 I_d)$,

$$\mathbf{x}(\boldsymbol{\beta}, s, \nu) = \mathbf{x}^*(\beta, s, \nu) + \boldsymbol{\epsilon}. \tag{2.7}$$

2.4 Properties of Agent Best Response

Using the following two assumptions, we establish a condition on the noise distribution which guarantees that the agent best response is a well-defined function and is continuously differentiable in $\boldsymbol{\beta}, s$. We also establish a related condition on the noise distribution which guarantees that the score of the agent best response is a contraction mapping.

Assumption 1. The cost function $c_\nu(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable. In addition, c_ν is also α_ν-strongly convex function for $\alpha_\nu > 0$ and $c_\nu(\mathbf{0})$ is its minimum.

Assumption 1 provides structure to the agent's cost of covariate modification by requiring that it is an α_ν-strongly convex function. The cost is minimized when the agent does not deviate from their raw covariates $\boldsymbol{\eta}$.

In the following lemma, we give a condition on the noise distribution under which the agent best response exists and is unique. This is essential so that we can treat the best response mapping as a well-defined function of $\boldsymbol{\beta}$ and s.

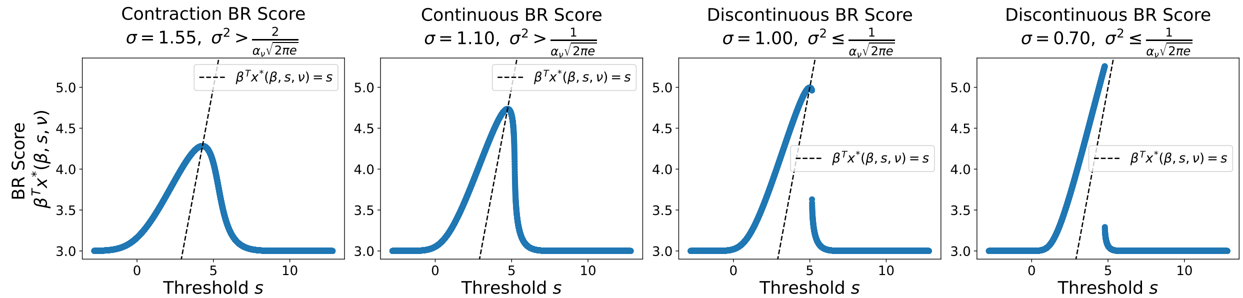

Figure 1: We plot $\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ vs. s at different noise levels. **Left**: If $\sigma^2 > \frac{2}{\alpha_\nu \cdot \sqrt{2\pi e}}$, then the score of the best response mapping is guaranteed to be contraction. **Middle Left**: If $\sigma^2 > \frac{1}{\alpha_\nu \cdot \sqrt{2\pi e}}$, then the best response mapping is guaranteed to be continuous. **Middle Right**: If $\sigma^2 \leq \frac{1}{\alpha_\nu \cdot \sqrt{2\pi e}}$, then the best response mapping may be discontinuous. **Right**: In cases where the best response mapping is discontinuous, the score of the best response mapping may not have a fixed point.

Lemma 1. *Consider an agent with type $\nu \in \mathcal{X} \times \mathcal{G}$. Under Assumption 1, the best response $\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ exists. Furthermore, if $\sigma^2 > \frac{1}{\alpha_\nu \cdot \sqrt{2\pi e}}$, then the best response $\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ is uniquely defined. Proof in Appendix C.1.*

Furthermore, using the Implicit Function Theorem, we can show that under the same conditions, the best response mapping is continuously differentiable in $\boldsymbol{\beta}, s$.

Lemma 2. *Consider an agent with type $\nu \in \mathcal{X} \times \mathcal{G}$. Under Assumption 1, if $\sigma^2 > \frac{1}{\alpha_\nu \cdot \sqrt{2\pi e}}$ and a best response $\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) \in Int(\mathcal{X})$, then the best response is continuously differentiable in $\boldsymbol{\beta}$ and s. Proof in Appendix C.2.*

We can loosely interpret Lemma 1 as follows: if G has sufficiently high variance, then the best response mapping exists and is unique.

Given a slightly stronger bound on M, we can strengthen our result and verify that the score of the agent best response mapping is a contraction in s, i.e., there is $\kappa \in (0, 1)$ such that, for any fixed $\boldsymbol{\beta} \in \mathcal{B}$ and $\nu \in \mathcal{X} \times \mathcal{G}$,

$$|\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s', \nu)| \leq \kappa|s - s'| \quad \forall s, s' \in \mathcal{S}.$$

The contraction property is useful because fixed-point iteration is known to converge for functions that are contractions (see Theorem 23).

Lemma 3. *Consider an agent with type $\nu \in \mathcal{X} \times \mathcal{G}$. Under Assumption 1, if $\sigma^2 > \frac{2}{\alpha_\nu \cdot \sqrt{2\pi e}}$ and a best response $\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) \in Int(\mathcal{X})$, then for fixed $\boldsymbol{\beta} \in \mathcal{B}$, the score of an agent's best response $\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ is a contraction mapping in s. Proof in Appendix C.3.*

We can loosely interpret Lemma 3 as follows: if G has sufficiently high variance (twice as high as that required for continuity of the best response), then the score of the best response mapping is a contraction.

We end this section by numerically investigating the role of noise on the agents' best response functions, and verify that in the absence of sufficient noise unstable behaviors may occur. Qualitatively, the reason why instability may arise is that, in a zero-noise setting, there are two modes of agent behavior. In one mode, the agent does not deviate from their raw covariates at all, so $\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) = \boldsymbol{\beta}^T\boldsymbol{\eta}$. This is either because the threshold is low enough that the agent expects to receive the treatment without deviating from their raw covariates or because the threshold is so high that the benefit of receiving the treatment does not outweigh the cost of modifying their covariates. In the other mode, the threshold takes on intermediate values, so the agent will invest the minimum effort to ensure that they receive the treatment under the previous policy, meaning that $\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) = s$. A discontinuity in the best response arises when an agent no longer finds modifying their covariates beneficial. Introducing noise increases the agent's uncertainty in whether they will receive the treatment, which causes agents to be less reactive to the previous policy and smooths the agent best response.

Under different noise settings, we analyze the score of an agent's best response with a fixed selection criteria $\boldsymbol{\beta}$ while the threshold s varies. We consider an agent with type $\nu = (\boldsymbol{\eta}, \boldsymbol{\gamma})$ where $\boldsymbol{\eta} = [3., 0.]^T$ and $\boldsymbol{\gamma} = [0.1, 1.]^T$. We suppose the decision maker's model is $\boldsymbol{\beta} = [1., 0.]^T$. Let the agent have an expected utility function with quadratic cost of covariate modification, as given in (2.5).

We visualize the score of the agent best response, $\boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$, as a function of s, the previous threshold for receiving treatment. We plot $\boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ vs. s at four different noise levels σ. In the left plot of Figure 1, the noise distribution satisfies $\sigma^2 > \frac{2}{\alpha_\nu \cdot \sqrt{2\pi e}}$, so Lemma 3 is applicable, and we observe that the score of the best response is a contraction in s. In the middle left plot, the noise distribution satisfies $\sigma^2 > \frac{1}{\alpha_\nu \cdot \sqrt{2\pi e}}$, so Lemma 2 is applicable, and the score of the best response is continuous. In the plots on the right of Figure 1, the noise distributions satisfy $\sigma^2 \leq \frac{1}{\alpha_\nu \cdot \sqrt{2\pi e}}$. In such cases, the best response mapping may be discontinuous and may not necessarily have a fixed point.

The lack of a fixed point in the score of an agent best response in in low-noise regimes (rightmost plot, Figure 1) implies that there are distributions F over agent types for which there is no equilibrium in our dynamic model in low-noise regimes. As a result, when the noise condition for continuity of the agent best response does not hold, an equilibrium of our dynamic model may not exist. In Section 3, when we establish uniqueness and existence of equilibria of our dynamic model, we assume a noise condition that guarantees continuity properties of the agents' best response mappings.

3 Mean-Field Results

Thus far, we have presented a dynamic model for capacity-constrained treatment assignment in the presence of generic strategic behavior and specified the type of strategic behavior we consider in this work. Recall that the decision maker's objective, as outlined in Section 2, is to find a selection criteria $\boldsymbol{\beta}$ that minimizes the equilibrium policy loss $L_{\text{eq}}(\boldsymbol{\beta})$. This is a sensible goal in settings where an equilibrium exists and is unique for each selection criteria $\boldsymbol{\beta}$ in consideration. In this section, we give conditions for existence and uniqueness of an equilibrium in the mean-field regime, where there are an infinite number of agents. We describe a plausible mechanism through which the equilibrium will arise in the mean-field regime. Finally, we show that the mean-field equilibrium threshold is differentiable with respect to $\boldsymbol{\beta}$, which is crucial for defining the policy effect in Section 5.

We instantiate the dynamic model from Section 2 in the mean-field regime. An infinite population of agents with types sampled from F is considered for the treatment at each time step t. Let $\boldsymbol{\beta}$ be the decision maker's fixed selection criteria. At time step $t+1$, suppose all agents best respond with knowledge of the same selection criteria $\boldsymbol{\beta}$ and previous threshold for receiving treatment s^t, and noise distribution $N(0, \sigma^2 I_d)$. Let $P_{\boldsymbol{\beta}, s^t}$ be the marginal distribution over scores of the form $\boldsymbol{\beta}^T \mathbf{x}(\boldsymbol{\beta}, s^t, \nu)$. Let $q(P_{\boldsymbol{\beta}, s^t})$ denote the q-th quantile of $P_{\boldsymbol{\beta}, s^t}$. Then, agents who score above $s^{t+1} = q(P_{\boldsymbol{\beta}, s^t})$ will receive the treatment. Iterating this procedure gives a fixed-point iteration process

$$s^{t+1} = q(P_{\boldsymbol{\beta}, s^t}) \quad t = 0, 1, 2, \ldots \tag{3.1}$$

As described in Section 2, the system is at the selection criteria $\boldsymbol{\beta}$'s equilibrium if the threshold for receiving treatment is fixed over time. The equilibrium induced by $\boldsymbol{\beta}$ is characterized by an *equilibrium threshold* s^* for which $s^* = q(P_{\boldsymbol{\beta}, s^*})$. In the iterative process in (3.1) if $s^0 = s^*$, then $s^t = s^*$ for all t.

To give conditions under which the equilibrium is unique, we use the following three assumptions.

Assumption 2. There are finitely many distinct types ν that occur with positive probability in F.

Assumption 2 is made for convenience. In combination with Assumption 1 for all agent types $\nu \sim F$, Assumption 2 guarantees that $\alpha_*(F)$, as defined below, is positive.

$$\alpha_*(F) = \inf_{\nu \in \text{supp}(F)} \alpha_\nu. \tag{3.2}$$

We will omit the dependence of $\alpha_*(F)$ on F when it is clear that there is only one type distribution of interest.

Assumption 3. For any agent type ν in the support of F and any choice of $\boldsymbol{\beta} \in \mathcal{B}$ and $s \in \mathbb{R}$, we assume that $\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) \in \text{Int}(\mathcal{X})$. In other words, we require that agent best responses fall in the interior of the set $\mathcal{X}$.

We require Assumption 3 to ensure that the best response mapping for each agent type $\nu \sim F$ is uniquely defined, so that $P_{\boldsymbol{\beta},s}$ is a valid distribution function. These assumptions, along with a noise condition that $\sigma^2 > \frac{1}{\alpha_* \cdot \sqrt{2\pi e}}$, guarantee uniqueness of the equilibrium. Note that this condition ensures that for all agent types in F, their best response mappings are well-defined (Lemma 1) and continuously differentiable in $\boldsymbol{\beta}, s$ (Lemma 2).

Theorem 4. *Fix $\boldsymbol{\beta} \in \mathcal{B}$. Under Assumption 1, 2, and 3, if $\sigma^2 > \frac{1}{\alpha_* \cdot \sqrt{2\pi e}}$ and $q(P_{\boldsymbol{\beta},s})$ has a fixed point ($s = q(P_{\boldsymbol{\beta},s})$ has a solution), then the fixed point must be unique. Proof in Appendix D.1.*

The proof of uniqueness relies on exhibiting useful properties of the distribution function $P_{\boldsymbol{\beta},s}(r)$, namely that it is continuously differentiable in its arguments and has a well-defined inverse function. When this holds, we have that a fixed point of $q(P_{\boldsymbol{\beta},s})$ is given by a value of s that solves the equation $P_{\boldsymbol{\beta},s}(s) = q$. Finally, we observe that $P_{\boldsymbol{\beta},s}(s)$ is a monotonically increasing function, so it can intersect the horizontal line $y = q$ in at most one point, yielding uniqueness.

Under the same assumptions, we can also show that the $q(P_{\boldsymbol{\beta},s})$ is continuously differentiable in $\boldsymbol{\beta}$ and s. This result follows from the Implicit Function Theorem.

Lemma 5. *Under Assumption 1, 2, and 3, if $\sigma^2 > \frac{1}{\alpha_* \cdot \sqrt{2\pi e}}$, then $q(P_{\boldsymbol{\beta},s})$ is continuously differentiable in $\boldsymbol{\beta}$ and s. Proof in Appendix D.2.*

With the result that $q(P_{\boldsymbol{\beta},s})$ is continuous, we can establish the existence of the equilibrium in the mean-field model through an application of Intermediate Value Theorem.

Theorem 6. *Fix $\boldsymbol{\beta} \in \mathcal{B}$. Under Assumption 1, 2, and 3, if $\sigma^2 > \frac{1}{\alpha_* \cdot \sqrt{2\pi e}}$, then there exists a threshold s such that $q(P_{\boldsymbol{\beta},s}) = s$. In other words, $q(P_{\boldsymbol{\beta},s})$ has at least one fixed point. Proof in Appendix D.3.*

The next two results give conditions under which the equilibrium arises via fixed-point iteration (3.1). Corollary 7 is a direct application of Banach's Fixed-Point Theorem.

Corollary 7. *Fix $\boldsymbol{\beta} \in \mathcal{B}$. Under Assumptions 1, 2, and 3, if $q(P_{\boldsymbol{\beta},s})$ is a contraction mapping in s where s^* is the unique fixed point of $q(P_{\boldsymbol{\beta},s})$, then fixed-point iteration (3.1) converges to s^*. Proof in Appendix D.4.*

In Corollary 8, we give a sufficient condition for ensuring that $q(P_{\boldsymbol{\beta},s})$ is a contraction. The sufficient condition is equivalent to ensuring that for all agent types ν in the support of F, their best response mappings are contractions in s. In the proof of this corollary, we use the fact that the derivative of $q(P_{\boldsymbol{\beta},s})$ with respect to s is a convex combination of the derivatives of the agents' best response mappings with respect to s. We note that a function is a contraction if and only if its derivative is bounded between -1 and 1 (Lemma 22). So, ensuring that each agent's best response mapping is a contraction guarantees that $q(P_{\boldsymbol{\beta},s})$ is a contraction. We note that this condition is sufficient but not necessary for $q(P_{\boldsymbol{\beta},s})$ to be a contraction.

Corollary 8. *Fix $\boldsymbol{\beta} \in \mathcal{B}$. Under Assumptions 1, 2, and 3, if $\sigma^2 > \frac{2}{\alpha_* \cdot \sqrt{2\pi e}}$, then $q(P_{\boldsymbol{\beta},s})$ is a contraction in s and fixed-point iteration (3.1) converges to s^*, the unique fixed point of $q(P_{\boldsymbol{\beta},s})$. Proof in Appendix D.5.*

Thus far, we have demonstrated that under sufficient regularity conditions, for a fixed selection criteria $\boldsymbol{\beta}$, an equilibrium exists and is unique in the mean-field limit, and fixed-point iteration is a mechanism through which this equilibrium arises. Crucially, the existence and uniqueness of equilibria induced by fixed selection criteria allows us to define a function $s(\boldsymbol{\beta}) : \mathcal{B} \to \mathcal{S}$ that maps selection criteria $\boldsymbol{\beta} \in \mathcal{B}$ to the equilibrium threshold $s(\boldsymbol{\beta}) \in \mathcal{S}$ that characterizes the criteria's mean-field equilibrium. The following theorem establishes the differentiability of s.

Corollary 9. *Under Assumption 1, 2, and 3, if $\sigma^2 > \frac{1}{\alpha_* \cdot \sqrt{2\pi e}}$, then we can define a function $s : \mathcal{B} \to \mathbb{R}$ that maps model parameters $\boldsymbol{\beta}$ to the unique fixed point $s^* \in \mathbb{R}$ that satisfies $q(P_{\boldsymbol{\beta},s^*}) = s^*$. The function s is continuously differentiable in $\boldsymbol{\beta}$. Proof in Appendix D.6.*

We conclude this section by observing that a selection criteria impacts the level of competition that agents experience through its impact on the equilibrium threshold for receiving the treatment. Although different selection criteria will induce different equilibrium thresholds in the mean-field limit, under regularity conditions the equilibrium thresholds will vary smoothly with the selection criteria. These results make it possible to define and estimate policy effects in Section 5.

4 Finite Sample Approximation

Understanding equilibrium behavior of our dynamic model in the finite regime is of interest because our ultimate goal is to learn optimal equilibrium policies in finite samples. In this section, we instantiate the dynamic model from Section 2 in the regime where a finite number of agents are considered for the treatment at each time step. A difficulty of the finite regime is that deterministic equilibria do not exist. Instead, we give conditions under which stochastic equilibria arise and show that, in large samples, these stochastic equilibria sharply approximate the mean-field limit derived above.

Let $\boldsymbol{\beta}$ be the decision maker's fixed selection criteria. At each time step, n new agents with types sampled i.i.d. from F are considered for the treatment. For example, in the application of college admissions, the sampled agents at each time step represent a class of students applying for admission each year. At time step $t+1$, the n agents who are being considered for the treatment best respond with knowledge of the same selection criteria $\boldsymbol{\beta}$, previous threshold for receiving treatment $\hat{s}_n^t$, and noise distribution $N(0, \sigma^2 I_d)$. In the finite model, the decision maker observes an *empirical* score distribution $P^n_{\boldsymbol{\beta},\hat{s}_n^t}$. Let $q(P^n_{\boldsymbol{\beta},\hat{s}_n^t})$, denote the q-th quantile of $P^n_{\boldsymbol{\beta},\hat{s}_n^t}$. Then, agents who score above $\hat{s}_n^{t+1} = q(P^n_{\boldsymbol{\beta},\hat{s}_n^t})$ will receive the treatment. Iterating this procedure gives a stochastic version of fixed-point iteration

$$\hat{s}_n^{t+1} = q(P^n_{\boldsymbol{\beta},\hat{s}_n^t}), \quad t = 0, 1, 2, \dots \tag{4.1}$$

Since new agents are sampled at each time step, $q(P^n_{\boldsymbol{\beta},\cdot})$ is a random operator. Iterating the random operator $q(P^n_{\boldsymbol{\beta},\cdot})$ given some initial threshold $\hat{s}_n^0$ yields a stochastic process $\{\hat{s}_n^t\}_{t\geq 0}$. We note that for any fixed $\boldsymbol{\beta}$, the random operator $q(P^n_{\boldsymbol{\beta},\cdot})$ approximates the deterministic function $q(P_{\boldsymbol{\beta},\cdot})$.

In Section 3, we showed that there are conditions under which fixed-point iteration of the mean-field model's deterministic operator $q(P_{\boldsymbol{\beta},\cdot})$ converges to s^*, the mean-field equilibrium threshold. In the finite model, if $q(P^n_{\boldsymbol{\beta},\cdot})$ closely approximates $q(P_{\boldsymbol{\beta},\cdot})$, we may expect that there are conditions under which the stochastic process $\{\hat{s}_n^t\}_{t\geq 0}$ will eventually oscillate in a small neighborhood about s^*. This is illustrated in Figure 2.

We define a constant that will be used in our concentration inequality and convergence result for the behavior of the finite system. For $\epsilon > 0$

$$M_\epsilon = \inf_{s\in\mathbb{R}} \min\{P_{\boldsymbol{\beta},s}(q(P_{\boldsymbol{\beta},s}) + \epsilon) - q, q - P_{\boldsymbol{\beta},s}(q(P_{\boldsymbol{\beta},s}) - \epsilon)\}. \tag{4.2}$$

The following result guarantees that M_ϵ is positive and gives a finite-sample concentration inequality for the behavior of $q(P^n_{\boldsymbol{\beta},s})$.

Lemma 10. *Under Assumption 1, 2, 3, if* $\sigma^2 > \frac{1}{\alpha_* \sqrt{2\pi e}}$, *then* $M_\epsilon > 0$ *and*

$$P(|q(P_{\boldsymbol{\beta},s}) - q(P^n_{\boldsymbol{\beta},s})| < \epsilon) \geq 1 - 4e^{-2nM_\epsilon^2}.$$

Proof in Appendix E.1.

Notably, the bound in the concentration inequality does not depend on the particular choice of s. We use this lemma to characterize the behavior of the system of n agents for sufficiently large iterates t and number of agents n in Theorem 11. Theorem 11 shows under the same conditions that enable fixed-point iteration in the mean-field model to converge to the mean-field equilibrium threshold (3.1), sufficiently large iterates of the stochastic fixed-point iteration in the finite model (4.1) will lie in a small neighborhood about the mean-field equilibrium threshold with high probability. We can view these iterates as stochastic equilibria of the finite system.

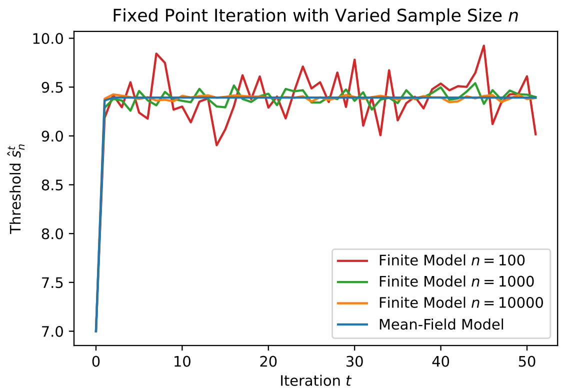

Figure 2: Given a fixed distribution F over agent types, we consider the finite model for various n and the mean-field model. In this example, the conditions of Theorem 7 are satisfied, so fixed-point iteration in the mean-field model (3.1) converges to its unique fixed point. Fixed-point iteration in the finite models (4.1) oscillates about the fixed point of the mean-field model. For large n, we observe that the iterates $\{\hat{s}_n^t\}$ are more concentrated about the fixed point of the mean-field model.

Theorem 11. *Fix $\boldsymbol{\beta} \in \mathcal{B}$. Suppose Assumptions 1, 2, 3 hold. Let $\epsilon \in (0,1), \delta \in (0,1)$, and s^* is the mean-field equilibrium threshold induced by selection criteria $\boldsymbol{\beta}$. Let κ is the Lipschitz constant of $q(P_{\boldsymbol{\beta},s})$. Let $\epsilon_g = \frac{\epsilon(1-\kappa)}{2}$. Let $S = |\hat{s}_n^0 - s^*|$. If $\sigma^2 > \frac{2}{\alpha_* \sqrt{2\pi e}}$, then for t such that*

$$t \geq \left\lceil \frac{\log(\frac{\epsilon}{2S})}{\log \kappa} \right\rceil$$

and n such that

$$n \geq \frac{1}{2M_{\epsilon_g}^2} \log(\frac{4t}{\delta}),$$

we have that

$$P(|\hat{s}_n^t - s^*| \geq \epsilon) \leq \delta.$$

Proof in Appendix E.2.

The main idea of the proof of this result is at each time step the quantity $|\hat{s}_n^t - s^*|$ can be decomposed into two terms,

$$|\hat{s}_n^t - s^*| \leq |q(P^n_{\boldsymbol{\beta},\hat{s}_n^{t-1}}) - q(P_{\boldsymbol{\beta},\hat{s}_n^{t-1}})| + |q(P_{\boldsymbol{\beta},\hat{s}_n^{t-1}}) - s^*|.$$

The first term on the right side is a noise term that arises due to the difference between an empirical quantile and a population quantile. The second term on the right side can be upper bounded by $\kappa|\hat{s}_n^{t-1} - s^*|$ because $q(P_{\boldsymbol{\beta},\cdot})$ is assumed to be a contraction with Lipschitz constant κ. Recursively applying this decomposition k times leaves a vanishing series of dependent noise terms and a term that depends on the distance of the $t-k$-th iterate from s^*. Analyzing the series of noise terms is difficult due to the dependence between the noise terms. We sidestep this challenge by introducing a sequence of independent random variables each of which stochastically dominates the corresponding noise terms in our series of interest. Analysis of the series of the independent random variables yields our result.

Also, the following corollary is a building block for our consistency results in Section 5.

Corollary 12. *Fix $\boldsymbol{\beta} \in \mathcal{B}$. Let S, κ be defined as in Theorem 11. Let $\{t_n\}$ be a sequence such that $t_n \uparrow \infty$ as $n \to \infty$ and $t_n \prec \exp(n)$ (t_n grows slower than exponentially fast in n) Under the conditions of Theorem 11, $\hat{s}_n^{t_n} \xrightarrow{p} s^*$, where s^* is the unique fixed point of $q(P_{\boldsymbol{\beta},s})$. Proof in Appendix E.3.*

5 Learning Policies via Gradient Descent

In this section, we apply the equilibrium concepts developed in Sections 3 and 4 to define and estimate the policy effect, the gradient of the equilibrium policy loss with respect to the selection criteria. To enable learning of the optimal policy, we rely on estimation of the derivative of the policy loss, a method that is motivated by prior works [Chetty, 2009, Wager and Xu, 2021].

First, we give conditions under which the loss function is continuously differentiable as a function of $\boldsymbol{\beta}$ and define the policy effect in terms of the mean-field equilibrium threshold. Next, using results from Section 4, we give methods for estimating these effects in finite samples in a unit-level randomized experiment as in Munro et al. [2021], Wager and Xu [2021]. Finally, we propose a method for learning the optimal policy by using the policy effect estimator.

5.1 Policy Effect

Recall the equilibrium policy loss defined in Section 2.

Lemma 13. *Under the conditions of Corollary 9, $L_{eq}(\boldsymbol{\beta})$ is continuously differentiable in $\boldsymbol{\beta}$. Proof in Appendix F.1.*

From the definition of $L_{\text{eq}}(\boldsymbol{\beta})$ in Definition 1, we have that the total derivative of $L_{\text{eq}}(\boldsymbol{\beta})$ can be written as

$$\frac{dL_{\text{eq}}}{d\boldsymbol{\beta}}(\boldsymbol{\beta}) = \frac{\partial L}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})) + \Big(\frac{\partial L}{\partial s}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})) + \frac{\partial L}{\partial r}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta}))\Big) \cdot \frac{\partial s}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}). \tag{5.1}$$

We decompose the total derivative of $L_{\text{eq}}(\boldsymbol{\beta})$, or the policy effect, into two parts. The first term corresponds to the model effect and the second term corresponds to the equilibrium effect.

Definition 3 (Model Effect)**.** Let τ_{ME} denote the model effect of deploying selection criteria $\boldsymbol{\beta}$ on the equilibrium policy loss the decision maker incurs.

$$\tau_{\text{ME}}(\boldsymbol{\beta}) = \frac{\partial L}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})).$$

The selection criteria $\boldsymbol{\beta}$ impacts the decision maker's loss because agents modify their covariates in response to the criteria and the criteria is also used to score the agents. Both of these influence the treatments that the agents receive and thus the loss the decision maker incurs.

In the absence of capacity constraints, the model effect is sufficient for capturing the policy effect. However, due to the decision maker's capacity constraint, the equilibrium threshold for receiving treatment also depends on the selection criteria. So, we must also account for how the decision maker's loss changes with respect to the equilibrium threshold and how the equilibrium threshold changes with respect to the selection criteria. Following notation from (2.2), we write $\partial L/\partial s$ and $\partial L/\partial r$ for the partial derivatives of L in its second and third arguments respectively.

Definition 4 (Equilibrium Effect)**.** Let τ_{EE} denote the equilibrium effect of deploying selection criteria $\boldsymbol{\beta}$ on the equilibrium policy loss the decision maker incurs.

$$\tau_{\text{EE}}(\boldsymbol{\beta}) = \Big(\frac{\partial L}{\partial s}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})) + \frac{\partial L}{\partial r}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta}))\Big) \cdot \frac{\partial s}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}).$$

The previous threshold for receiving treatment s impacts the decision maker's loss because agents modify their covariates in response to s. This influences the treatments that agents receive and thus the loss the decision maker incurs. The realized threshold for receiving treatment r impacts the decision maker's loss because it determines agents' treatment assignments, which influences the loss the decision maker incurs. At equilibrium, we have that $s = r = s(\boldsymbol{\beta})$, so we can account for both of these effects simultaneously.

Definition 5 (Policy Effect)**.** Let τ_{PE} denote the policy effect of deploying selection criteria $\boldsymbol{\beta}$ on the equilibrium policy loss the decision maker incurs.

$$\tau_{\text{PE}}(\boldsymbol{\beta}) = \tau_{\text{ME}}(\boldsymbol{\beta}) + \tau_{\text{EE}}(\boldsymbol{\beta}).$$

The policy captures both the model effect and equilibrium effect.

5.2 Estimation of Policy Effect

We derive estimators for the model, equilibrium, policy effects through a unit-level randomized experiment in a finite samples. In a system consisting of n agents, we apply symmetric, mean-zero perturbations to the parameters of the policy that each agent responds to. Let R represent the distribution of Rademacher random variables and let R^d represent a distribution over d-dimensional Rademacher random variables. For agent i, we perturb the policy parameters as follows

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + b\boldsymbol{\zeta}_i, \quad \boldsymbol{\zeta}_i \sim R^d,$$
$$s_i = s + b\zeta_i, \quad \zeta_i \sim R.$$

In practice, these perturbations can be applied by telling agent i that they will be scored according to $\boldsymbol{\beta}_i$ instead of $\boldsymbol{\beta}$ and a small shock of size $-b\zeta_i$ will be applied to their score. Instead of reporting covariates in response to the previous policy $\pi(\mathbf{x}; \boldsymbol{\beta}, s)$, we presume that with information about the perturbations, agent i will report covariates in response to a policy $\pi(\mathbf{x}; \boldsymbol{\beta}_i, s_i)$ as follows:

$$\mathbf{x}(\boldsymbol{\beta}_i, s_i, \nu_i) = \mathbf{x}^*(\boldsymbol{\beta}_i, s_i, \nu_i) + \boldsymbol{\epsilon}_i \quad \boldsymbol{\epsilon}_i \sim N(0, \sigma^2 I_d),$$

where

$$\mathbf{x}^*(\boldsymbol{\beta}_i, s_i, \nu_i) = \operatorname*{argmax}_{\mathbf{x}\in\mathcal{X}} \mathbb{E}_{\boldsymbol{\epsilon}} \left[u(\mathbf{x}; \boldsymbol{\beta}_i, s_i, \nu_i) \right]. \tag{5.2}$$

Let $P_{\boldsymbol{\beta},s,b}$ denote the distribution over scores when each agent i responds to $(\boldsymbol{\beta}_i, s_i)$ and the prescribed perturbation is applied to the agent's score. For clarity, we contrast the form of a score sampled from $P_{\boldsymbol{\beta},s}$ to that of the form of a score sampled from $P_{\boldsymbol{\beta},s,b}$. An agent with type ν_i who best responds to $\boldsymbol{\beta}, s$ will obtain a score $\boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta}, s, \nu_i)$ in the unperturbed setting. An agent with type ν_i who best responds to a perturbed version of $\boldsymbol{\beta}, s$ will obtain a score $\boldsymbol{\beta}_i^T\mathbf{x}(\boldsymbol{\beta}_i, s_i, \nu_i) - b\zeta_i$.

The purpose of applying these perturbations is so that we can recover the relevant gradient terms by running a linear regression from the perturbations to outcomes of interest, which include the decision maker's loss and the proportion of agents whose score exceeds a threshold r. To construct the estimators of the model and equilibrium effects, we rely on gradient estimates of the loss function $L(\boldsymbol{\beta}, s, r)$ and gradient estimates of the complementary CDF of the score distribution $\Pi(\boldsymbol{\beta}, s; r)$, which is defined as

$$\Pi(\boldsymbol{\beta}, s; r) = 1 - P_{\boldsymbol{\beta},s}(r). \tag{5.3}$$

In this experiment, we suppose that thresholds evolve by the stochastic fixed-point iteration process below. Note that it differs slightly from the process given in (4.1).

$$\hat{s}_{b,n}^{t+1} = \begin{cases} q(P^n_{\boldsymbol{\beta},\hat{s}_n^t,b}) & q(P^n_{\boldsymbol{\beta},\hat{s}_n^t,b}) \in [-D, D] \\ -D & q(P^n_{\boldsymbol{\beta},\hat{s}_n^t,b}) < -D \\ D & q(P^n_{\boldsymbol{\beta},\hat{s}_n^t,b}) > D \end{cases} \tag{5.4}$$

This process differs from (4.1) because it includes perturbations of size b to the selection criteria and threshold and restricts the threshold to a bounded set $\mathcal{S} = [-D, D]$ where D is sufficiently large constant so that the equilibrium threshold $s^* \in \mathcal{S}$. Such a set exists because it can be shown there exists $D > 0$ such that $|q(P_{\boldsymbol{\beta},s})| < D$ for all $s \in \mathbb{R}$.

Analyzing the stochastic process $\{\hat{s}_n^t\}_{t\geq 0}$ generated by the iteration above presents two technical challenges. First, the above stochastic process truncates the threshold values so that they lie in $\mathcal{S}$, whereas the results from Section 4 do not involve truncation. Nevertheless, the truncation is a contraction map to the equilibrium threshold, so the results of Section 4 also apply to the stochastic fixed point iteration process with truncated threshold values. The other challenge is that the results from Section 3 and Section 4 focus on the setting where all agents best respond to the same policy $\pi(\mathbf{x}; \boldsymbol{\beta}, s)$. Nevertheless, under the following assumption, we can show that for sufficiently small b, analogous results hold under unit-level perturbations, where each agent i best responds to the policy $\pi(\mathbf{x}; \boldsymbol{\beta}_i, s_i)$.

Assumption 4. For all types $\nu = (\boldsymbol{\eta}, \boldsymbol{\gamma})$ that have positive probability in F, we have that $\boldsymbol{\eta} \in \mathrm{Int}(\mathcal{X})$.

To show that results from Section 4 transfer to the setting with unit-level perturbations, we can define a new distribution over agent types $\tilde{F}$ and new cost functions. When agents with types sampled from $\tilde{F}$ best respond to $\pi(\mathbf{x};\boldsymbol{\beta}, s)$ according to the new cost function, the score distribution that results equals $P_{\boldsymbol{\beta},s,b}$.

Now, we can define the model effect estimator.

Definition 6 (Model Effect Estimator)**.** We consider an experiment where n agents are considered for the treatment. Let b be the size of the perturbation. Let each row of $\mathbf{Z} \in \mathbb{R}^{n\times d}$ correspond to $b\boldsymbol{\zeta}_i^T$, the perturbation applied to the selection criteria observed by the i-th agent. Since the n agents will best respond to these perturbations as in (5.2), we observe an empirical distribution over scores $P^n_{\boldsymbol{\beta},s,b}$. Let each entry of $\boldsymbol{\ell}(\boldsymbol{\beta}, s, r) \in \mathbb{R}^n$ correspond to the loss the decision maker incurs on the i-th agent as follows

$$\boldsymbol{\ell}_i(\boldsymbol{\beta}, s, r) = \ell(\pi(\mathbf{x}(\boldsymbol{\beta}_i, s_i, \nu_i); \boldsymbol{\beta}_i, r + b\zeta_i), \nu_i). \tag{5.5}$$

Let $\hat{\Gamma}_{\ell,\boldsymbol{\beta}}(\boldsymbol{\beta}, s)$ be the regression coefficient that is obtained by running OLS of $\boldsymbol{\ell}$ on $\mathbf{Z}$. In particular,

$$\hat{\Gamma}_{\ell,\boldsymbol{\beta}}(\boldsymbol{\beta}, s, r) = b^{-1}\Big(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\zeta}_i\boldsymbol{\ell}_i(\boldsymbol{\beta}, s, r)\Big).$$

The model effect estimator with sample size n, perturbation size b, and iteration t as

$$\hat{\tau}^t_{\mathrm{ME},b,n}(\boldsymbol{\beta}) = \hat{\Gamma}_{\ell,\boldsymbol{\beta}}(\boldsymbol{\beta}, \hat{s}^t_{b,n}, \hat{s}^{t+1}_{b,n}), \tag{5.6}$$

where $\hat{s}^t_{b,n}$ is given by (5.4).

To prove consistency of this estimator, we require additional conditions on the loss function ℓ.

Assumption 5. The functions $\ell(0, \nu)$ and $\ell(1, \nu)$ are continuous on $\mathcal{X} \times \mathcal{G}$. In addition, $\ell(\pi, \nu)$ is bounded on $\{0, 1\} \times \mathcal{X} \times \mathcal{G}$.

Theorem 14. *Let $\{t_n\}$ be a sequence such that $t_n \uparrow \infty$ as $n \to \infty$ and $t_n \prec \exp(n)$. Let $\mathcal{S} = [-D, D]$ for a sufficiently large constant $D > 0$, so that the equilibrium threshold $s^* \in \mathcal{S}$. We consider the sequence of model effect estimators given by $\hat{\tau}^{t_n}_{ME,n}(\boldsymbol{\beta})$. Under Assumptions 1, 2, 3, 4, and 5, if $\sigma^2 > \frac{2}{\alpha_*\sqrt{2\pi e}}$, then there exists a sequence $\{b_n\}$ such that $b_n \to 0$ so that $\hat{\tau}^{t_n}_{ME,b_n,n}(\boldsymbol{\beta}) \xrightarrow{p} \tau_{ME}(\boldsymbol{\beta})$. Proof in Appendix F.2.*

Second, we define the equilibrium effect estimator. Although the same approach applies, estimating the equilibrium effect is more complicated than estimating the model effect. We estimate the equilibrium effect by estimating the two components of the equilibrium effect, $\frac{\partial L}{\partial s} + \frac{\partial L}{\partial r}$ and $\frac{\partial s}{\partial \boldsymbol{\beta}}$.

Definition 7 (Equilibrium Effect Estimator)**.** We consider an experiment where n agents are considered for the treatment. Let b be the size of the perturbation. Let each row of $\mathbf{Z}_{\boldsymbol{\beta}} \in \mathbb{R}^{n\times d}$ and of $\mathbf{Z}_s \in \mathbb{R}^{n\times d}$ correspond to the perturbation applied to the linear model and baseline score, respectively for the i-th agent. Since the n agents will best respond to these perturbations as in (5.2), we observe an empirical distribution over scores $P^n_{\boldsymbol{\beta},s,b}$. Let each entry of $\boldsymbol{\ell}, \boldsymbol{\pi} \in \mathbb{R}^n$ correspond to the following outcomes for the i-th agent

$$\boldsymbol{\ell}_i(\boldsymbol{\beta}, s, r) = \ell(\pi(\mathbf{x}(\boldsymbol{\beta}_i, s_i, \nu_i); \boldsymbol{\beta}_i, r + b\zeta_i), \nu_i),$$
$$\boldsymbol{\pi}_i(\boldsymbol{\beta}, s, r) = \pi(\mathbf{x}(\boldsymbol{\beta}_i, s_i, \nu_i); \boldsymbol{\beta}_i, r).$$

Let $\hat{\Gamma}_{\ell,s,\ell,r}(\boldsymbol{\beta}, s), \hat{\Gamma}_{\pi,\boldsymbol{\beta}}(\boldsymbol{\beta}, s, r)$, and $\hat{\Gamma}_{\pi,s}(\boldsymbol{\beta}, s, r)$ correspond to the regression coefficients from running OLS of $\boldsymbol{\ell}$ on $\mathbf{Z}_s$, $\boldsymbol{\pi}$ on $\mathbf{Z}_{\boldsymbol{\beta}}$, and $\boldsymbol{\pi}$ on $\mathbf{Z}_s$, respectively. In particular,

$$\hat{\Gamma}_{\ell,s,\ell,r}(\boldsymbol{\beta}, s, r) = b^{-1}\Big(\frac{1}{n}\sum_{i=1}^n \zeta_i\zeta_i^T\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^n \zeta_i\boldsymbol{\ell}_i(\boldsymbol{\beta}, s, r)\Big),$$

$$\hat{\Gamma}_{\pi,\boldsymbol{\beta}}(\boldsymbol{\beta}, s, r) = b^{-1}\Big(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\zeta}_i\boldsymbol{\pi}_i(\boldsymbol{\beta}, s, r)\Big),$$

$$\hat{\Gamma}_{\pi,s}(\boldsymbol{\beta}, s, r) = b^{-1}\Big(\frac{1}{n}\sum_{i=1}^n \zeta_i\zeta_i^T\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^n \zeta_i\boldsymbol{\pi}_i(\boldsymbol{\beta}, s, r)\Big).$$

Let $\{h_n\}$ be a sequence such that $h_n \to 0$ and $nh_n \to \infty$. Let $p^n_{\boldsymbol{\beta},s,b_n}(r)$ denote a kernel density estimate of $p_{\boldsymbol{\beta},s,b_n}(r)$ with kernel function $k(z) = \mathbb{I}(z \in [-\frac{1}{2}, \frac{1}{2}))$ and bandwidth h_n.

We define the model effect estimator with sample size n and iteration t as

$$\hat{\tau}^t_{\mathrm{EE},b,n}(\boldsymbol{\beta}) = \hat{\Gamma}_{\ell,s,\ell,r}(\boldsymbol{\beta}, \hat{s}^t_{b,n}, \hat{s}^{t+1}_{b,n}) \cdot \left(\frac{1}{p^n_{\boldsymbol{\beta},\hat{s}^t_{b,n},b}(\hat{s}^t_{b,n}) - \hat{\Gamma}_{\pi,s}(\boldsymbol{\beta}, \hat{s}^t_{b,n}, \hat{s}^t_{b,n})} \cdot \hat{\Gamma}_{\pi,\boldsymbol{\beta}}(\boldsymbol{\beta}, \hat{s}^t_{b,n}, \hat{s}^t_{b,n}) \right). \tag{5.7}$$

In Theorem 15, we show that these three linear approximations and the density estimate of the score distribution enable us to estimate the equilibrium effect.

Theorem 15. *Let $\{t_n\}$ be a sequence such that $t_n \uparrow \infty$ as $n \to \infty$ and $t_n \prec \exp(n)$. We consider the sequence of equilibrium effect estimators given by $\hat{\tau}^{t_n}_{EE,n}(\boldsymbol{\beta})$. Under the conditions of Theorem 14, there exists a sequence $\{b_n\}$ such that $b_n \to 0$ so that $\hat{\tau}^{t_n}_{EE,n}(\boldsymbol{\beta}) \xrightarrow{p} \tau_{EE}(\boldsymbol{\beta})$. Proof in Appendix F.3.*

Finally, we can sum the estimators of the model and equilibrium effects to estimate the policy effect.

Corollary 16. *Let $\{t_n\}$ be a sequence such that $t_n \uparrow \infty$ as $n \to \infty$ and $t_n \prec \exp(n)$. We consider the sequence of approximate policy effects given by*

$$\hat{\tau}^{t_n}_{PE,b_n,n}(\boldsymbol{\beta}) = \hat{\tau}^{t_n}_{ME,b_n,n}(\boldsymbol{\beta}) + \hat{\tau}^{t_n}_{EE,b_n,n}(\boldsymbol{\beta}).$$

Under the conditions of Theorem 14, there exists a sequence $\{b_n\}$ such that $b_n \to 0$ so that $\hat{\tau}^{t_n}_{PE,n}(\boldsymbol{\beta}) \xrightarrow{p} \tau_{PE}(\boldsymbol{\beta})$. Proof in Appendix F.4.

5.3 Learning the Optimal Policy

We now describe an algorithm (see Algorithm 1) for learning the optimal policy. Following Wager and Xu [2021], the algorithm entails first learning equilibrium-adjusted gradients of the policy loss as discussed above and then updating the selection criteria via gradient descent. In this paper, we will only investigate empirical properties of this approach, and refer to Wager and Xu [2021] for formal results for this type of gradient-based learning.

The decision maker runs the algorithm for J epochs. In Section 2, we describe that it may be infeasible for the decision maker to update the selection criteria at each time step. This algorithm requires the decision maker to deploy an updated selection criteria at each epoch j. In other words, updates to the selection criteria are necessary but infrequent. We emphasize that deploying different selection criteria is only necessary for the learning procedure, and ultimately, we aim to learn a fixed selection criteria that minimizes the equilibrium policy loss.

In epoch j, the decision maker deploys a policy $\boldsymbol{\beta}^j$. Through the stochastic fixed-point iteration process with perturbations (5.4), a stochastic equilibrium induced by $\boldsymbol{\beta}^j$ emerges, yielding the threshold for receiving treatment s^j. Each agent best responds to their perturbed policy and the decision maker observes their reported covariates. Following the procedure from Section 5.2, the decision maker can then use the outcomes and the perturbations to estimate the policy effect of $\boldsymbol{\beta}^j$ on the equilibrium policy loss (Algorithm 2). The decision maker can set $\boldsymbol{\beta}^{j+1}$ by taking a gradient descent step from $\boldsymbol{\beta}^j$ using the policy effect estimator as the gradient. Any first-order variant of stochastic gradient descent can be used. In our experiments, we use vanilla stochastic gradient descent and projected stochastic gradient descent.

6 Numerical Experiments

In this section, we demonstrate that the policy effect estimator defined Section 5 can be used to learn a capacity-constrained policy that achieves lower equilibrium policy loss compared to approaches that do not account for strategic behavior or only account for the model effect. First, we give a one-dimensional toy example, where we suppose that F contains *cross-types*, which are pairs of agent types where one agent has higher ability to modify their covariates and the other has more favorable raw covariates. In the toy example, we demonstrate that using the policy effect estimator $\hat{\tau}_{\mathrm{PE}}$ enables a decision maker to learn the optimal solution. Second, in a high-dimensional ($d = 10$) simulation with a generic distribution F over agent types, we also demonstrate that learning with the policy effect estimator $\hat{\tau}_{\mathrm{PE}}$ yields solutions with lower equilibrium policy loss than just the model effect estimator $\hat{\tau}_{\mathrm{ME}}$.

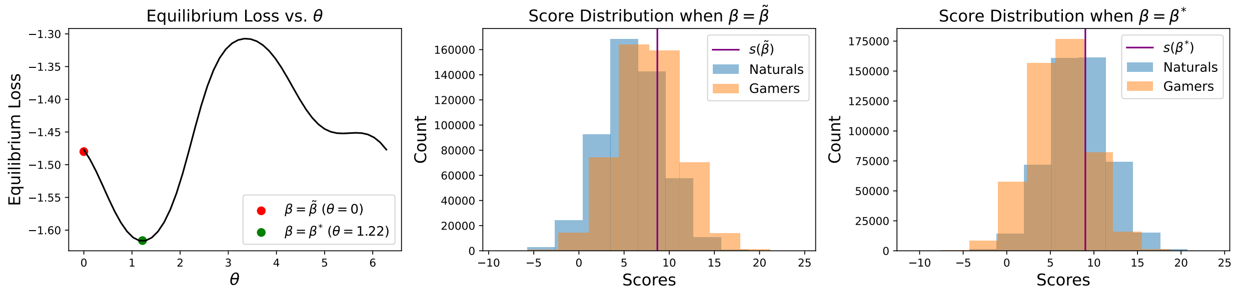

Figure 3: **Left**: We plot the (expected) equilibrium policy loss across $\theta = \arctan(\boldsymbol{\beta}_1/\boldsymbol{\beta}_0)$. Deploying $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$ (equivalently, $\theta = 0$) does not yield an optimal loss. We note that the equilibrium policy loss has a global minimum $\boldsymbol{\beta}^* = [0.345, 0.938]^T$(equivalently, $\theta^* = 1.22$). **Middle**: When $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$, the naturals make up only 34% of agents who score above the threshold. **Right**: When $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, the naturals make up 69% of agents who score above the threshold.

6.1 Toy Example

For a one-dimensional example, we consider policies with the following parametrization

$$\boldsymbol{\beta} = [\cos\theta, \sin\theta]^T, \text{ where } \theta \in \mathbb{S}^1.$$

Algorithm 1: Gradient Descent with $\hat{\tau}_{\text{PE}}$

---

while $j \leq J$ **do**
- Decision maker deploys $\boldsymbol{\beta}^j$;
- Stochastic fixed-point iteration for sufficiently many iterations with unit-level perturbations (see (5.4)) until s^j is reached ;
- **for** $i \in \{1 \dots n\}$ **do**
  - Sample random perturbation $\boldsymbol{\zeta}_i \sim R^d$ and $\zeta_i \sim R$;
  - $\boldsymbol{\beta}_i^j \leftarrow \boldsymbol{\beta}^j + b_n \boldsymbol{\zeta}_i$;
  - $s_i^j \leftarrow s^j + b_n \zeta_i$;
  - Agent i best responds to $\boldsymbol{\beta}_i^j, s_i^j$;
  - Decision maker observes best response $\mathbf{x}_i^j$;
- **end**
- Given the scores $\{\boldsymbol{\beta}_i^j \mathbf{x}_i^j - b_n \zeta_i^j\}_{i=1}^n$, the decision maker computes the q-th quantile of the scores q^j and density of scores at q^j, yields ρ^j;
- **for** $i \in \{1 \dots n\}$ **do**
  - Decision incurs loss $\boldsymbol{\ell}_i^j$ and measures
  - $\boldsymbol{\pi}_i^j \leftarrow \mathbb{I}((\boldsymbol{\beta}_i^j)^T \mathbf{x}_i^j > s^j)$;
- **end**
- $\mathbf{Z}_{\boldsymbol{\beta}}^j \leftarrow b_n \boldsymbol{\zeta}^j$ is the $n \times d$ matrix of perturbations $\boldsymbol{\zeta}$;
- $\mathbf{Z}_s^j \leftarrow b_n \zeta^j$ is the $n \times 1$ matrix of perturbations ζ;
- $\boldsymbol{\ell}^j$ is the n-length vector of losses $\boldsymbol{\ell}_i$;
- $\boldsymbol{\pi}^j$ is the n-length vector of indicators $\boldsymbol{\pi}_i$;
- Construct gradient estimate Γ^j from $\mathbf{Z}_{\boldsymbol{\beta}}^j, \mathbf{Z}_s^j, \boldsymbol{\ell}^j, \boldsymbol{\pi}^j, \rho^j$ (See Algorithm 2);
- Take a projected gradient descent step
- $\boldsymbol{\beta}^{j+1} \leftarrow \text{Proj}_{\mathcal{B}}(\boldsymbol{\beta}^j - a \cdot \Gamma^j)$;

end

Algorithm 2: Construct gradient estimates

Run OLS of $\boldsymbol{\ell}^j$ on $\mathbf{Z}^j_{\boldsymbol{\beta}}$: $\Gamma^j_{\ell,\boldsymbol{\beta}} \leftarrow ((\mathbf{Z}^j_{\boldsymbol{\beta}})^T \mathbf{Z}^j_{\boldsymbol{\beta}})^{-1} (\mathbf{Z}^j_{\boldsymbol{\beta}})^T \boldsymbol{\ell}^j)$;
Run OLS of $\boldsymbol{\ell}^j$ on $\mathbf{Z}^j_s$: $\Gamma^j_{\ell,s,\ell,r} \leftarrow ((\mathbf{Z}^j_s)^T \mathbf{Z}^j_s)^{-1} ((\mathbf{Z}^j_s)^T \boldsymbol{\ell}^j)$;
Run OLS of $\boldsymbol{\pi}^j$ on $\mathbf{Z}^j_{\boldsymbol{\beta}}$: $\Gamma^j_{\pi,\boldsymbol{\beta}} \leftarrow ((\mathbf{Z}^j_{\boldsymbol{\beta}})^T \mathbf{Z}^j_{\boldsymbol{\beta}})^{-1} ((\mathbf{Z}^j_{\boldsymbol{\beta}})^T \boldsymbol{\pi}^j)$;
Run OLS of $\boldsymbol{\pi}^j$ on $\mathbf{Z}^j_s$: $\Gamma^j_{\pi,s} \leftarrow ((\mathbf{Z}^j_s)^T \mathbf{Z}^j_s)^{-1} ((\mathbf{Z}^j_s)^T \boldsymbol{\pi}^j)$;
$\Gamma^j_{s,\boldsymbol{\beta}} \leftarrow \frac{1}{\rho^j - \Gamma^j_{\pi,s}} \cdot \Gamma^j_{\pi,\boldsymbol{\beta}}$.
$\Gamma^j \leftarrow \Gamma^j_{\ell,\boldsymbol{\beta}} + \Gamma^j_{\ell,s,\ell,r} \cdot \Gamma^j_{s,\boldsymbol{\beta}}$

We suppose that the capacity constraint limits the decision maker to accept only 30% of the agent population. To define the decision maker's loss, suppose ℓ is specified as follows

$$\ell(\pi, \nu) = \begin{cases} -\boldsymbol{\eta}_1 & \pi = 1 \\ 0 & \pi = 0 \end{cases}.$$

The decision maker's equilibrium policy loss $L_{\text{eq}}(\boldsymbol{\beta})$ is given by Definition 1.

We consider an agent distribution where agents are heterogeneous in their raw covariates and ability to modify their observed covariates. We suppose that

$$\boldsymbol{\eta} \in [0, 10]^2, \quad \boldsymbol{\gamma} \in [0.01, 20]^2, \quad \mathbf{x} \in [0, 10]^2.$$

The variance of the noise distribution σ^2 is set to ensure the continuous differentiability property of the quantile mapping of the score distribution; we set $\sigma = 3.30$. Agents optimize the quadratic utility function in (2.5). So, the entry $\boldsymbol{\gamma}_i$ quantifies the cost of gaming $\boldsymbol{\eta}_i$.

Motivated by Frankel and Kartik [2019b], we consider an agent distribution with two groups of agent types in the population of equal proportion, the *naturals* and the *gamers*. The naturals have

$$\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \sim \text{Uniform}[5, 7], \quad \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \sim \text{Uniform}[10, 20].$$

In contrast, the gamers have

$$\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \sim \text{Uniform}[3, 5], \quad \boldsymbol{\gamma}_1 \sim \text{Uniform}[0.01, 0.02], \quad \boldsymbol{\gamma}_2 \sim \text{Uniform}[10, 20].$$

In this simulation, there are 10 agent types, 5 naturals and 5 gamers. The naturals and gamers are cross types as in Frankel and Kartik [2019b] because the naturals have higher values of $\boldsymbol{\eta}$ compared to the gamers and the gamers have lower cost to modifying $\boldsymbol{\eta}_1$ compared to the naturals.

Note that the decision maker incurs lower loss when they admit any natural compared to when they admit any gamers because naturals have higher $\boldsymbol{\eta}_1$ compared to gamers. Under the naive assumption that agents will report $\mathbf{x} = \boldsymbol{\eta}$, the decision maker minimizes their policy loss by using the selection criteria $\tilde{\boldsymbol{\beta}} = \mathbf{e}_1$. However, the gamers have high ability to deviate from $\boldsymbol{\eta}_1$ when reporting $\mathbf{x}_1$. So, a naive application of $\tilde{\boldsymbol{\beta}}$ as the selection criteria could potentially result in the decision maker accepting many gamers, yielding suboptimal policy loss.

Intuitively, there should exist a better policy in this setting. We note that all agents are relatively homogenous in their ability to deviate from $\boldsymbol{\eta}_2$ when reporting $\mathbf{x}_2$ because $\boldsymbol{\gamma}_2 \sim \text{Uniform}[10, 20]$ for all agents. At the same time, $\boldsymbol{\eta}_2$ is correlated with $\boldsymbol{\eta}_1$. So, a selection criteria that places some weight on $\mathbf{x}_2$ should allow the decision maker to obtain lower policy loss by accepting more naturals.

We plot the equilibrium policy loss of decision maker as a function of $\theta = \arctan(\boldsymbol{\beta}_1/\boldsymbol{\beta}_0)$ in (Figure 3, left plot). As expected, we observe that deploying the naive policy $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$, which corresponds to $\theta = 0$, is suboptimal for minimizing the equilibrium policy loss. When $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$, we observe that only 35% of agents who score above $s(\boldsymbol{\beta})$ are naturals (Figure 3, middle plot). The policy $\boldsymbol{\beta}^* = [0.345, 0.938]^T$ achieves the optimal equilibrium policy loss, and as expected it places considerable weight on $\mathbf{x}_2$. When $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, we observe that 69% of agents who score above $s(\boldsymbol{\beta})$ are naturals (Figure 3, right plot).

We compare the solutions obtained by running stochastic gradient descent with $\hat{\tau}_{\text{PE}}$ and $\hat{\tau}_{\text{ME}}$. We optimize $\boldsymbol{\beta}$ via vanilla stochastic gradient descent on θ, the polar-coordinate representation of $\boldsymbol{\beta}$. We initialize gradient

Method	$\lvert L_{\text{eq}}(\hat{\boldsymbol{\beta}}) - L_{\text{eq}}(\boldsymbol{\beta}^*)\rvert$	$\lvert\hat{\theta} - \theta^*\rvert$
None (Set $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$)	0.19 ± 0.04	1.24 ± 0.08
GD with $\hat{\tau}_{\text{ME}}$	0.08 ± 0.02	0.68 ± 0.13
GD with $\hat{\tau}_{\text{PE}}$	0.00 ± 0.00	0.03 ± 0.02

Table 1: Over 10 random trials, we observe that gradient descent with the policy effect converges to the optimal θ^* (or in Cartesian coordinates, $\boldsymbol{\beta}^*$). However, gradient descent with the model effect does not converge to θ^*.

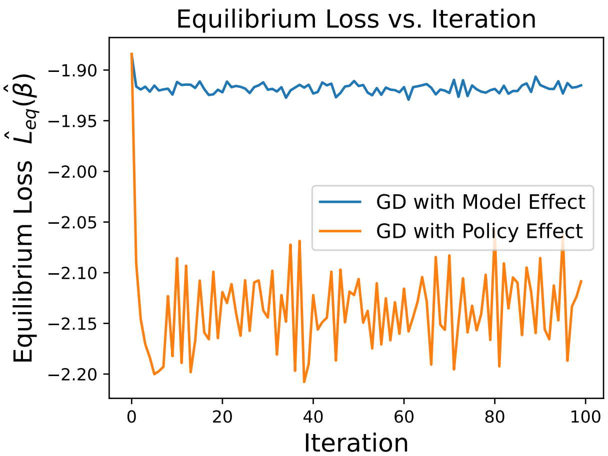

Figure 4: We plot the equilibrium policy loss obtained from iterates of gradient descent with $\hat{\tau}_{\text{ME}}$ and $\hat{\tau}_{\text{PE}}$ in our high-dimensional simulation ($d = 10$) and find that gradient descent with $\hat{\tau}_{\text{PE}}$ converges to a solution that obtains lower equilibrium policy loss.

descent with $\tilde{\boldsymbol{\beta}}$ ($\theta = 0$). We report the equilibrium policy loss obtained by the learned parameters $\hat{\boldsymbol{\beta}}$ after 100 iterations of gradient descent. We assume that $n = 1000000$ agents are observed by the decision maker at each iteration. We report the absolute difference between the equilibrium policy loss of the optimal solution and the learned solution. In addition, we report the absolute difference between the polar coordinate representations of the solutions $\lvert\hat{\theta} - \theta^*\rvert$, where $\hat{\theta} = \arctan(\hat{\boldsymbol{\beta}}_1/\hat{\boldsymbol{\beta}}_0)$ and $\theta^* = \arctan(\boldsymbol{\beta}^*_1/\boldsymbol{\beta}^*_0)$.

Across 10 random trials (where the randomness is over the sampled agent types and sampled agents) , we observe that gradient descent using the policy effect $\hat{\tau}_{\text{PE}}$ from Corollary 16 converges to θ^* (Table 1). Meanwhile, gradient descent with the model effect $\hat{\tau}_{\text{ME}}$ from Theorem 14 converges to a policy that attains suboptimal equilibrium policy loss (Table 1). This demonstrates the value of accounting for the equilibrium effect. Nevertheless, we note that $\hat{\tau}_{\text{ME}}$ is a relatively strong baseline because it accounts for agents' strategic behavior with knowledge of the selection criteria $\boldsymbol{\beta}$'s impact on the decision maker's loss. In absence of capacity constraints, gradient descent with $\hat{\tau}_{\text{ME}}$ will enable learning of the optimal solution.

6.2 High-Dimensional Simulation

For $d = 10$, in this simulation we consider d-dimensional linear policies

$$\boldsymbol{\beta} \in \mathcal{B}, \text{ where } \mathcal{B} = \{\mathbf{y} \in \mathbb{R}^d \mid ||\mathbf{y}||_2 = 1\}.$$

We suppose the capacity constraint only allows the decision maker to accept 30% of the agent population. To define the decision maker's loss, suppose that ℓ is specified as follows

$$\ell(\pi, \nu) = \begin{cases} -\boldsymbol{\eta}_1 & \pi = 1 \\ 0 & \pi = 0 \end{cases}.$$

The decision maker's equilibrium policy loss $L_{\text{eq}}(\boldsymbol{\beta})$ is given by Definition 1.

Method	Equilibrium Policy Loss $\hat{L}_{\text{eq}}(\hat{\boldsymbol{\beta}})$
GD with $\hat{\tau}_{\text{ME}}$	-1.81 ± 0.14
GD with $\hat{\tau}_{\text{PE}}$	-2.05 ± 0.14

Table 2: Across 10 random trials, we find that GD with $\hat{\tau}_{\text{PE}}$ attains lower equilibrium policy loss than GD with $\hat{\tau}_{\text{ME}}$ in our high-dimensional simulation ($d = 10$). A one-sided paired t-test, where we compare the final loss incurred of the policy learned via GD with $\hat{\tau}_{\text{PE}}$ and GD with $\hat{\tau}_{\text{ME}}$ with the same random seed, yields a p-value of 1e-5.

We suppose that

$$\boldsymbol{\eta} \in [0, 10]^d, \quad \boldsymbol{\gamma} \in [0.05, 5]^d, \quad \mathbf{x} \in [0, 10]^d.$$

The variance of the noise distribution σ^2 is set to ensure the continuous differentiability property of the quantile mapping of the score distribution; we set $\sigma = 1.10$. Agents optimize the quadratic utility function in (2.5), and the entry $\boldsymbol{\gamma}_i$ quantifies the cost of gaming $\boldsymbol{\eta}_i$. We consider a population with 10 agent types. For each agent type $(\boldsymbol{\eta}, \boldsymbol{\gamma})$, we have that

$$\boldsymbol{\eta}_i \sim \text{Uniform}[3, 8], \quad \boldsymbol{\gamma}_i \sim \text{Uniform}[0.05, 5], \quad i \in \{1, \ldots d\}.$$

We optimize $\boldsymbol{\beta}$ via projected stochastic gradient descent, initialized with

$$\hat{\boldsymbol{\beta}} = [\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}, \ldots \frac{1}{\sqrt{d}}]^T.$$

We compare the equilibrium policy loss of the gradient descent iterates obtained by using $\hat{\tau}_{\text{ME}}$ as the gradient to those obtained by using $\hat{\tau}_{\text{PE}}$ as the gradient (Figure 4). We assume that $n = 1000000$ agents are observed by the decision maker at each iteration. Across 10 random trials (where the randomness is over the sampled agent types and the sampled agents), we observe that gradient descent with $\hat{\tau}_{\text{PE}}$ finds a solution with lower equilibrium policy loss than gradient descent with $\hat{\tau}_{\text{ME}}$ (Table 2). Again, we note that GD with $\hat{\tau}_{\text{ME}}$ is a relatively strong baseline because it captures how the decision maker's loss changes with respect to the selection criteria in a way that accounts for agents' strategic behavior.

References

Daron Acemoglu and Martin Kaae Jensen. Robust comparative statics in large static games. In *49th IEEE Conference on Decision and Control (CDC)*, pages 3133–3139. IEEE, 2010.

Daron Acemoglu and Martin Kaae Jensen. Robust comparative statics in large dynamic economies. *Journal of Political Economy*, 123(3):587–640, 2015.

Saba Ahmadi, Hedyeh Beyhaghi, Avrim Blum, and Keziah Naggita. On classification of strategic agents who can both game and improve. *arXiv preprint arXiv:2203.00124*, 2022.

Susan Athey, Dean Eckles, and Guido W Imbens. Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240, 2018.

Debopam Bhattacharya and Pascaline Dupas. Inferring welfare maximizing treatment assignment under budget constraints. *Journal of Econometrics*, 167(1):168–196, 2012.

Daniel Björkegren, Joshua E Blumenstock, and Samsun Knight. Manipulation-proof machine learning. *arXiv preprint arXiv:2004.03865*, 2020.

Thomas Blake and Dominic Coey. Why marketplace experimentation is harder than it seems: The role of test-control interference. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 567–582, 2014.

John Bound, Brad Hershbein, and Bridget Terry Long. Playing the admissions game: Student reactions to increasing college competition. *Journal of Economic Perspectives*, 23(4):119–46, 2009.

Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Michael Brückner, Christian Kanzow, and Tobias Scheffer. Static prediction games for adversarial learning problems. *The Journal of Machine Learning Research*, 13(1):2617–2654, 2012.

HE Buchanan and TH Hildebrandt. Note on the convergence of a sequence of functions of a certain type. *The Annals of Mathematics*, 9(3):123–126, 1908.

Raj Chetty. Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annu. Rev. Econ.*, 1(1):451–488, 2009.

Luis C Corchón. Comparative statics for aggregative games the strong concavity case. *Mathematical Social Sciences*, 28(3):151–165, 1994.

Augustin Cournot. *Researches into the Mathematical Principles of the Theory of Wealth.* Routledge, 1982.

Nilesh Dalvi, Pedro Domingos, Mausam Sumit, and Sanghai Deepak Verma. Adversarial classification. In *In Proceedings of the Tenth International Conference on Knowledge Discovery and Data Mining*, pages 99–108. ACM Press, 2004.

Phoebus J Dhrymes. *Mathematics for econometrics*, volume 984. Springer, 1978.

Jinshuo Dong, Aaron Roth, Zachary Schutzman, Bo Waggoner, and Zhiwei Steven Wu. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 55–70, 2018.

Alex Frankel and Navin Kartik. Improving information from manipulable data. *Journal of the European Economic Association*, 2019a.

Alex Frankel and Navin Kartik. Muddled information. *Journal of Political Economy*, 127(4):1739–1776, 2019b.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM conference on innovations in theoretical computer science*, pages 111–122, 2016.

James J Heckman, Lance Lochner, and Christopher R Taber. General equilibrium treatment effects: A study of tuition policy, 1998.

Meena Jagadeesan, Celestine Mendler-Dünner, and Moritz Hardt. Alternative microfoundations for strategic classification. In *International Conference on Machine Learning*, pages 4687–4697. PMLR, 2021.

Nathan Kallus and Angela Zhou. Minimax-optimal policy learning under unobserved confounding. *Management Science*, 67(5):2870–2890, 2021.

Michihiro Kandori, George J Mailath, and Rafael Rob. Learning, mutation, and long run equilibria in games. *Econometrica: Journal of the Econometric Society*, pages 29–56, 1993.

Toru Kitagawa and Aleksey Tetenov. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.

Jon Kleinberg and Manish Raghavan. How do classifiers induce agents to invest effort strategically? *ACM Transactions on Economics and Computation (TEAC)*, 8(4):1–23, 2020.

Sagi Levanon and Nir Rosenfeld. Generalized strategic classification and the case of aligned incentives. *arXiv preprint arXiv:2202.04357*, 2022.

Lydia T Liu, Nikhil Garg, and Christian Borgs. Strategic ranking. *arXiv preprint arXiv:2109.08240*, 2021.

Charles F Manski. Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4):1221–1246, 2004.

Dov Monderer and Lloyd S Shapley. Fictitious play property for games with identical interests. *Journal of economic theory*, 68(1):258–265, 1996.

Evan Munro. Learning to personalize treatments when agents are strategic. *arXiv preprint arXiv:2011.06528*, 2020.

Evan Munro, Stefan Wager, and Kuang Xu. Treatment effects in market equilibrium. *arXiv preprint arXiv:2109.11647*, 2021.

Whitney K Newey and Daniel McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994.

James M Ortega. *Numerical analysis: a second course.* SIAM, 1990.

Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

David Pollard. *Convergence of stochastic processes.* Springer Science & Business Media, 2012.

R Tyrrell Rockafellar. *Convex analysis*, volume 18. Princeton university press, 1970.

Jun Shao. *Mathematical statistics.* Springer Science & Business Media, 2003.

Kyle Thayer and Amy J Ko. Barriers faced by coding bootcamp students. In *Proceedings of the 2017 ACM Conference on International Computing Education Research*, pages 245–253, 2017.

Stefan Wager and Kuang Xu. Experimenting in equilibrium. *Management Science*, 2021.

A Experiment Details

A.1 Toy Experiment

We use a learning rate of $a = 1$ in GD with $\hat{\tau}_{\mathrm{PE}}$. We use a learning rate of $a = 0.25$ in GD with $\hat{\tau}_{\mathrm{ME}}$. We use a perturbation size $b = 0.025$ for $\boldsymbol{\beta}$ and $b = 0.2$ for s.

A.2 High-Dimensional Experiment

We use a learning rate of $a = 1$ in GD with $\hat{\tau}_{\mathrm{ME}}$ and in GD with $\hat{\tau}_{\mathrm{PE}}$. We use a perturbation size $b = 0.025$ for $\boldsymbol{\beta}$ and $b = 0.2$ for s.

B Standard Results

Lemma 17. *Let $\mathcal{X} \subset \mathbb{R}^d$ is a convex set. Let $f : \mathcal{X} \to \mathbb{R}$ be a strictly concave function. If f has a global maximizer, then the maximizer is unique (Boyd et al. [2004]).*

Lemma 18. *Let $f : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} \subset \mathbb{R}^d$ is a convex set, be a twice-differentiable function. If f is a strictly concave function and $\mathbf{x}^*$ is in the interior of $\mathcal{X}$, then $\mathbf{x}^*$ is the unique global maximizer of f on $\mathcal{X}$ if and only if $\nabla f(\mathbf{x}^*) = 0$ (Boyd et al. [2004]).*

Theorem 19 (Implicit Function Theorem)**.** *Suppose $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is continuously differentiable in an open set containing $(\mathbf{x}_0, \mathbf{y}_0)$ and $\mathbf{f}(\mathbf{x}_0, \mathbf{y}_0) = 0$. Let $\mathbf{M}$ be the $m \times m$ matrix*

$$D_{n+j}\mathbf{f}^i(\mathbf{x}, \mathbf{y}) \quad 1 \leq i, j \leq m.$$

If det$(\mathbf{M}) \neq 0$, then there is an open set $X \subset \mathbb{R}^n$ containing $\mathbf{x}_0$ and an open set $Y \subset \mathbb{R}^m$ containing $\mathbf{y}_0$, with the following property: for each $\mathbf{x} \in X$ there is a unique $\mathbf{g}(\mathbf{x}) \in Y$ such that $\mathbf{f}(\mathbf{x}, \mathbf{g}(\mathbf{x})) = 0$. The function g is continuously differentiable.

Theorem 20 (Sherman-Morrison Formula)**.** *Suppose* $\mathbf{A} \in \mathbb{R}^{d\times d}$ *is an invertible square matrix,* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$*. Then* $\mathbf{A} + \mathbf{u}\mathbf{v}^T$ *is invertible iff* $1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u} \neq 0$*. In this case,*

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}.$$

Lemma 21. *Suppose* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{d\times d}$ *are positive definite matrices. If* $\mathbf{A} - \mathbf{B}$ *is positive semidefinite, then* $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ *is positive semidefinite (Dhrymes [1978]).*

Lemma 22. *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a differentiable real function. The function* f *is a contraction with modulus* $\kappa \in (0, 1)$ *if and only if* $|f'(x)| \leq \kappa$ *for all* $x \in \mathbb{R}$ *(Ortega [1990]).*

Theorem 23 (Banach's Fixed-Point Theorem)**.** *Let* (X, d) *be a non-empty complete metric space with a contraction mapping* $T : X \to X$*. Then* T *admits a unique fixed-point* x^* *such that* $T(x^*) = x^*$*. Furthermore, for any number* $x_0 \in X$*, the sequence defined by* $x_n = T(x_{n-1}), n \geq 1$ *converges to the unique fixed point* x^**.*

Theorem 24. *Let* $X_1, X_2, \dots, X_n$ *be i.i.d. random variables from a c.d.f.* F*. Let* θ_p *be the p-th quantile of* F *and let* $\hat{\theta}_p$ *be the p-th quantile of* F_n*. Suppose* F *satisfies* $p < F(\theta_p + \epsilon)$ *for any* $\epsilon > 0$*. Then for every* $\epsilon > 0$*, then*

$$P(|\hat{\theta}_p - \theta_p| > \epsilon) \leq 4e^{-2nM_\epsilon^2},$$

where $M_\epsilon = \min\{F(\theta_p + \epsilon) - p, p - F(\theta_p - \epsilon)\}$ *(Theorem 5.9, Shao [2003]).*

Theorem 25 (Bernoulli's Inequality)**.** *For every* $r \geq 0$ *and* $x \geq -1$*,* $(1 + x)^r \geq 1 + rx$*.*

Lemma 26. *If the* w_i *i.i.d.,* Θ *is compact,* $a(\cdot, \theta)$ *is continuous at each* $\theta \in \Theta$ *with probability one, and there is* $d(w)$ *with* $||a(w, \theta)|| \leq d(w)$ *for all* $\theta \in \Theta$ *and* $\mathbb{E}[d(w)] < \infty$*, then* $E[a(w, \theta)]$ *is continuous and*

$$\sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{i=1}^{n} a(w_i, \theta) - \mathbb{E}\left[a(w, \theta)\right] \right| \xrightarrow{p} 0$$

(Lemma 2.4, Newey and McFadden [1994]).

Lemma 27. *Suppose* Θ *is compact and* $f(\theta)$ *is continuous. Then* $\sup_{\theta\in\Theta} |\hat{f}_n(\theta) - f(\theta)| \to 0$ *if and only if* $\hat{f}_n(\theta) \xrightarrow{p} f(\theta)$ *for all* $\theta \in \Theta$ *and* $\{\hat{f}_n(\theta)\}$ *is stochastically equicontinuous (Lemma 2.8, Newey and McFadden [1994]).*

Lemma 28. *Suppose* $\{Z_n(t)\}$ *is a collection of stochastic processes indexed by* $t \in \mathcal{T}$*. Suppose* $\{Z_n(t)\}$ *is stochastically equicontinuous at* $t_0 \in \mathcal{T}$*. Let* τ_n *be a sequence of random elements of* $\mathcal{T}$ *known to satisfy* $\tau_n \xrightarrow{p} t_0$*. It follows that* $Z_n(\tau_n) - Z_n(t_0) \xrightarrow{p} 0$*, (Pollard [2012]).*

Lemma 29. *Let* $f_n : \mathcal{X} \to \mathbb{R}$ *where* $\mathcal{X} \subset \mathbb{R}$ *is a compact set. Let* $\{f_n\}$ *be a sequence of continuous, monotonic functions that converge pointwise to a continuous function* f*. Then* $f_n \to f$ *uniformly (Buchanan and Hildebrandt [1908]).*

Lemma 30. *Let* $f_n : \mathcal{X} \to \mathbb{R}$ *where* $\mathcal{X} \subset \mathbb{R}^d$ *is a compact set. Let* $\{f_n\}$ *be a sequence of continuous, concave functions that converge pointwise to* f*. Furthermore, assume that* f *is continuous. Then* $f_n \to f$ *uniformly (Rockafellar [1970]).*

Lemma 31. *Let* $f_n : \mathcal{X} \to \mathbb{R}$ *where* $\mathcal{X} \subset \mathbb{R}$ *is a compact set. Let* $\{f_n\}$ *be a sequence of continuous functions that converge uniformly to* f*. Suppose each* f_n *has exactly one root* $x_n \in \mathcal{X}$ *and* f *has exactly one root* $x^* \in \mathcal{X}$*. Then* $x_n \to x^*$*. Proof in Appendix G.1.*

Lemma 32. *Let* $f_n : \mathcal{X} \to \mathbb{R}$ *where* $\mathcal{X} \subset \mathbb{R}^d$ *is a compact set. Let* $\{f_n\}$ *be a sequence of continuous functions that converge uniformly to* f*. Suppose each* f_n *has exactly one maximizer* $x_n \in \mathcal{X}$ *and* f *has exactly one maximizer* $x^* \in \mathcal{X}$*. Then* $x_n \to x^*$*. Proof in Appendix G.2.*

Theorem 33. *Let us assume the following:*

1. K *vanishes at infinity, and* $\int_{-\infty}^{\infty} K^2(x)dx < \infty$*,*

2. $h_n \to 0$ *as* $n \to \infty$,

3. $nh_n \to \infty$ *as* $n \to \infty$.

Let $f^n(x)$ *be a kernel density estimate of the density function* f *with* n *samples, kernel* K*, and bandwidth* h_n*. Then* $f^n(x) \xrightarrow{p} f(x)$*, as* $n \to \infty$ *(Parzen [1962]).*

C Proofs of Agent Results

Lemma 34. *Under Assumption 1, the expected utility (2.4) is twice continuously differentiable in* $\mathbf{x}, \boldsymbol{\beta}, s$*. Proof in Appendix G.3.*

Lemma 35. *Under Assumption 1, if* $\sigma^2 > \frac{1}{\alpha_\nu\sqrt{2\pi e}}$*, then* $\nabla^2_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};\boldsymbol{\beta},s,\nu)\right]$ *is negative definite. Proof in Appendix G.4.*

Lemma 36. *Under Assumption 1, if* $\sigma^2 > \frac{1}{\alpha_\nu\sqrt{2\pi e}}$ *and* $\mathbf{x} \in Int(\mathcal{X})$*, then* $\mathbf{x} = \mathbf{x}^*(\boldsymbol{\beta},s,\nu)$ *if and only if* $\nabla_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};\boldsymbol{\beta},s,\nu)\right] = 0$*. Proof in Appendix G.5.*

Lemma 37. *Under Assumption 1, if* $\sigma^2 > \frac{1}{\alpha_\nu\sqrt{2\pi e}}$ *and* $\mathbf{x}^*(\boldsymbol{\beta},s,\nu) \in Int(\mathcal{X})$*, then*

$$\boldsymbol{\beta}^T\nabla_s\mathbf{x}^* = G''(s-\boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} - \Big(\frac{(G''(s-\boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta})^2}{1+G''(s-\boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}\Big), \tag{C.1}$$

where $\nu = (\boldsymbol{\eta},\boldsymbol{\gamma}), \mathbf{x}^*(s) := \mathbf{x}^*(\boldsymbol{\beta},s,\nu)$*, and* $\mathbf{H} := \nabla^2 c_\nu(\mathbf{x}^*(s)-\boldsymbol{\eta};\boldsymbol{\gamma})$*. Proof in Appendix G.6.*

Lemma 38. *Let* $\mathbf{H} = \nabla^2 c_\nu(\mathbf{y})$ *for some* $\mathbf{y} \in \mathcal{X}$*. Under Assumption 1, we have that* $\mathbf{H}$ *is positive definite,* $\mathbf{H}^{-1}$ *is positive definite, and*

$$\sup_{\mathbf{z}\in\mathcal{B}} \mathbf{z}^T\mathbf{H}^{-1}\mathbf{z} \le \frac{1}{\alpha_\nu}. \tag{C.2}$$

Proof in Appendix G.7.

Lemma 39. *Under Assumptions 1, if* $\sigma^2 > \frac{1}{\alpha_\nu\sqrt{2\pi e}}$ *and* $\mathbf{x}^*(\boldsymbol{\beta},s,\nu) \in Int(\mathcal{X})$*, then the function* $h(s;\boldsymbol{\beta},\nu) = s - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta},s,\nu)$ *is strictly increasing in* s*. Proof in Appendix G.8.*

Lemma 40. *Let* $\nu = (\boldsymbol{\eta},\boldsymbol{\gamma})$*. Consider* $\omega(s;\boldsymbol{\beta},\nu) = \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta},s,\nu)$*. Under Assumption 1, if* $\sigma^2 > \frac{1}{\alpha_\nu\sqrt{2\pi e}}$ *and* $\mathbf{x}^*(\boldsymbol{\beta},s,\nu) \in Int(\mathcal{X})$*,*

$$\lim_{s\to\infty} \omega(s;\boldsymbol{\beta},\nu) = \boldsymbol{\beta}^T\boldsymbol{\eta}. \tag{C.3}$$

$$\lim_{s\to-\infty} \omega(s;\boldsymbol{\beta},\nu) = \boldsymbol{\beta}^T\boldsymbol{\eta}. \tag{C.4}$$

Proof in Appendix G.9.

Lemma 41. *Let* $\nu = (\boldsymbol{\eta},\boldsymbol{\gamma})$*. Consider* $\omega(s;\boldsymbol{\beta},\nu) = \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta},s,\nu)$*. Under Assumption 1, if* $\sigma^2 > \frac{1}{\alpha_\nu\sqrt{2\pi e}}$ *and* $\mathbf{x}^*(\boldsymbol{\beta},s,\nu) \in Int(\mathcal{X})$*, then* $\omega(s;\boldsymbol{\beta},\nu)$ *is maximized at a point* s^**,* $\omega(s)$ *is increasing when* $s < s^*$ *and is decreasing on* $s > s^*$*. Proof in Appendix G.10.*

C.1 Proof of Lemma 1

We can apply Lemma 34 to show that the expected utility (2.4) is twice continuously differentiable in $\mathbf{x}$, and thus continuous in $\mathbf{x}$. Since $\mathcal{X}$ is compact, the expected utility attains a maximum value on $\mathcal{X}$ because a continuous function attains a maximum value on a compact set. Thus, there exists $\mathbf{x} \in \mathcal{X}$ that maximizes the expected utility.

From Lemma 35, $\nabla^2_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};\boldsymbol{\beta},s,\nu)\right]$ is negative definite everywhere. This implies that the expected utility is strictly concave. Since $\mathcal{X}$ is a convex set and the expected utility is strictly concave, we can apply Lemma 17 to conclude that the best response is the unique maximizer of the expected utility on $\mathcal{X}$.

C.2 Proof of Lemma 2

We use the following abbreviations for the expected utility and best response

$$\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};s)\right] := \mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};\boldsymbol{\beta},s,\nu)\right]$$
$$\mathbf{x}^*(s) := \mathbf{x}^*(\boldsymbol{\beta},s,\nu),$$

where $\boldsymbol{\beta},\nu$ are fixed. We aim to show that if a best response $\mathbf{x}^*(s) \in \text{Int}(\mathcal{X})$, then $\mathbf{x}^*$ is continuously differentiable in s. By Lemma 36, if a best response $\mathbf{x}^*(s) \in \text{Int}(\mathcal{X})$, then it satisfies $\nabla_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};s)\right] = 0$. Our goal is to apply the Implicit Function Theorem (Theorem 19) to show that $\mathbf{x}$ that satisfies $\nabla_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};s)\right] = 0$ can be written as a continuously differentiable function of s.

Now, we verify the conditions of the Implicit Function Theorem. The conditions include $\nabla_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};s)\right]$ is continuously differentiable in its arguments and that at the point $(\mathbf{x}_0,s_0)$ where the theorem is applied, we have $\det(\nabla^2_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x}_0;s_0)\right]) \neq 0$. The first condition follows from Lemma 34, which in fact states that $\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};s)\right]$ is twice continuously differentiable in its arguments. For the second condition, we note that $\nabla^2_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};s)\right]$ is always negative definite everywhere from Lemma 35.

As a result, the conditions of the Implicit Function Theorem are satisfied. Let $(\mathbf{x}_0,s_0)$ be any point that that satisfies $\nabla_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};s)\right] = 0$. In an open neighborhood $V \times W \subset \mathbb{R}^d \times \mathbb{R}$ of $(\mathbf{x}_0,s_0)$, for each $s \in W$ there is a unique $\mathbf{g}(s) \in V$ such that $\nabla_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{g}(s);s)\right] = 0$ and $\mathbf{g}$ is a continuously differentiable function of s. If $\mathbf{g}(s) \in \text{Int}(\mathcal{X})$, then $\mathbf{g}(s)$ coincides with the unique best response $\mathbf{x}^*(s)$ by Lemma 36. This implies that for $\mathbf{x}(s) \in \text{Int}(\mathcal{X})$, then $\mathbf{x}^*$ is continuously differentiable in s.

An analogous proof can be used to show that the best response $\mathbf{x}^*$ is continuously differentiable in $\boldsymbol{\beta}$.

C.3 Proof of Lemma 3

Without loss of generality, we fix $\boldsymbol{\beta},\nu$. We abbreviate

$$\mathbf{x}^*(s) := \mathbf{x}^*(\boldsymbol{\beta},s,\nu).$$

To show that $\boldsymbol{\beta}^T\mathbf{x}^*(s)$ is a contraction, it is sufficient to show that $|\boldsymbol{\beta}^T\nabla_s\mathbf{x}^*| < 1$ (Lemma 22). We show this result in two steps. First, we use Lemma 39 to show that $\boldsymbol{\beta}^T\nabla_s\mathbf{x}^* < 1$. Second, we can use our assumption that $\sigma^2 > \frac{2}{\alpha_\nu\sqrt{2\pi e}}$ to show $\boldsymbol{\beta}^T\nabla_s\mathbf{x}^* > -1$

We first show that $\boldsymbol{\beta}^T\nabla_s\mathbf{x}^* < 1$. Since we assume that $\sigma^2 > \frac{2}{\alpha_\nu\sqrt{2\pi e}}$, we certainly have that $\sigma^2 > \frac{1}{\alpha_\nu\sqrt{2\pi e}}$, so we can apply Lemma 39. This gives us that $h(s;\boldsymbol{\beta},\nu) = s - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta},s,\nu)$ is strictly increasing. We can apply Lemma 2 to establish the differentiability of the best response, which consequently gives the differentiability of h. Since h is also strictly increasing, we have that

$$\frac{dh}{ds} = 1 - \boldsymbol{\beta}^T\nabla_s\mathbf{x}^*(s) > 0.$$

This gives us that $\boldsymbol{\beta}^T\nabla_s\mathbf{x}^*(s) < 1$.

Now, we establish that $\boldsymbol{\beta}^T\nabla_s\mathbf{x}^* > -1$. We use Lemma 37 to get an expression ((C.1)) for the $\boldsymbol{\beta}^T\nabla\mathbf{x}^*_s$.

We can simplify (C.1) as follows,

$$\boldsymbol{\beta}^T\nabla_s\mathbf{x} = G''(s-\boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} - \Big(\frac{(G''(s-\boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta})^2}{1+G''(s-\boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}\Big) \tag{C.5}$$

$$= \frac{G''(s-\boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}{1+G''(s-\boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}. \tag{C.6}$$

We study the numerator of the term on the right side of (C.6).

$$G''(s-\boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} \geq \inf_y G''(y)\sup_{\mathbf{z}\in\mathcal{B}}\mathbf{z}^T\mathbf{H}^{-1}\mathbf{z} \tag{C.7}$$

$$\geq \Big(-\frac{1}{\sigma^2\sqrt{2\pi e}}\Big)\cdot\frac{1}{\alpha_\nu} \tag{C.8}$$

$$> -\frac{\alpha_\nu}{2}\cdot\frac{1}{\alpha_\nu} \tag{C.9}$$

$$= -\frac{1}{2}. \tag{C.10}$$

(C.7) follows from the observation that $G''(y)$ may take negative values while $\mathbf{H}^{-1}$ is positive definite (Lemma 38). (C.8) holds because $-\frac{1}{\sigma^2\sqrt{2\pi e}} \leq G''(y) \leq \frac{1}{\sigma^2\sqrt{2\pi e}}$. (C.8) is an application of Lemma 38. In (C.9), we use our assumption that $\sigma^2 > \frac{2}{\alpha_\nu\sqrt{2\pi e}}$.

Finally, we have that

$$G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} > -\frac{1}{2} \tag{C.11}$$

$$2G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} > -1 \tag{C.12}$$

$$G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} > -1 - G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}. \tag{C.13}$$

Since

$$1 + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} > \frac{1}{2} > 0,$$

we can divide both sides of (C.13) by $1 + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}$ to see that

$$\frac{G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}{1 + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}} > -1. \tag{C.14}$$

We realize that the term on the left side of (C.14) matches our expression for the the gradient of the score of the best response from (C.6), so we conclude that

$$\boldsymbol{\beta}^T\nabla_s\mathbf{x}^* > -1.$$

Thus, we have that $|\boldsymbol{\beta}^T\nabla_s\mathbf{x}^*| < 1$, so $\boldsymbol{\beta}^T\nabla_s\mathbf{x}^*$ is a contraction in s.

D Proofs of Mean-Field Results

We state technical lemmas that will be used in many of our results. The proofs of these lemmas can be found in Appendix G.

Lemma 42. *The distribution $P_{\boldsymbol{\beta},s}$ is given by*

$$P_{\boldsymbol{\beta},s}(r) = \int_{\mathcal{X}\times\mathcal{G}} G(r - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu))dF. \tag{D.1}$$

Under Assumptions 1, 2, and 3, if $\sigma^2 > \frac{1}{\alpha_*\sqrt{2\pi e}}$, then $P_{\boldsymbol{\beta},s}(r)$ is a well-defined function. Furthermore, it is strictly increasing in r, continuously differentiable in $\boldsymbol{\beta}, s, r$, and has a unique continuous inverse distribution function. Proof in Appendix G.11.

Lemma 43. *Fix $\boldsymbol{\beta} \in \mathcal{B}$. Suppose Assumptions 1, 2, and 3 hold. If $\sigma^2 > \frac{1}{\alpha_*\sqrt{2\pi e}}$, then $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s} < 1$. If $\sigma^2 > \frac{2}{\alpha_*\sqrt{2\pi e}}$, then $\left|\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}\right| < 1$. Proof in Appendix G.12.*

Lemma 44. *Let $\boldsymbol{\beta} \in \mathcal{B}$. Let P be the distribution over $\boldsymbol{\beta}^T(\boldsymbol{\eta} + \boldsymbol{\epsilon})$ where $\nu = (\boldsymbol{\eta}, \boldsymbol{\gamma}) \sim F$ and $\boldsymbol{\epsilon} \sim N(0, \sigma^2 I_d)$. Under Assumption 1, 2, 3, if $\sigma^2 > \frac{1}{\alpha_*\sqrt{2\pi e}}$, then*

$$\lim_{s\to\infty} P_{\boldsymbol{\beta},s} = P,$$

$$\lim_{s\to-\infty} P_{\boldsymbol{\beta},s} = P,$$

$$\lim_{s\to\infty} q(P_{\boldsymbol{\beta},s}) = q(P),$$

$$\lim_{s\to-\infty} q(P_{\boldsymbol{\beta},s}) = q(P).$$

Proof in Appendix G.13.

D.1 Proof of Theorem 4

Our main goal is show that $P_{\boldsymbol{\beta},s}(s)$ is a continuous and strictly increasing function of s. A continuous and strictly increasing function can intersect a horizontal line (e.g. $y = q$) in at most one point, so if a fixed point of $q(P_{\boldsymbol{\beta},s})$ exists, then it must be unique.

From Lemma 42, $P_{\boldsymbol{\beta},s}$ has a unique inverse. So, if there is a fixed point s, which means that $s = q(P_{\boldsymbol{\beta},s})$, then the fixed point satisfies $P_{\boldsymbol{\beta},s}(s) = q$.

Applying (D.1) from Lemma 42, we have

$$\begin{aligned} P_{\boldsymbol{\beta},s}(s) &= \int_{\mathcal{X}\times\mathcal{G}} G(s - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu))dF \\ &= \int_{\mathcal{X}\times\mathcal{G}} G(h(s;\boldsymbol{\beta},\nu))dF, \end{aligned}$$

where $h(s;\boldsymbol{\beta},\nu) = s - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$.

The continuity of $P_{\boldsymbol{\beta},s}(s)$ in s follows from the continuity of $P_{\boldsymbol{\beta},s}(r)$ in (s, r) (Lemma 42).

We show that $P_{\boldsymbol{\beta},s}(s)$ is strictly increasing in s. From Lemma 39, we have that $h(s;\boldsymbol{\beta},\nu)$ is strictly increasing in s for any agent type (ν). Since G is a strictly increasing CDF, so we have that $G(h(s;\boldsymbol{\beta},\nu))$ is also strictly increasing. Finally, the sum of strictly increasing functions is strictly increasing, which gives that the integral is also a strictly increasing function of s.

Since $P_{\boldsymbol{\beta},s}(s)$ is continuous and strictly increasing in s, there is at most one point where it can equal q. Thus, if a fixed point of $q(P_{\boldsymbol{\beta},s})$ exists, then it is unique.

D.2 Proof of Lemma 5

To show that $q(P_{\boldsymbol{\beta},s})$ is continuously differentiable in s, we first show that $q(P_{\boldsymbol{\beta},s})$ can be expressed implicitly as a solution to

$$h(s, r) = P_{\boldsymbol{\beta},s}(r) - q = 0, \text{ where } r = q(P_{\boldsymbol{\beta},s}). \tag{D.2}$$

Second, we verify that the Implicit Function Theorem (Theorem 19) can be applied to $h(s, r) = 0$, so that r can be expressed as a continuously differentiable function of s. Since $r = q(P_{\boldsymbol{\beta},s})$, we can conclude that $q(P_{\boldsymbol{\beta},s})$ is continuously differentiable in s.

For the first step, we aim to show that $q(P_{\boldsymbol{\beta},s})$ can be expressed by (D.2). By Lemma 42, we have that $P_{\boldsymbol{\beta},s}$ has a unique inverse distribution function. So, there exists a unique r such that $r = q(P_{\boldsymbol{\beta},s})$. Equivalently, $P_{\boldsymbol{\beta},s}(r) = q$ for $r = q(P_{\boldsymbol{\beta},s})$, which yields (D.2).

In the second step, we aim to apply Implicit Function Theorem to $h(s, r) = 0$ at any point (s_0, r_0) that satisfies $h(s, r) = 0$ to show that r can be expressed as a continuously differentiable function of s. Since $r = q(P_{\boldsymbol{\beta},s})$, this is sufficient for showing that $q(P_{\boldsymbol{\beta},s})$ is continuously differentiable function of s.

The conditions of the Implicit Function Theorem include that $h(s, r)$ is continuously differentiable in its arguments and that $\frac{\partial h}{\partial r}(s_0, r_0) \neq 0$. We verify that these conditions hold as follows. Both of these conditions follow from Lemma 42, which gives that $P_{\boldsymbol{\beta},s}(r)$ is continuously differentiable in (s, r) and strictly increasing in r. We have that

$$\frac{\partial h}{\partial r} = \frac{\partial P_{\boldsymbol{\beta},s}(r)}{\partial r},$$

and $\frac{\partial P_{\boldsymbol{\beta},s}(r)}{\partial r} > 0$. So, for any (s_0, r_0), we have that $\frac{\partial h}{\partial r}(s_0, r_0) \neq 0$.

As a result, the conditions of the Implicit Function Theorem are satisfied. Let (s_0, r_0) be any point that satisfies $h(s, r) = 0$. In an open neighborhood $V \times W \subset \mathbb{R} \times \mathbb{R}$ of (s_0, r_0), for each $s \in V$ there is a unique $\mathbf{g}(s) \in W$ such that $h(s, \mathbf{g}(s)) = 0$ and $\mathbf{g}$ is a continuously differentiable function of s. Since $r = q(P_{\boldsymbol{\beta},s})$ satisfies $h(s, r) = 0$, we must have that $q(P_{\boldsymbol{\beta},s})$ is a continuously differentiable function of s.

An analogous proof can be used to show that $q(P_{\boldsymbol{\beta},s})$ is continuously differentiable in $\boldsymbol{\beta}$.

D.3 Proof of Theorem 6

We aim to apply the Intermediate Value Theorem to the function $g(s) = s - q(P_{\boldsymbol{\beta},s})$ to show that $q(P_{\boldsymbol{\beta},s})$ has at least one fixed point. We note that by Lemma 5 that $g(s)$ is continuous. It remains to show that there exists s_l such that $g(s_l) < 0$ and there exists s_h such that $s_h > s_l$ and $g(s_h) > 0$. Then, by the Intermediate Value Theorem, there must be $s \in [s_l, s_h]$ for which $g(s) = 0$, which gives that $q(P_{\boldsymbol{\beta},s})$ has at least one fixed point.

First, by Lemma 44, we have that

$$\lim_{s\to\infty} q(P_{\boldsymbol{\beta},s}) = q(P)$$
$$\lim_{s\to-\infty} q(P_{\boldsymbol{\beta},s}) = q(P).$$

So, for some $s \geq S_1$ where $S_1 < \infty$, we have that $|q(P_{\boldsymbol{\beta},s}) - q(P)| < \delta$. Let $s_h = \max(q(P) + \delta, S_1)$. Then we have for $s \geq s_h$,

$$\begin{aligned} g(s) &= s - q(P_{\boldsymbol{\beta},s}) \\ &> s - q(P) - \delta \\ &\geq q(P) + \delta - q(P) - \delta \\ &= 0. \end{aligned}$$

Similarly, for some $s \leq S_2$ where $S_2 > -\infty$, we have that $|q(P_{\boldsymbol{\beta},s}) - q(P)| < \delta$. Let $s_l = \min(q(P) - \delta, S_2)$. Then we have for $s \leq s_l$,

$$\begin{aligned} g(s) &= s - q(P_{\boldsymbol{\beta},s}) \\ &< s - q(P) + \delta \\ &\leq q(P) - \delta - q(P) + \delta \\ &= 0. \end{aligned}$$

So, by the Intermediate Value Theorem there must be $s \in [s_l, s_h]$ for which $g(s) = 0$, which gives that $q(P_{\boldsymbol{\beta},s})$ has at least one fixed point.

D.4 Proof of Corollary 7

Since we assumed that $q(P_{\boldsymbol{\beta},s})$ is a contraction in s and $q(P_{\boldsymbol{\beta},s}) : \mathbb{R} \to \mathbb{R}$, then we can apply Banach's Fixed Point Theorem (Theorem 23) to conclude that the process in (3.1) converges to s^*, the unique fixed point of $q(P_{\boldsymbol{\beta},s})$.

D.5 Proof of Corollary 8

If $\sigma^2 > \frac{2}{\alpha_* \sqrt{2\pi e}}$, we can apply the second part of Lemma 43 to conclude that $|\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}| < 1$. By Lemma 22, $q(P_{\boldsymbol{\beta},s})$ is a contraction in s. As a consequence of Theorem 7, we can conclude that fixed point iteration (3.1) converges to s^*, the unique fixed point of $q(P_{\boldsymbol{\beta},s})$.

D.6 Proof of Corollary 9

First, we show that we can a define a function that maps a linear model $\boldsymbol{\beta}$ to the equilibrium threshold s^* induced by $\boldsymbol{\beta}$. Second, we give an equation that implicitly expresses this function. We verify that this equation satisfies the conditions of the Implicit Function Theorem at any point $(\boldsymbol{\beta}, s^*)$, where $\boldsymbol{\beta}$ is a linear model and s^* is equilibrium threshold induced by $\boldsymbol{\beta}$, and apply the Implicit Function Theorem to arrive at the desired result. To prove one of the conditions of the Implicit Function Theorem, we will use the first part of Lemma 43.

Recall that for every $\boldsymbol{\beta} \in \mathcal{B}$, there exists a fixed point s^* that satisfies $q(P_{\boldsymbol{\beta},s^*}) = s^*$ (Theorem 6), and it is unique (Theorem 4). As a result, we can define a function $s : \mathcal{B} \to \mathbb{R}$ that maps $\boldsymbol{\beta}$ to the fixed point induced by $\boldsymbol{\beta}$.

Note that we can implicitly represent $s(\boldsymbol{\beta})$ by s in the following equation

$$h(\boldsymbol{\beta}, s) = s - q(P_{\boldsymbol{\beta},s}) = 0.$$

We aim to apply the Implicit Function Theorem to $h(\boldsymbol{\beta}, s)$ at any point $(\boldsymbol{\beta}_0, s)$ where $h(\boldsymbol{\beta}_0, s_0) = 0$. We verify that the conditions of the Implicit Function Theorem are satisfied–$h(\boldsymbol{\beta}, s)$ must be continuously differentiable in its arguments and $\frac{\partial h(\boldsymbol{\beta},s)}{\partial s}(\boldsymbol{\beta}_0, s_0) \neq 0$.

For the first condition, Lemma 5 gives us that $q(P_{\boldsymbol{\beta},s})$ is continuously differentiable in its arguments. As a result, $h(\boldsymbol{\beta}, s)$ is also continuously differentiable in $\boldsymbol{\beta}, s$.

For the second condition, we note that

$$\frac{\partial h(\boldsymbol{\beta}, s)}{\partial s} = 1 - \frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}.$$

From Lemma 43, we have that $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s} < 1$, so $\frac{\partial h(\boldsymbol{\beta},s)}{\partial s} > 0$. Thus, the conditions of the Implicit Function Theorem are satisfied.

Let $(\boldsymbol{\beta}_0, s_0)$ be a point that yields $h(\boldsymbol{\beta}_0, s_0) = 0$. In a open neighborhood $V \times W \subset \mathbb{R}^d \times \mathbb{R}$ of $(\boldsymbol{\beta}_0, s_0)$, for every $\boldsymbol{\beta} \in V$, there is a unique $g(\boldsymbol{\beta}) \in W$ such that $h(\boldsymbol{\beta}, g(\boldsymbol{\beta})) = 0$ and g is a continuously differentiable function of $\boldsymbol{\beta}$. We note that such $g(\boldsymbol{\beta})$ must correspond to the unique equilibrium threshold induced by $\boldsymbol{\beta}$, so $s(\boldsymbol{\beta}) = g(\boldsymbol{\beta})$. Thus, $s(\boldsymbol{\beta})$ is a continuously differentiable function of $\boldsymbol{\beta}$.

E Proofs of Finite Approximation Results

Lemma 45. *Suppose the conditions of Theorem 11. Let $\{z^t\}$ be a sequence of random variables where*

$$z^t = \begin{cases} \epsilon_g & w.p.\ p_n(\epsilon_g) \\ S_k & w.p.\ \frac{1-p_n(\epsilon_g)}{2^k}, k \geq 1, \end{cases}$$

where $p_n(\epsilon_g)$ is the bound from Lemma 10 and

$$S_k = \sqrt{\frac{1}{2nD^2} \cdot \log\Big(\frac{2^{k+1}}{1 - p_n(\epsilon_g)}\Big)}. \tag{E.1}$$

For any $s \in \mathbb{R}$, z^t stochastically dominates $|q(P^n_{\boldsymbol{\beta},s}) - q(P_{\boldsymbol{\beta},s})|$. Proof in Appendix G.14.

Lemma 46. *Suppose the conditions of Theorem 11 hold. Let $\{\hat{s}^t_n\}_{t\geq 0}$ be a stochastic process generated via (4.1). Let $\{z^t\}_{t\geq 1}$ be a sequence of random variables where*

$$z^t = \begin{cases} \epsilon_g & w.p.\ p_n(\epsilon_g) \\ S_k & w.p.\ \frac{1-p_n(\epsilon_g)}{2^k}, k \geq 1, \end{cases}$$

where $p_n(\epsilon_g)$ is the bound from Lemma 10 and S_k is as defined in Lemma 45. Let κ be the Lipschitz constant of $q(P_{\boldsymbol{\beta},s})$. Then $\sum_{i=0}^k z^{t-i}\kappa^i + \kappa^k|\hat{s}^{t-k}_n - s^*|$ stochastically dominates $|\hat{s}^t_n - s^*|$. Proof in Appendix G.15.

E.1 Proof of Lemma 10

We define notation that will be used in the rest of the proof. For agent type $\nu \in \text{supp}(F)$, s^*_ν to be the threshold $s \in \mathbb{R}$ that maximizes its best response function (without noise) $\mathbf{x}(\boldsymbol{\beta}, s, \nu)$. Let $s_L = \inf_{\nu\in\text{supp}(F)} s^*_\nu$ and let $s_H = \sup_{\nu\in\text{supp}(F)} s^*_\nu$. We also can define functions

$$f_1(s) := P_{\boldsymbol{\beta},s}(q(P_{\boldsymbol{\beta},s}) + \epsilon) - q$$
$$f_2(s) := q - P_{\boldsymbol{\beta},s}(q(P_{\boldsymbol{\beta},s}) - \epsilon).$$

We define

$$M_\epsilon = \inf_{s\in\mathbb{R}} \min\{f_1(s), f_2(s)\},$$

and we aim to show that $M_\epsilon > 0$. We note that

$$M_\epsilon = \min\{\inf_{s\in\mathbb{R}} f_1(s), \inf_{s\in\mathbb{R}} f_2(s)\}.$$

We note that

$$\inf_{s\in\mathbb{R}} f_1(s) = \min\{\inf_{s<s_L} f_1(s), \inf_{s>s_H} f_1(s), \inf_{s\in[s_L,s_H]} f_1(s)\}.$$

The differentiability of $P_{\boldsymbol{\beta},s}$ and $q(P_{\boldsymbol{\beta},s})$ in s is given by Lemma 42 and Lemma 5, respectively. Thus, we can write that

$$\frac{df_1}{ds} = p_{\boldsymbol{\beta},s}(q(P_{\boldsymbol{\beta},s}) + \epsilon) \cdot \frac{dq(P_{\boldsymbol{\beta},s})}{ds}.$$

By Lemma 41, the score of the best response $\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ for each agent ν is increasing on $s \le s_L$ and decreasing on $s \ge s_H$. By Lemma 43, $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}$ is a convex combination of $\boldsymbol{\beta}^T\nabla_s\mathbf{x}$. Thus, $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}$ is positive for $s < s_L$ and negative on $s > s_H$. This implies that $\frac{df_1}{ds}$ is positive for $s < s_L$ and negative on $s > s_H$. Thus, $f_1(s)$ is increasing on $(-\infty, s_L)$ and $f_1(s)$ is decreasing on (s_H, ∞). So,

$$\begin{aligned}\inf_{s<s_L} f_1(s) &= \lim_{s\to-\infty} f_1(s)\\ &= \lim_{s\to-\infty} P_{\boldsymbol{\beta},s}(q(P_{\boldsymbol{\beta},s}) + \epsilon) - q\\ &= P(q(P) + \epsilon) - q,\end{aligned}$$

where the last line follows from Lemma 44 and P is the distribution defined that lemma. Similarly, we can show that

$$\inf_{s>s_H} f_1(s) = \lim_{s\to\infty} f_1(s) = P(q(P) + \epsilon) - q.$$

Finally, because $[s_L, s_H]$ is a compact set, there is some $s_1 \in [s_L, s_H]$ for which $f_1(s)$ achieves its infimum on the interval. Thus,

$$\inf_{s\in\mathbb{R}} f_1(s) = \min\{P(q(P) + \epsilon) - q, P_{\boldsymbol{\beta},s_1}(q(P_{\boldsymbol{\beta},s_1}) + \epsilon) - q\}.$$

Thus, $\inf_{s\in\mathbb{R}} f_1(s) > 0$.

Similarly, we can compute $\inf_{s\in\mathbb{R}} f_2(s)$. We note that

$$\inf_{s\in\mathbb{R}} f_2(s) = \min\{\inf_{s<s_L} f_2(s), \inf_{s>s_H} f_2(s), \inf_{s\in[s_L,s_H]} f_2(s)\}.$$

We can write that

$$\frac{df_2}{ds} = -p_{\boldsymbol{\beta},s}(q(P_{\boldsymbol{\beta},s}) - \epsilon) \cdot \frac{dq(P_{\boldsymbol{\beta},s})}{ds}.$$

From this result, we can see that $f_2(s)$ is decreasing on $(-\infty, s_L)$ and $f_2(s)$ is increasing on (s_H, ∞). So,

$$\begin{aligned}\inf_{s<s_L} f_2(s) &= \lim_{s\to s_L} f_2(s)\\ &= f_2(s_L)\\ \inf_{s>s_H} f_2(s) &= \lim_{s\to s_H} f_2(s)\\ &= f_2(s_H).\end{aligned}$$

In addition, because $[s_L, s_H]$ is a compact set, there is some $s_2 \in [s_L, s_H]$ for which $f_2(s)$ achieves its infimum on the interval. Thus,

$$\inf_{s\in\mathbb{R}} f_2(s) = \min_{\{s_L,s_H,s_2\}} f_2(s),$$

so $\inf_{s\in\mathbb{R}} f_2(s) > 0$. Thus, $M_\epsilon > 0$.

Now, we proceed to show the second component of the lemma. From Theorem 24, we have that

$$P(|q(P_{\boldsymbol{\beta},s}) - q(P^n_{\boldsymbol{\beta},s})| < \epsilon) \ge 1 - 4e^{-2nM^2_{\epsilon,s}},$$

where $M_{\epsilon,s} = \min\{f_1(s), f_2(s)\}$. We can obtain a bound that is uniform over s by realizing that $M_\epsilon = \inf_{s\in\mathbb{R}} \min\{f_1(s), f_2(s)\}$ and $M_\epsilon > 0$. So, we have that

$$P(|q(P_{\boldsymbol{\beta},s}) - q(P^n_{\boldsymbol{\beta},s})| < \epsilon) \ge 1 - 4e^{-2nM^2_\epsilon}.$$

E.2 Proof of Theorem 11

Let $\{z^t\}_{t\geq 1}$ be a sequence of random variables where

$$z^t = \begin{cases} \epsilon_g & \text{w.p. } p_n(\epsilon_g) \\ S_k & \text{w.p. } \frac{1-p_n(\epsilon_g)}{2^k}, k \geq 1 \end{cases},$$

where $\epsilon_g = \frac{\epsilon(1-\kappa)}{2}$ and $p_n(\epsilon_g)$ is the bound from Lemma 10. We have that

$$|\hat{s}_n^t - s^*| \preceq_{\text{SD}} \sum_{i=0}^{t} z^{t-i}\kappa^i + \kappa^t|\hat{s}_n^0 - s^*| \tag{E.2}$$

$$\preceq_{\text{SD}} \sum_{i=0}^{t} z^{t-i}\kappa^i + \kappa^t S. \tag{E.3}$$

(E.2) follows from Lemma 46. (E.3) follows from the definition of S.

We note that

$$\begin{aligned}\sum_{i=0}^{t} \epsilon_g\kappa^i &< \sum_{i=0}^{\infty} \epsilon_g\kappa^i \\ &= \frac{\epsilon_g}{1-\kappa} \\ &= \frac{\epsilon(1-\kappa)}{2} \cdot \frac{1}{1-\kappa}. \\ &= \frac{\epsilon}{2}.\end{aligned}$$

In addition, let $t \geq \left\lceil \frac{\log(\frac{\epsilon}{2S})}{\log \kappa} \right\rceil$. For such t, we have that

$$t \geq \frac{\log(\frac{\epsilon}{2S})}{\log \kappa}.$$

Rearranging the above inequality gives

$$\kappa^t S \leq \frac{\epsilon}{2}.$$

As a result, we have that

$$\sum_{i=0}^{t} \epsilon_g\kappa^i + \kappa^t S < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

By (E.3) and the definition of stochastic dominance, we have that

$$\begin{aligned}P(|\hat{s}_n^t - s^*| \leq \epsilon) &\geq P(\sum_{i=0}^{t} z^{t-i}\kappa^i + \kappa^t S \leq \epsilon) \\ &\geq P(z^{t-i} = \epsilon_g \text{ for } i = 0\ldots t) \\ &\geq (p_n(\epsilon_g))^t\end{aligned}$$

If we have that

$$n \geq \frac{1}{2M_{\epsilon_g}^2} \log\left(\frac{4t}{\delta}\right), \tag{E.4}$$

then we can show that $p_n(\epsilon_g) \geq 1 - \frac{\delta}{t}$. We can rearrange (E.4)

$$e^{-2nM_{\epsilon_g}^2} \leq \frac{\delta}{4t}.$$

Rearranging again,

$$1 - 4e^{-2nM_{\epsilon_g}^2} \geq 1 - \frac{\delta}{t}.$$

Thus, we have that

$$p_n(\epsilon_g) \geq 1 - \frac{\delta}{t}.$$

So, $(p_n(\epsilon_g))^t \geq (1 - \frac{\delta}{t})^t$. Applying Theorem 25 gives that $(p_n(\epsilon_g))^t \geq 1 - \delta$. Therefore, we conclude that if $t \geq \left\lceil \frac{\log(\frac{\epsilon}{2S})}{\log \kappa} \right\rceil$ and $n \geq \frac{1}{2M_{\epsilon_g}^2} \log\left(\frac{4t}{\delta}\right)$, then

$$P(|\hat{s}_n^t - s^*| \leq \epsilon) \geq 1 - \delta,$$

as desired.

E.3 Proof of Corollary 12

To show that $\hat{s}_n^{t_n} \xrightarrow{p} s^*$, we must show that

$$\lim_{n \to \infty} P(|\hat{s}_n^{t_n} - s^*| > \epsilon) = 0.$$

It is sufficient to show that for any $\delta > 0$, there exists N such that for $n \geq N$,

$$P(|\hat{s}_n^{t_n} - s^*| > \epsilon) \leq \delta.$$

As in the statement of Theorem 11, let $S = |\hat{s}_n^0 - s^*|$. Let $N_1 \in \mathbb{N}$ be the smallest value of n such that $t_n \geq \lceil \frac{\log(\frac{\epsilon}{2S})}{\log \kappa} \rceil$.

We have that $t_n \prec \exp(n)$. So, there exists $N_2 \in \mathbb{N}$ such that for $n \geq N_2$,

$$t_n \leq \frac{\delta}{4} \exp\left(2nM_{\epsilon_g}^2\right).$$

Rearranging this equation, we have that for $n \geq N_2$,

$$\exp(2nM_{\epsilon_g}^2) \geq \frac{4t_n}{\delta}.$$

Taking log of both sides yields for $n \geq N_2$

$$2nM_{\epsilon_g}^2 \geq \log(\frac{4t_n}{\delta}).$$

So, for $n \geq N_2$, we have that

$$n \geq \frac{1}{2M_{\epsilon_g}^2} \log(\frac{4t_n}{\delta}).$$

We can take $N = \max\{N_1, N_2\}$. By Theorem 11, we have that for $n \geq N, P(|\hat{s}_n^{t_n} - s^*| > \epsilon) < \delta$. Thus, we have that $\hat{s}_n^{t_n} \xrightarrow{p} s^*$.

F Proofs of Learning Results

We state technical lemmas that will be used in many of our learning results.

Lemma 47. *Let $\boldsymbol{\beta} \in \mathcal{B}$. Let s^* be the mean-field equilibrium threshold. Define a truncated stochastic fixed point iteration process*

$$\hat{s}_n^{t+1} = \begin{cases} q(P^n_{\boldsymbol{\beta},\hat{s}_n^t}) & q(P^n_{\boldsymbol{\beta},\hat{s}_n^t}) \in \mathcal{S} \\ -D & q(P^n_{\boldsymbol{\beta},\hat{s}_n^t}) < -D \\ D & q(P^n_{\boldsymbol{\beta},\hat{s}_n^t}) > D \end{cases}. \tag{F.1}$$

Under Assumptions 1, 2, 3, if $\sigma^2 > \frac{2}{\alpha_* \sqrt{2\pi e}}$, then for any sequence $\{t_n\}$ such that $t_n \uparrow \infty$ as $n \to \infty$ and $t_n \prec \exp(n)$, we have that $\hat{s}_n^{t_n} \xrightarrow{p} s^*$.

Lemma 48. *Let $\tilde{\pi}$ and $\tilde{\ell}$ be functions $\mathcal{X} \times \mathcal{G} \times supp(G) \times \mathcal{B} \times \mathcal{S} \times \mathcal{S} \to \mathbb{R}$.*

$$\tilde{\pi}(\nu, \boldsymbol{\beta}, s, r) = \pi(\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) + \boldsymbol{\epsilon}; \boldsymbol{\beta}, r).$$
$$\tilde{\ell}(\nu, \boldsymbol{\beta}, s, r) = \ell(\pi(\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) + \boldsymbol{\epsilon}; \boldsymbol{\beta}, r)), \nu),$$
$$\tilde{k}(\nu, \epsilon, \boldsymbol{\beta}, s, r) = \mathbb{I}\Big(\frac{r - \boldsymbol{\beta}^T(\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) + \boldsymbol{\epsilon})}{h} \in \Big[-\frac{1}{2}, \frac{1}{2}\Big)\Big)$$

Let $(\nu, \boldsymbol{\epsilon})$ represent the data $\mathbf{w}$ and $(\boldsymbol{\beta}, s, r)$ represent parameters $\boldsymbol{\theta}$. Suppose the conditions of Theorem 14 hold. With $(\nu, \boldsymbol{\epsilon})$ i.i.d., $\tilde{\ell}, \tilde{\pi}$, and $\tilde{k}$ satisfy the requirements on the function $a(\mathbf{w}; \boldsymbol{\theta})$ from Lemma 26. Proof in Appendix G.17.

Lemma 49. *Let $\boldsymbol{\beta} \in \mathcal{B}, s \in \mathcal{S}, \boldsymbol{\zeta} \in \{-1, 1\}^d, \zeta \in \{-1, 1\}$, and $b > 0$. Define $T : (\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta) \to (\nu'_{b,\boldsymbol{\zeta},\zeta}, c_{\nu'})$, where $\nu = (\boldsymbol{\eta}, \boldsymbol{\gamma}) \in \mathcal{X} \times \mathcal{G}$ and cost function c_ν satisfies Assumption 1. Let*

$$\mathbf{x}_1 := \mathbf{x}^*(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu)$$
$$r := (\boldsymbol{\beta} + b\boldsymbol{\zeta})^T \mathbf{x}_1 - b\zeta.$$

Let $\nu'_{b,\boldsymbol{\zeta},\zeta} = (\boldsymbol{\eta}'_{b,\boldsymbol{\zeta},\zeta}, \boldsymbol{\gamma})$ and cost function $c_{\nu'}$ defined as follows.

$$\boldsymbol{\eta}'_{b,\boldsymbol{\zeta},\zeta} := \boldsymbol{\eta} + \boldsymbol{\beta} \cdot b \cdot (\boldsymbol{\zeta}^T \mathbf{x}_1 - \zeta) \tag{F.2}$$
$$c_{\nu'}(\mathbf{y}) := c_\nu(\mathbf{y}) - G(s - r)\boldsymbol{\beta}^T \mathbf{y}. \tag{F.3}$$

If $\boldsymbol{\eta}, \mathbf{x}_1 \in Int(\mathcal{X})$ and b sufficiently small, then $\nu'_{b,\boldsymbol{\zeta},\zeta} \in \mathcal{X} \times \mathcal{G}$, $c_{\nu'}$ is α_ν-strongly convex,

$$\mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) = \mathbf{x}_1 + b \cdot \boldsymbol{\beta}(\boldsymbol{\zeta}^T \mathbf{x}_1 - \zeta),$$

$\mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) \in Int(\mathcal{X})$, and $\boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) = r$. In other words, when the agent with type and cost function $T(\nu, c_\nu)$ best responds to the unperturbed model $\boldsymbol{\beta}$ and threshold s, they obtain the same raw score (without noise) as the agent with type and cost function (ν, c_ν) who responds to a perturbed model $\boldsymbol{\beta} + b\zeta$ and threshold $s + b\zeta$. Proof in Appendix G.18.

Lemma 50. *Suppose the conditions of Theorem 14 hold. Fix $\boldsymbol{\beta} \in \mathcal{B}, s \in \mathcal{S}$. For sufficiently small b, there exists a distribution over agent types, $\tilde{F}_b$ and corresponding cost functions $c_{\nu'}$ for each type $\nu' \sim \tilde{F}_b$ such that when agents with types $\nu' \sim \tilde{F}_b$ and cost functions $c_{\nu'}$ best respond to the unperturbed model $\boldsymbol{\beta}$ and threshold s the induced score distribution is equal to $P_{\boldsymbol{\beta},s,b}$. Furthermore, the support of $\tilde{F}_b$ is contained in $\mathcal{X} \times \mathcal{G}$, each $c_{\nu'}$ satisfies Assumption 1, $\tilde{F}_b$ has a finite number of agent types, $\alpha_*(\tilde{F}_b) = \alpha_*(F)$, and for any agent type $\nu' \sim \tilde{F}_b$, we have $\mathbf{x}(\boldsymbol{\beta}, s, \nu') \in Int(\mathcal{X})$. Proof in Appendix G.19.*

Lemma 51. *Fix $\boldsymbol{\beta} \in \mathcal{B}$. Suppose the conditions of Theorem 14 hold. If b is sufficiently small, then $q(P_{\boldsymbol{\beta},s,b})$ has a unique fixed point $s(\boldsymbol{\beta}, b)$. As $b \to 0$, $s(\boldsymbol{\beta}, b) \to s(\boldsymbol{\beta})$, where $s(\boldsymbol{\beta})$ is the unique fixed point of $q(P_{\boldsymbol{\beta},s})$. Proof in Appendix G.20.*

Lemma 52. *Fix $\boldsymbol{\beta} \in \mathcal{B}$. Let $\{t_n\}$ be a sequence such that $t_n \uparrow \infty$ as $n \to \infty$. Under the conditions of Theorem 14, if b sufficiently small,*

$$\hat{s}_{b,n}^{t_n} \xrightarrow{p} s(\boldsymbol{\beta}, b), \quad \hat{s}_{b,n}^{t_n+1} \xrightarrow{p} s(\boldsymbol{\beta}, b)$$

where $s(\boldsymbol{\beta}, b)$ is the unique fixed point of $q(P_{\boldsymbol{\beta},s,b})$. Proof in Appendix G.21.

F.1 Proof of Lemma 13

Let $\Delta\ell(\nu) = \ell(1, \nu) - \ell(0, \nu)$. We have that

$$\begin{aligned}
L_{\text{eq}}(\boldsymbol{\beta}) =& L(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})) \\
=& \mathbb{E}_{\boldsymbol{\epsilon},\nu}\left[\ell(\pi(\mathbf{x}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), \nu); \boldsymbol{\beta}, s(\boldsymbol{\beta})), \nu)\right] \\
=& \mathbb{E}_{\boldsymbol{\epsilon},\nu}\left[\ell(\mathbb{I}(\boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), \nu) \geq s(\boldsymbol{\beta})), \nu)\right] \\
=& \mathbb{E}_{\nu}\left[\mathbb{E}_{\boldsymbol{\epsilon}|\nu}\left[\ell(1,\nu)\cdot\mathbb{I}(\boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), \nu) \geq s(\boldsymbol{\beta}))\right]\right] \\
&+ \mathbb{E}_{\nu}\left[\mathbb{E}_{\boldsymbol{\epsilon}|\nu}\left[\ell(0,\nu)\cdot\mathbb{I}(\boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), \nu) < s(\boldsymbol{\beta}))\right]\right] \\
=& \mathbb{E}_{\nu}\left[\ell(1,\nu)(1 - G(s(\boldsymbol{\beta}) - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s(\boldsymbol{\beta}), \nu))\right] \\
&+ \mathbb{E}_{\nu\sim F}\left[\ell(0,\nu)G(s(\boldsymbol{\beta}) - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s(\boldsymbol{\beta}), \nu))\right] \\
=& \mathbb{E}_{\nu\sim F}\left[\ell(1,\nu) - \Delta\ell(\nu)G(s(\boldsymbol{\beta}) - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s(\boldsymbol{\beta}), \nu))\right].
\end{aligned}$$

Under the assumed conditions, we have that $\mathbf{x}^*$ is continuously differentiable in its first and second arguments by Lemma 2. We also have that s is continuously differentiable $\boldsymbol{\beta}$ by Corollary 9. Thus, $L_{\text{eq}}(\boldsymbol{\beta})$ continuously differentiable in $\boldsymbol{\beta}$.

F.2 Proof of Theorem 14

Let $s(\boldsymbol{\beta})$ be the unique fixed point of $q(P_{\boldsymbol{\beta},s})$. We introduce the following quantities.

$$\begin{aligned}
\tilde{\ell}_i(\boldsymbol{\beta}, s, r) &:= \ell(\pi(\mathbf{x}^*(\boldsymbol{\beta}, s, \nu_i) + \boldsymbol{\epsilon}_i; \boldsymbol{\beta}, r), \nu_i) \\
\tilde{\ell}_n(\boldsymbol{\beta}, s, r) &:= \frac{1}{n}\sum_{i=1}^{n}\tilde{\ell}_i(\boldsymbol{\beta}, s, r). \\
\tilde{L}(\boldsymbol{\beta}, s, r) &:= \mathbb{E}_{\nu,\boldsymbol{\epsilon}}\left[\tilde{\ell}_i(\boldsymbol{\beta}, s, r)\right]
\end{aligned}$$

We note that

$$\boldsymbol{\ell}_i(\boldsymbol{\beta}, s, r) = \tilde{\boldsymbol{\ell}}_i(\boldsymbol{\beta} + b_n\boldsymbol{\zeta}_i, s + b_n\zeta_i, r + b_n\zeta_i),$$

where $\boldsymbol{\zeta}_i$ and ζ_i are the perturbations applied to agent i. When $s = r = s(\boldsymbol{\beta})$, $\tilde{L}(\boldsymbol{\beta}, s, r) = L_{\text{eq}}(\boldsymbol{\beta})$. Through an identical argument as in the proof of Lemma 13, we can see that $\tilde{L}(\boldsymbol{\beta}, s, r)$ is continuously differentiable in s and r.

The model effect estimator $\hat{\tau}^{t_n}_{\text{ME},n}(\boldsymbol{\beta})$ is the regression coefficient obtained by running OLS of $\boldsymbol{\ell}(\boldsymbol{\beta}, \hat{s}^{t_n}_n, \hat{s}^{t_n+1}_n)$ on $\mathbf{Z}$. The regression coefficient must have the following form.

$$\hat{\tau}^{t_n}_{\text{ME},n}(\boldsymbol{\beta}) = (\mathbf{S}^n_{zz})^{-1}\mathbf{s}^n_{zy}, \text{ where } \mathbf{S}^n_{zz} := \frac{1}{b_n^2 n}\mathbf{Z}^T\mathbf{Z}, \quad \mathbf{s}^n_{zy} := \frac{1}{b_n^2 n}\mathbf{Z}^T\boldsymbol{\ell}(\boldsymbol{\beta}, \hat{s}^{t_n}_{b_n,n}, \hat{s}^{t_n+1}_{b_n,n}). \tag{F.4}$$

In this proof, we establish convergence in probability of the two terms above separately. The bulk of the proof is the first step, which entails showing that

$$\mathbf{s}^n_{zy} \xrightarrow{p} \frac{\partial L}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})).$$

Due to $\boldsymbol{\ell}$'s dependence on the stochastic processes $\{\hat{s}^{t_n}_{b_n,n}\}$ and $\{\hat{s}^{t_n+1}_{b_n,n}\}$, the main workhorse of this result is Lemma 28. To apply this lemma, we must establish stochastic equicontinuity for the collection of stochastic processes $\{\tilde{\ell}_n(\boldsymbol{\beta}, s, r)\}$. Second, through a straightforward application of the Weak Law of Large Numbers, we show that

$$\mathbf{S}_{zz} \xrightarrow{p} \mathbf{I}_d.$$

Finally, we use Slutsky's Theorem to establish the convergence of the model effect estimator.

We proceed with the first step of establishing convergence of $\mathbf{s}_{zy}^n$. We have that

$$\begin{aligned}\mathbf{s}_{zy}^n &= \frac{1}{b_n^2 n}\mathbf{Z}^T \boldsymbol{\ell}(\boldsymbol{\beta}, \hat{s}_{b_n,n}^{t_n}, \hat{s}_{b_n,n}^{t_n+1}) \\ &= \frac{1}{b_n^2 n}\sum_{i=1}^n b_n \boldsymbol{\zeta}_i \boldsymbol{\ell}_i(\boldsymbol{\beta}, \hat{s}_{b_n n}^{t_n}, \hat{s}_{b_n,n}^{t_n+1}) \\ &= \frac{1}{b_n}\cdot\frac{1}{n}\sum_{i=1}^n \boldsymbol{\zeta}_i \boldsymbol{\ell}_i(\boldsymbol{\beta}, \hat{s}_{b_n,n}^{t_n}, \hat{s}_{b_n,n}^{t_n+1}).\end{aligned}$$

We fix j and $b_n = b$ where $b > 0$ and is small enough to satisfy the hypothesis of Lemma 52. For each $\boldsymbol{\zeta} \in \{-1,1\}^d$ and $\zeta \in \{-1,1\}$, let

$$n_{\boldsymbol{\zeta},\zeta} = \sum_{i=1}^n \mathbb{I}(\boldsymbol{\zeta}_i = \boldsymbol{\zeta}, \zeta_i = \zeta).$$

Let $z(\boldsymbol{\zeta})$ map a perturbation $\boldsymbol{\zeta} \in \{-1,1\}^d$ to the identical vector $\boldsymbol{\zeta}$, except with j-th entry set to 0. So, if the j-th entry of $\boldsymbol{\zeta}$ is 1, then $\boldsymbol{\zeta} = \mathbf{e}_j + z(\boldsymbol{\zeta})$. If the j-th entry of $\boldsymbol{\zeta}$ is -1, then $\boldsymbol{\zeta} = -\mathbf{e}_j + z(\boldsymbol{\zeta})$. So, we have that

$$\begin{aligned}\boldsymbol{\ell}_i(\boldsymbol{\beta}, \hat{s}_{b,n}^{t_n}, \hat{s}_{b,n}^{t_n+1}) &= \tilde{\ell}_i(\boldsymbol{\beta} + b\boldsymbol{\zeta}_i, \hat{s}_n^{t_n} + b\zeta_i, \hat{s}_{b,n}^{t_n+1} + b\zeta_i) \\ &= \tilde{\ell}_i(\boldsymbol{\beta} + b\boldsymbol{\zeta}_{i,j}\mathbf{e}_j + b\cdot z(\boldsymbol{\zeta}_i), \hat{s}_n^{t_n} + b\zeta_i, \hat{s}_{b,n}^{t_n+1} + b\zeta_i).\end{aligned}$$

As a result, we have that

$$\frac{1}{n}\sum_{i=1}^n \boldsymbol{\zeta}_{i,j}\boldsymbol{\ell}_i(\boldsymbol{\beta}, \hat{s}_{b,n}^{t_n}, \hat{s}_{b,n}^{t_n+1}) \tag{F.5}$$

$$=\frac{1}{n}\sum_{i=1}^n \boldsymbol{\zeta}_{i,j}\cdot\tilde{\ell}_i(\boldsymbol{\beta} + b\boldsymbol{\zeta}_{i,j}\mathbf{e}_j + b\cdot z(\boldsymbol{\zeta}_i), \hat{s}_{b,n}^{t_n} + b\zeta_i, \hat{s}_{b,n}^{t_n+1} + b\zeta_i) \tag{F.6}$$

$$= \sum_{\substack{\boldsymbol{\zeta}\in\{-1,1\}^d \text{ s.t. } \boldsymbol{\zeta}_j = 1 \\ \zeta\in\{-1,1\}}} \frac{n_{\boldsymbol{\zeta},\zeta}}{n}\sum_{i=1}^{n_{\boldsymbol{\zeta},\zeta}} \tilde{\ell}_i(\boldsymbol{\beta} + b\mathbf{e}_j + b\cdot z(\boldsymbol{\zeta}), \hat{s}_{b,n}^{t_n} + b\zeta, \hat{s}_{b,n}^{t_n+1} + b\zeta) \tag{F.7}$$

$$- \sum_{\substack{\boldsymbol{\zeta}\in\{-1,1\}^d \text{ s.t. } \boldsymbol{\zeta}_j = -1 \\ \zeta\in\{-1,1\}}} \frac{n_{\boldsymbol{\zeta},\zeta}}{n}\sum_{i=1}^{n_{\boldsymbol{\zeta},\zeta}} \tilde{\ell}_i(\boldsymbol{\beta} - b\mathbf{e}_j + b\cdot z(\boldsymbol{\zeta}), \hat{s}_{b,n}^{t_n} + b\zeta, \hat{s}_{b,n}^{t_n+1} + b\zeta) \tag{F.8}$$

To establish convergence properties of each term in the double sum in (F.7) and (F.8), we must establish stochastic equicontinuity of the collection of stochastic processes $\{\tilde{\ell}_n(\boldsymbol{\beta}, s, r)\}$ indexed by $(s,r) \in \mathcal{S}\times\mathcal{S}$. Because $\mathcal{S}\times\mathcal{S}$ compact and $\tilde{L}(\boldsymbol{\beta}, s, r)$ is continuous in s and r, we can show that $\{\tilde{\ell}_n(\boldsymbol{\beta}, s, r)\}$ is stochastically equicontinuous by establishing that $\tilde{\ell}_n(\boldsymbol{\beta}, s, r)$ converges uniformly in probability (with respect to (s,r)) to $\tilde{L}(\boldsymbol{\beta}, s, r)$ (Lemma 27). We will show that $\tilde{\ell}_n(\boldsymbol{\beta}, s, r)$ converges uniformly (with respect to (s,r)) in probability to $\tilde{L}(\boldsymbol{\beta}, s, r)$ via Lemma 26.

By Lemma 48, we have that $\tilde{\ell}$ satisfies the conditions of Lemma 26. Thus, we can apply Lemma 26 to establish uniform convergence in probability of $\tilde{\ell}_n(\boldsymbol{\beta}, s, r)$ with respect to (s,r). As a consequence, the collection of stochastic processes $\{\tilde{\ell}_n(\boldsymbol{\beta}, s, r)\}$ is stochastically equicontinuous. In particular, $\tilde{\ell}_n(\boldsymbol{\beta}, s, r)$ is stochastically equicontinuous at $(s(\boldsymbol{\beta}, b), s(\boldsymbol{\beta}, b))$, where $s(\boldsymbol{\beta}, b)$ is the unique fixed point of $q(P_{\boldsymbol{\beta},s,b})$ (see Lemma 51). By Lemma 52, we have that

$$\begin{aligned}\hat{s}_{b,n}^{t_n} &\xrightarrow{p} s(\boldsymbol{\beta}, b) \\ \hat{s}_{b,n}^{t_n+1} &\xrightarrow{p} s(\boldsymbol{\beta}, b).\end{aligned}$$

Now, we can apply Lemma 28 to establish convergence in probability for each term in the double sum of (F.7), (F.8). As an example, for a perturbation $\boldsymbol{\zeta} \in \{-1,1\}^d$ with j-th entry equal to 1 and arbitrary $\zeta \in \{-1,1\}$, Lemma 28 gives that

$$\begin{aligned}
&\tilde{\ell}_{n_{\boldsymbol{\zeta},\zeta}}(\boldsymbol{\beta} + b\mathbf{e}_j + b \cdot z(\boldsymbol{\zeta}), \hat{s}^{t_n}_{b_n,n} + b\zeta, \hat{s}^{t_n+1}_{b,n} + b\zeta) \\
&\xrightarrow{p} \tilde{\ell}_{n_{\boldsymbol{\zeta},\zeta}}(\boldsymbol{\beta} + b\mathbf{e}_j + b \cdot z(\boldsymbol{\zeta}), s(\boldsymbol{\beta}, b) + b\zeta, s(\boldsymbol{\beta}, b) + b\zeta),
\end{aligned}$$

and by the Weak Law of Large Numbers, we have that

$$\begin{aligned}
&\tilde{\ell}_{n_{\boldsymbol{\zeta},\zeta}}(\boldsymbol{\beta} + b\mathbf{e}_j + b \cdot z(\boldsymbol{\zeta}), s(\boldsymbol{\beta}, b) + b\zeta, s(\boldsymbol{\beta}, b) + b\zeta) \\
&\xrightarrow{p} \tilde{L}(\boldsymbol{\beta} + b\mathbf{e}_j + b \cdot z(\boldsymbol{\zeta}), s(\boldsymbol{\beta}, b) + b\zeta, s(\boldsymbol{\beta}, b) + b\zeta).
\end{aligned}$$

Analogous statements hold for the remaining terms in (F.7) and (F.8). Also,

$$\frac{n_{\boldsymbol{\zeta},\zeta}}{n} \xrightarrow{p} \frac{1}{2^{d+1}}, \quad \boldsymbol{\zeta} \in \{-1,1\}^d, \zeta \in \{-1,1\}.$$

By Slutsky's Theorem, when any j and b fixed, we have

$$\mathbf{s}^n_{zy,j} \xrightarrow{p} \sum_{\substack{\boldsymbol{\zeta} \in \{-1,1\}^d \text{ s.t. } \boldsymbol{\zeta}_j = 1 \\ \zeta \in \{-1,1\}}} \frac{\tilde{L}(\boldsymbol{\beta} + b\mathbf{e}_j + b \cdot z(\boldsymbol{\zeta}), s(\boldsymbol{\beta}, b) + b\zeta, s(\boldsymbol{\beta}, b) + b\zeta)}{2^{d+1} \cdot b} \tag{F.9}$$

$$- \sum_{\substack{\boldsymbol{\zeta} \in \{-1,1\}^d \text{ s.t. } \boldsymbol{\zeta}_j = -1 \\ \zeta \in \{-1,1\}}} \frac{\tilde{L}(\boldsymbol{\beta} - b\mathbf{e}_j + b \cdot z(\boldsymbol{\zeta}), s(\boldsymbol{\beta}, b) + b\zeta, s(\boldsymbol{\beta}, b) + b\zeta)}{2^{d+1} \cdot b}. \tag{F.10}$$

Let R_b denote the expression on the right side of the above equation. If there is a sequence $\{b_n\}$ such that $b_n \to 0$, then by Lemma 51, $s(\boldsymbol{\beta}, b_n) \to s(\boldsymbol{\beta})$, where $s(\boldsymbol{\beta})$ is the unique fixed point of $q(P_{\boldsymbol{\beta},s})$. Furthermore, by the continuity of L, we have that

$$R_{b_n} \to \frac{\partial L}{\partial \boldsymbol{\beta}_j}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})).$$

Using the definition of convergence in probability, we show that there exists such a sequence $\{b_n\}$. From (F.9) and (F.10), we have that for each $\epsilon, \delta > 0$ and $b > 0$ and sufficiently small, there exists $n(\epsilon, \delta, b)$ such that for $n \geq n(\epsilon, \delta, b)$

$$P(|\mathbf{s}^n_{zy,j} - R_b| \leq \epsilon) \geq 1 - \delta.$$

So, we can fix $\delta > 0$. For $k = 1, 2, \ldots$, let $N(k) = n(\frac{1}{k}, \delta, \frac{1}{k})$. Then, we can define a sequence such that $b_n = \epsilon_n = \frac{1}{k}$ for all $N(k) \leq n \leq N(k+1)$. So, we have that $\epsilon_n \to 0$ and $b_n \to 0$. Finally, this gives that

$$\mathbf{s}^n_{zy,j} \xrightarrow{p} \frac{\partial L}{\partial \boldsymbol{\beta}_j}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})).$$

Considering all indices j,

$$\mathbf{s}^n_{zy} \xrightarrow{p} \frac{\partial L}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})).$$

It remains to establish convergence in probability for $\mathbf{S}^n_{zz}$. We have that

$$\begin{aligned}
\mathbf{Z}^n &= \frac{1}{b_n^2 n} \mathbf{Z}^T \mathbf{Z} \\
&= \frac{1}{b_n^2 n} \sum_{i=1}^n (b_n \boldsymbol{\zeta}_i)^T (b_n \boldsymbol{\zeta}_i). \\
&= \frac{1}{n} \sum_{i=1}^n \boldsymbol{\zeta}_i^T \boldsymbol{\zeta}_i.
\end{aligned}$$

We note that

$$\mathbb{E}_{\boldsymbol{\zeta}_i \sim R^d}\left[\boldsymbol{\zeta}_{i,j}\boldsymbol{\zeta}_{i,k}\right] = \begin{cases} 1 \text{ if } j = k \\ 0 \text{ if } j \neq k \end{cases}$$

because $\boldsymbol{\zeta}_i$ is a vector of independent Rademacher random variables. So, $\mathbb{E}\left[\boldsymbol{\zeta}_i^T \boldsymbol{\zeta}_i\right] = \mathbf{I}_d$. By the Weak Law of Large Numbers, we have that

$$\mathbf{S}_{zz}^n \xrightarrow{p} \mathbf{I}_d.$$

Finally, we can use Slutsky's Theorem to show that

$$\hat{\tau}_{\mathrm{ME},b_n,n}^{t_n}(\boldsymbol{\beta}) = (\mathbf{S}_{zz}^n)^{-1}\mathbf{s}_{zy}^n \xrightarrow{p} (I_d)^{-1}\frac{\partial L}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})) = \frac{\partial L}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})) = \tau_{\mathrm{ME}}(\boldsymbol{\beta}).$$

F.3 Proof of Theorem 15

Let $s^* = s(\boldsymbol{\beta})$. Let $\hat{\Gamma}_{\boldsymbol{\ell},s}^n, \hat{\Gamma}_{\boldsymbol{\pi},s}^n, \hat{\Gamma}_{\boldsymbol{\pi},\boldsymbol{\beta}}^n$ be the regression coefficients defined in Definition 7. Let $p_{\boldsymbol{\beta},s,b}^n(s)$ be the density estimate defined in Definition 7. In this proof, we rely on the results on the following convergence results for these estimators.

Corollary 53. *Let $\{t_n\}$ be a sequence such that $t_n \uparrow \infty$ as $n \to \infty$. Let $\mathbf{Z}_s, \boldsymbol{\ell}, \hat{\Gamma}_{\ell,s}^n(\boldsymbol{\beta}, s, r)$ be as defined in Definition 7. Under the conditions of Theorem 14, there exists a sequence $\{b_n\}$ such that $b_n \to 0$ so that*

$$\hat{\Gamma}_{\ell,s,\ell,r}^n(\boldsymbol{\beta}, \hat{s}_{b_n,n}^{t_n}, \hat{s}_{b_n,n}^{t_n+1}) \xrightarrow{p} \frac{\partial L}{\partial s}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})) + \frac{\partial L}{\partial r}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})). \tag{F.11}$$

Proof in Appendix G.22.

Lemma 54. *Let $\{t_n\}$ be a sequence such that $t_n \uparrow \infty$ as $n \to \infty$. Let $\mathbf{Z}_{\boldsymbol{\beta}}, \boldsymbol{\pi}, \hat{\Gamma}_{\pi,\boldsymbol{\beta}}^n(\boldsymbol{\beta}, s, r)$ be as defined in Definition 7. Under the conditions of Theorem 14, there exists a sequence $\{b_n\}$ such that $b_n \to 0$ so that*

$$\hat{\Gamma}_{\pi,\boldsymbol{\beta}}^n(\boldsymbol{\beta}, \hat{s}_{b_n,n}^{t_n}, \hat{s}_{b_n,n}^{t_n}) \xrightarrow{p} \frac{\partial \Pi}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s(\boldsymbol{\beta}); s(\boldsymbol{\beta})). \tag{F.12}$$

Proof in Appendix G.23.

Corollary 55. *Let $\{t_n\}$ be a sequence such that $t_n \uparrow \infty$ as $n \to \infty$. Let $\mathbf{Z}_s, \boldsymbol{\pi}, \hat{\Gamma}_{\pi,s}^n(\boldsymbol{\beta}, s, r)$ be as defined in Definition 7. Under the conditions of Theorem 14, there exists a sequence $\{b_n\}$ such that $b_n \to 0$ so that*

$$\hat{\Gamma}_{\pi,s}^n(\boldsymbol{\beta}, \hat{s}_{b_n,n}^{t_n}, \hat{s}_{b_n,n}^{t_n}) \xrightarrow{p} \frac{\partial \Pi}{\partial s}(\boldsymbol{\beta}, s(\boldsymbol{\beta}); s(\boldsymbol{\beta})). \tag{F.13}$$

Proof in Appendix G.24.

Lemma 56. *Fix $\boldsymbol{\beta} \in \mathcal{B}$. Let $\{h_n\}$ be a sequence such that $h_n \to 0$ and $nh_n \to \infty$. Let $p_{\boldsymbol{\beta},s,b}^n(r)$ denote a kernel density estimate of $p_{\boldsymbol{\beta},s,b}(r)$ with kernel function $k(z) = \mathbb{I}(z \in [-\frac{1}{2}, \frac{1}{2})$ and bandwidth h_n. Let $\{t_n\}$ be a sequence such that $t_n \uparrow \infty$ as $n \to \infty$. Under the conditions of Theorem 14, there exists a sequence $\{b_n\}$ such that $b_n \to 0$ so that*

$$p_{\boldsymbol{\beta},\hat{s}_n^{t_n},b_n}^n(\hat{s}_{b_n,n}^{t_n}) \xrightarrow{p} p_{\boldsymbol{\beta},s^*}(s^*), \tag{F.14}$$

where s^* is the unique fixed point of $q(P_{\boldsymbol{\beta},s})$. Proof in Appendix G.25.

Finally, we use the following lemma to show that we recover the equilibrium effect.

Lemma 57. *Let $\Pi(\boldsymbol{\beta}, s; r)$ be defined as in (5.3). Under Assumptions 1, 2, and 3, if $\sigma^2 > \frac{1}{\alpha_* \sqrt{2\pi e}}$ then*

$$\frac{\partial s}{\partial \boldsymbol{\beta}} = \frac{1}{p_{\boldsymbol{\beta},s^*}(s^*) - \frac{\partial \Pi}{\partial s}(\boldsymbol{\beta}, s^*; s^*)} \cdot \frac{\partial \Pi}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s^*; s^*), \tag{F.15}$$

where $s^* = s(\boldsymbol{\beta})$, the unique fixed point induced by the model $\boldsymbol{\beta}$. Proof in Appendix G.26.

We proceed with the main proof. The equilibrium effect estimator in (5.7) consists of two terms. We see that the convergence of the first term is immediately given by (F.11) above. It remains to show that the second term converges in probability to $\frac{\partial s}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta})$. We have that

$$\frac{\hat{\Gamma}^n_{\pi,\boldsymbol{\beta}}(\boldsymbol{\beta}, \hat{s}^{t_n}_n; \hat{s}^{t_n}_{b_n,n})}{p^n_{\boldsymbol{\beta},\hat{s}^{t_n}_{b_n,n},b_n}(\hat{s}^{t_n}_{b_n,n}) - \hat{\Gamma}^n_{\pi,s}(\boldsymbol{\beta}, \hat{s}^{t_n}_n; \hat{s}^{t_n}_{b_n,n})} \xrightarrow{p} \frac{1}{p_{\boldsymbol{\beta},s^*}(s^*) - \frac{\partial \Pi}{\partial s}(\boldsymbol{\beta}, s^*; s^*)} \cdot \frac{\partial \Pi}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s^*; s^*) \tag{F.16}$$

$$= \frac{\partial s}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}). \tag{F.17}$$

(F.16) follows by Slutsky's Theorem given (F.13), (F.12), and (F.14). (F.17) follows from Lemma (57). Combining (F.11) and (F.17) using Slutsky's Theorem, yields

$$\begin{aligned}
\hat{\tau}^{t_n}_{\mathrm{EE},n}(\boldsymbol{\beta}) &= \hat{\Gamma}_{\ell,s,\ell,r}(\boldsymbol{\beta}, \hat{s}^{t_n}_{b_n,n}, \hat{s}^{t_n+1}_{b_n,n}) \cdot \frac{\hat{\Gamma}_{\pi,\boldsymbol{\beta}}(\boldsymbol{\beta}, \hat{s}^{t_n}_{b_n,n}; \hat{s}^{t_n}_n)}{p^n_{\boldsymbol{\beta},\hat{s}^{t_n}_{b_n,n},b_n}(\hat{s}^{t_n}_{b_n,n}) - \hat{\Gamma}_{\pi,s}(\boldsymbol{\beta}, \hat{s}^{t_n}_{b_n,n}; \hat{s}^{t_n}_{b_n,n})} \\
&\xrightarrow{p} \Big(\frac{\partial L}{\partial s}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})) + \frac{\partial L}{\partial r}(\boldsymbol{\beta}, s(\boldsymbol{\beta}), s(\boldsymbol{\beta})\Big) \cdot \frac{\partial s}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}) \\
&= \tau_{\mathrm{EE}}(\boldsymbol{\beta}).
\end{aligned}$$

F.4 Proof of Corollary 16

This result follows from applying Slutsky's Theorem to the results of Theorem 14 and Theorem 15.

G Proofs of Technical Results

G.1 Proof of Lemma 31

Since $f_n \to f$ uniformly and f_n's are defined on a compact domain, then f must be continuous. By assumption, f has only one zero x^* in $[a, b]$. We can choose

$$\epsilon = \inf\{|f(x)| \mid |x - x^*| > \delta\}.$$

By uniform convergence, there exists N such that for $n \geq N$, $\sup_{x\in\mathcal{X}} |f_n(x) - f(x)| < \frac{\epsilon}{2}$. By the triangle inequality we have that

$$\begin{aligned}
|f(x)| &= |f(x) - f_n(x) + f_n(x)| \\
&\leq |f(x) - f_n(x)| + |f_n(x)| \\
|f_n(x)| &\geq |f(x)| - |f(x) - f_n(x)|.
\end{aligned}$$

For $n \geq N$ and x such that $|x - x^*| > \delta$, we realize that $|f_n(x)| > \frac{\epsilon}{2}$. So x cannot be a fixed point of f_n. Thus, if x is a fixed point of f_n, then we have that $|x_n - x^*| < \delta$. This implies that $x_n \to x^*$.

G.2 Proof of Lemma 32

By uniform convergence, we have that for any $\epsilon > 0$, for every $x \in \mathcal{X}$, there is $n \geq N$ so that

$$f(x) - \epsilon < f_n(x) < f(x) + \epsilon.$$

So, for all $x \in \mathcal{X}$, we have that $f_n(x) < f(x) + \epsilon < x^* + \epsilon$. In addition, for all $x \in \mathcal{X}$, we have that $f(x) - \epsilon < x_n$. We realize that this implies that $x_n \leq x + \epsilon$ and $x^* - \epsilon \leq x_n$. Thus, we have that $|x_n - x| < \epsilon$. So, we have that $x_n \to x^*$.

G.3 Proof of Lemma 34

From Assumption 1, we have that c_ν is twice continuously differentiable. Since G is the Normal CDF, we have that G is twice continuously differentiable. Since the composition and sum of twice continuously differentiable functions is also twice continuously differentiable, we have that $\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};\boldsymbol{\beta},s,\nu)\right]$ is twice continuously differentiable in $\mathbf{x},\boldsymbol{\beta},s$.

G.4 Proof of Lemma 35

We abbreviate

$$\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right] := \mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};\boldsymbol{\beta},s,\nu)\right].$$

From Lemma 34, we have that $\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right]$ is twice continuously differentiable in $\mathbf{x}$, so

$$\nabla_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right] = -\nabla c_\nu(\mathbf{x}-\boldsymbol{\eta};\boldsymbol{\gamma}) + G'(s-\boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta}^T.$$
$$\nabla^2_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right] = -\nabla^2 c_\nu(\mathbf{x}-\boldsymbol{\eta};\boldsymbol{\gamma}) - \boldsymbol{\beta}G''(s-\boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta}^T.$$

To show that $\nabla^2_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right]$ is negative definite at any point $(\mathbf{x},s)$, we can show that $-\nabla^2_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right]$ is positive definite at any point $(\mathbf{x},s)$. Let $\mathbf{z}\in\mathbb{S}^{d-1}$,

$$\begin{aligned}
\mathbf{z}^T(-\nabla^2_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right])\mathbf{z} &= \mathbf{z}^T[\nabla^2 c_\nu(\mathbf{x}-\boldsymbol{\eta};\boldsymbol{\gamma}) + \boldsymbol{\beta}G''(s-\boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta}^T]\mathbf{z} && \text{(G.1)}\\
&= \mathbf{z}^T\nabla^2 c_\nu(\mathbf{x}-\boldsymbol{\eta};\boldsymbol{\gamma})\mathbf{z} + G''(s-\boldsymbol{\beta}^T\mathbf{x})\cdot\mathbf{z}^T\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{z} && \text{(G.2)}\\
&\geq \inf_{\mathbf{y}}\mathbf{z}^T\nabla^2 c_\nu(\mathbf{y})\mathbf{z} + \inf_{y} G''(y)\cdot\mathbf{z}^T\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{z} && \text{(G.3)}\\
&\geq \alpha_\nu + (-\frac{1}{\sigma^2\sqrt{2\pi e}})\cdot\mathbf{z}^T\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{z} && \text{(G.4)}\\
&\geq \alpha_\nu + (-\alpha_\nu)\cdot\mathbf{z}^T\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{z} && \text{(G.5)}\\
&> 0 && \text{(G.6)}
\end{aligned}$$

We check the above inequality as follows. By Assumption 1, c_ν is α_ν-strongly convex and twice differentiable. So, we can lower bound the first term in (G.2). (G.4) holds because G is $N(0,\sigma^2)$ and the $-\frac{1}{\sigma^2\sqrt{2\pi e}} \leq G''(y) \leq \frac{1}{\sigma^2\sqrt{2\pi e}}$. The assumption that $\sigma^2 > \frac{1}{\alpha_\nu\sqrt{2\pi e}}$ yields (G.5). Finally, $\mathbf{z}^T\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{z} = (\mathbf{z}^T\boldsymbol{\beta})^2$, and the dot product of two unit vectors, $\boldsymbol{\beta}$ and $\mathbf{z}$, must be between -1 and 1, so $0 \leq \mathbf{z}^T\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{z} \leq 1$. Thus, $-\nabla^2_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right]$ is positive definite.

G.5 Proof of Lemma 36

Without loss of generality, we fix $\boldsymbol{\beta},s,\nu$. So, we abbreviate

$$\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right] := \mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};\boldsymbol{\beta},s,\nu)\right].$$

We apply Lemma 18 to $\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right]$ to establish the claim. The conditions of Lemma 18 include twice-differentiability in $\mathbf{x}$ and strict concavity in $\mathbf{x}$ of $\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right]$. Twice-differentiability follows from Lemma 34 and strict concavity follows from Lemma 35, which establishes that $\nabla^2\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right]$ is negative definite everywhere. Since $\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right]$ satisfies the conditions of Lemma 18, we have that if $\mathbf{x}\in\text{Int}(\mathcal{X})$, then $\mathbf{x}$ is the unique global maximizer of $\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right]$ on $\mathcal{X}$ if and only if $\nabla_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right]=0$. Under these conditions, if $\mathbf{x}\in\text{Int}(\mathcal{X})$, then $\mathbf{x}=\mathbf{x}^*(\boldsymbol{\beta},s,\nu)$ if and only if $\nabla_{\mathbf{x}}\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x})\right]=0$, as desired.

G.6 Proof of Lemma 37

Without loss of generality, we fix $\boldsymbol{\beta},\nu$. We use the following abbreviations

$$\mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};s)\right] := \mathbb{E}_{\boldsymbol{\epsilon}}\left[u(\mathbf{x};\boldsymbol{\beta},s,\nu)\right]$$
$$\mathbf{x}^*(s) := \mathbf{x}^*(\boldsymbol{\beta},s,\nu)$$
$$h(s) := h(s;\boldsymbol{\beta},\nu).$$

We state an additional lemma that will be used in the proof of Lemma 37.

Lemma 58. *Let* $\mathbf{x} \in \mathcal{X}$ *and* $\nu \in \mathcal{X} \times \mathcal{G}$, *and* $s \in \mathcal{S}$. *Let* $\mathbf{H} = \nabla^2 c_\nu(\mathbf{x} - \boldsymbol{\eta}; \boldsymbol{\gamma})$. *Under Assumption 1, we have that*

$$(\mathbf{H} + G''(s - \boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta}\boldsymbol{\beta}^T)^{-1} = \mathbf{H}^{-1} - \frac{G''(s - \boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{H}^{-1}}{1 + G''(s - \boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}.$$

Proof in Appendix G.27.

Now, we proceed with the main proof. We compute $\nabla_s \mathbf{x}^*$ by using the implicit expression for $\mathbf{x}^*(s)$ given by the first-order condition in Lemma 36. From Lemma 36, we note that $\mathbf{x}^*(s)$ must satisfy $\nabla_\mathbf{x} \mathbb{E}_\boldsymbol{\epsilon}[u(\mathbf{x}; s)] = 0$. We have that

$$\nabla_\mathbf{x} \mathbb{E}_\boldsymbol{\epsilon}[u(\mathbf{x}; s)] = -\nabla c_\nu(\mathbf{x} - \boldsymbol{\eta}; \boldsymbol{\gamma}) + G'(s - \boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta}^T.$$

So, the best response $\mathbf{x}(s)$ satisfies

$$-\nabla c_\nu(\mathbf{x}^*(s) - \boldsymbol{\eta}; \boldsymbol{\gamma}) + G'(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T = 0.$$

From Lemma 2, we have that the best response $\mathbf{x}^*(s)$ is continuously differentiable in s, so we can differentiate the above equation with respect to s. This yields the following equation

$$(\nabla^2 c_\nu(\mathbf{x}^*(s) - \boldsymbol{\eta}) + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}\boldsymbol{\beta}^T)\nabla_s \mathbf{x}^* = G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}.$$

Let $\mathbf{H} = \nabla^2 c_\nu(\mathbf{x}^*(s) - \boldsymbol{\eta})$. The above equation can be rewritten as

$$(\mathbf{H} + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}\boldsymbol{\beta}^T)\nabla_s \mathbf{x}^* = G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}.$$

From Lemma 58, we realize that the matrix term on the left side of the equation is invertible. We multiply both sides of the equation by the inverse of the matrix to compute $\nabla_s \mathbf{x}$.

$$\nabla_s \mathbf{x}^* = (\mathbf{H} + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}\boldsymbol{\beta}^T)^{-1} G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}.$$

We can substitute the expression for $(\mathbf{H} + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}\boldsymbol{\beta}^T)^{-1}$ from Lemma 58 into the above equation.

$$\nabla_s \mathbf{x}^* = \Big(\mathbf{H}^{-1} - \frac{G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\mathbf{H}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{H}^{-1}}{1 + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}\Big) G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}.$$

This gives us that

$$\boldsymbol{\beta}^T \nabla_s \mathbf{x}^* = \boldsymbol{\beta}^T \Big(\mathbf{H}^{-1} - \frac{G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\mathbf{H}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{H}^{-1}}{1 + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}\Big) G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta} \tag{G.7}$$

$$= G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} - \Big(\frac{(G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta})^2}{1 + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}\Big), \tag{G.8}$$

as desired.

G.7 Proof of Lemma 38

Since c_ν is α_ν-strongly convex (with $\alpha_\nu > 0$) and twice differentiable, we can see that for $\mathbf{z} \in \mathbb{R}^d$,

$$\mathbf{z}^T \nabla^2 c_\nu(\mathbf{y})\mathbf{z} \geq \alpha_\nu |\mathbf{z}|^2 > 0$$

So, $\mathbf{H} = \nabla^2 c_\nu(\mathbf{y})$ is positive definite. Since $\mathbf{H}$ is positive definite, it is invertible and its inverse $\mathbf{H}^{-1}$ is also positive definite. Assumption 1 gives

$$\mathbf{H} \succeq \alpha_\nu \mathbf{I}.$$

Since $\mathbf{H}$ and $\alpha_\nu \mathbf{I}$ are positive definite and $\mathbf{H} - \alpha_\nu \mathbf{I}$ is positive semidefinite, Lemma 21 gives us that $(\alpha_\nu \mathbf{I})^{-1} - \mathbf{H}^{-1}$ is positive semidefinite. As a result,

$$\frac{1}{\alpha_\nu}\mathbf{I} \succeq \mathbf{H}^{-1}.$$

We conclude that (C.2) holds because

$$\begin{aligned}\sup_{\mathbf{z}\in\mathcal{B}} \mathbf{z}^T\mathbf{H}^{-1}\mathbf{z} &\leq \sup_{\mathbf{z}\in\mathcal{B}} \frac{1}{\alpha_\nu}\mathbf{z}^T\mathbf{z} \\ &\leq \frac{1}{\alpha_\nu}.\end{aligned}$$

The last line follows because $\mathcal{B} = \mathbb{S}^{d-1}$.

G.8 Proof of Lemma 39

Without loss of generality, we fix $\boldsymbol{\beta}, \nu$. We use the following abbreviations

$$\begin{aligned}\mathbf{x}^*(s) &:= \mathbf{x}^*(\boldsymbol{\beta}, s, \nu) \\ h(s) &:= h(s; \boldsymbol{\beta}, \nu).\end{aligned}$$

First, we establish that h is differentiable and compute $\frac{dh}{ds}$ as follows

$$\frac{dh}{ds} = 1 - \boldsymbol{\beta}^T\nabla_s\mathbf{x}^*. \tag{G.9}$$

Second, we use the expression for $\boldsymbol{\beta}^T\nabla_s\mathbf{x}^*$ from Lemma 37 to show that under our conditions, $\boldsymbol{\beta}^T\nabla_s\mathbf{x}^* < 1$. Finally, using this fact along with (G.9), we conclude that h has a positive derivative, so it must be strictly increasing.

Now, we proceed with the main proof. First, we observe that h is differentiable in s because Lemma 2 gives that $\mathbf{x}^*(s)$ is continuously differentiable in s. Differentiating with respect to s yields (G.9).

Next, we use (C.1) to upper bound $\boldsymbol{\beta}^T\nabla_s\mathbf{x}^*$. We show that the second term on the right side of (C.1) is nonnegative. Let N and D be the numerator and denominator of the term, respectively. In particular,

$$\begin{aligned}N &:= (G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta})^2 \\ D &:= 1 + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}\end{aligned}$$

Clearly, we must have $N \geq 0$ because it consists of a squared term. We show that $D > 0$, as well.

$$\begin{aligned}D &= 1 + G''(s - \boldsymbol{\beta}^T\mathbf{x}^*(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} && \text{(G.10)} \\ &\geq 1 + \inf_y G''(y) \sup_{\mathbf{z}\in\mathcal{B}} \mathbf{z}^T\mathbf{H}^{-1}\mathbf{z} && \text{(G.11)} \\ &\geq 1 + \left(-\frac{1}{\sigma^2\sqrt{2\pi e}}\right)\cdot\frac{1}{\alpha_\nu} && \text{(G.12)} \\ &> 1 + (-\alpha_\nu)\cdot\frac{1}{\alpha_\nu} && \text{(G.13)} \\ &> 0 && \text{(G.14)}\end{aligned}$$

(G.11) follows from the observation that $\mathbf{H}^{-1}$ is positive definite (Lemma 38), so $\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} > 0$, while $G''(y)$ may take negative values. In (G.12), we apply Lemma 38 and we note that G is $N(0, \sigma^2)$, so $-\frac{1}{\sigma^2\sqrt{2\pi e}} \leq G''(y) \leq \frac{1}{\sigma^2\sqrt{2\pi e}}$. In (G.13), we use the condition that $\sigma^2 > \frac{1}{\alpha_\nu\sqrt{2\pi e}}$.

Since $D > 0$ and $N \geq 0$, we have that the second term on the right side of (C.1) is nonnegative. As a result, we can upper bound $\boldsymbol{\beta}^T\nabla_s\mathbf{x}^*$ as follows

$$\begin{aligned}
\boldsymbol{\beta}^T \nabla_s \mathbf{x}^* &= G''(s - \boldsymbol{\beta}^T \mathbf{x}^*(s)) \boldsymbol{\beta}^T \mathbf{H}^{-1} \boldsymbol{\beta} - \frac{(G''(s - \boldsymbol{\beta}^T \mathbf{x}^*(s)) \boldsymbol{\beta}^T \mathbf{H}^{-1} \boldsymbol{\beta})^2}{1 + G''(s - \boldsymbol{\beta}^T \mathbf{x}^*(s)) \boldsymbol{\beta}^T \mathbf{H}^{-1} \boldsymbol{\beta}} \\
&\leq G''(s - \boldsymbol{\beta}^T \mathbf{x}^*(s)) \boldsymbol{\beta}^T \mathbf{H}^{-1} \boldsymbol{\beta} \\
&\leq \sup_{\mathbf{z} \in \mathcal{B}} \mathbf{z}^T \mathbf{H}^{-1} \mathbf{z} \cdot \sup_y G''(y) \\
&\leq \sup_{\mathbf{z} \in \mathcal{B}} \mathbf{z}^T \mathbf{H}^{-1} \mathbf{z} \cdot \frac{1}{\sigma^2 \sqrt{2\pi e}} \\
&< \frac{1}{\alpha_\nu} \cdot \alpha_\nu \\
&< 1.
\end{aligned}$$

We can apply $\boldsymbol{\beta}^T \nabla_s \mathbf{x}^* < 1$ to (G.9) to find that

$$\frac{dh}{ds} = 1 - \boldsymbol{\beta} \nabla_s \mathbf{x}^* > 0.$$

Thus, h has a positive derivative, so it must be strictly increasing.

G.9 Proof of Lemma 40

Define $\bar{u}(\mathbf{x}; \boldsymbol{\beta}, \nu) = \lim_{s \to \infty} \mathbb{E}_{\boldsymbol{\epsilon}} [u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)]$. Note that

$$\bar{u}(\mathbf{x}; \boldsymbol{\beta}, \nu) = -c_\nu(\mathbf{x} - \boldsymbol{\eta}; \boldsymbol{\gamma}) + 1.$$

We realize that

$$\operatorname*{argmax}_{\mathbf{x} \in \mathcal{X}} \bar{u}(\mathbf{x}; \boldsymbol{\beta}, \nu) = \boldsymbol{\eta}.$$

To show that (C.3) holds, we establish that $\mathbb{E}_{\boldsymbol{\epsilon}} [u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)] \to \bar{u}(\mathbf{x}; \boldsymbol{\beta}, \nu)$ uniformly in $\mathbf{x}$ as $s \to \infty$. Then, we show that the maximizer of $\mathbb{E}_{\boldsymbol{\epsilon}} [u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)]$ must converge to the maximizer of $\bar{u}(\mathbf{x}; \boldsymbol{\beta}, \nu)$ as $s \to \infty$, which gives the desired result.

First, we verify the conditions of Lemma 30 to establish the uniform convergence of $\mathbb{E}_{\boldsymbol{\epsilon}} [u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)]$. For every s, we have that $\mathbb{E}_{\boldsymbol{\epsilon}} [u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)]$ is continuous (Lemma 34). In addition, for every s, Lemma 35 gives that the expected utility's second derivative is negative definite, which implies that it is strictly concave. Also, $\bar{u}(\mathbf{x}; \boldsymbol{\beta}, \nu)$ is continuous and concave. We note that $\mathbb{E}_{\boldsymbol{\epsilon}} [u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)] \to \bar{u}(\mathbf{x}; \boldsymbol{\beta}, \nu)$ pointwise in $\mathbf{x}$ as $s \to \infty$. Thus, Lemma 30 implies that $\mathbb{E}_{\boldsymbol{\epsilon}} [u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)] \to \bar{u}(\mathbf{x}; \boldsymbol{\beta}, \nu)$ converges uniformly in $\mathbf{x}$ as $s \to \infty$.

Second, we verify the conditions of Lemma 32 to show that

$$\lim_{s \to \infty} \mathbf{x}^*(\boldsymbol{\beta}, s, \nu) \to \boldsymbol{\eta} \tag{G.15}$$

We note that $\mathbb{E}_{\boldsymbol{\epsilon}} [u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)]$ has a unique maximizer $\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ for every s (Lemma 1), and $\bar{u}(\mathbf{x}; \boldsymbol{\beta}, \nu)$ is uniquely maximized at $\boldsymbol{\eta}$. As shown in the previous part, $\mathbb{E}_{\boldsymbol{\epsilon}} [u(\mathbf{x}; \boldsymbol{\beta}, s, \nu)]$ converges uniformly in $\mathbf{x}$ as $s \to \infty$. So, we can apply Lemma 32 to conclude that (G.15). This implies (C.3). An identical argument implies (C.4).

G.10 Proof of Lemma 41

We will use the following lemma in this proof.

Lemma 59. *Consider $\omega(s; \boldsymbol{\beta}, \nu) = \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$. Under Assumption 1, if $\sigma^2 > \frac{1}{\alpha_\nu \sqrt{2\pi e}}$ and $\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) \in$ Int($\mathcal{X}$), $\omega(s; \boldsymbol{\beta}, \nu)$ has a unique fixed point.* *Proof in Appendix G.28.*

By Lemma 59, $\omega(s;\boldsymbol{\beta},\nu)$ has a unique fixed point. Let the unique fixed point be s^*. We show that $\nabla_s\omega(s^*) = 0$ by an application of Lemma 37. Let $\mathbf{H} = \nabla^2 c(\mathbf{x}^*(s^*) - \eta;\boldsymbol{\gamma})$.

$$\begin{aligned}
\nabla_s\omega(s^*) &= \boldsymbol{\beta}^T\nabla_s\mathbf{x}^*(s^*) \\
&= G''(s^* - \boldsymbol{\beta}^T\mathbf{x}^*(s^*))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} - \Big(\frac{(G''(s^* - \boldsymbol{\beta}^T\mathbf{x}^*(s^*))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta})^2}{1 + G''(s^* - \boldsymbol{\beta}^T\mathbf{x}^*(s^*))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}\Big) \\
&= G''(0)\cdot(\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}) - \Big(\frac{(G''(0)\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta})^2}{1 + (G''(0))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}\Big) \\
&= 0.
\end{aligned}$$

The last line follows from the fact that G is $N(0,\sigma^2)$, so $G''(0) = 0$. Thus, we have that $\nabla_s\omega(s^*) = 0$.

When $s < s^*$, we have that $h(s;\boldsymbol{\beta},\nu) < h(s^*;\boldsymbol{\beta},\nu)$ because by Lemma 39, $h(s;\boldsymbol{\beta},\nu)$ is strictly increasing in s. Since $h(s^*;\boldsymbol{\beta},\nu) = 0$, this implies that $h(s;\boldsymbol{\beta},\nu) < 0$. Thus, we have that

$$\begin{aligned}
\nabla_s\omega(s) &= \boldsymbol{\beta}^T\nabla_s\mathbf{x}^*(s) \\
&= G''(h(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} - \Big(\frac{(G''(h(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta})^2}{1 + G''(h(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}\Big) \\
&= \frac{G''(h(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}{1 + G''(h(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}} \\
&> 0.
\end{aligned}$$

The last line follows because $G''(h(s)) > 0$ when $h(s) < 0$ because G is $N(0,\sigma^2)$. Thus, $\omega(s)$ is increasing when $s < s^*$.

When $s > s^*$, we have that $h(s^*;\boldsymbol{\beta},\nu) < h(s;\boldsymbol{\beta},\nu)$, again because h is strictly increasing. This implies that $h(s;\boldsymbol{\beta},\nu) > 0$. So, when $s > s^*$, $G''(h(s)) < 0$. Meanwhile,

$$\begin{aligned}
1 + G''(h(s))\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta} &\geq 1 + \inf_y G''(y)\cdot\sup_{\mathbf{z}\in\mathcal{B}}\mathbf{z}^T\mathbf{H}\mathbf{z} \\
&\geq 1 + (-\frac{1}{\sigma^2\sqrt{2\pi e}})\cdot\frac{1}{\alpha_\nu} \\
&\geq 1 + (-\alpha_\nu)\cdot\frac{1}{\alpha_\nu} \\
&> 0.
\end{aligned}$$

This means that for $s > s^*$, we have that $\nabla_s\omega(s) < 0$. So, $\omega(s)$ is decreasing when $s > s^*$.

Since $\omega(s)$ is increasing when $s < s^*$ and is decreasing when $s > s^*$, then $\omega(s)$ is maximized when $s = s^*$.

G.11 Proof of Lemma 42

Recall that $P_{\boldsymbol{\beta},s}$ denotes the distribution over the noisy scores, and the noisy score for an agent with type ν is denoted by $\boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta},s,\nu) = \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta},s,\nu) + \boldsymbol{\beta}^T\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon},\nu$ are random variables. In addition, recall that the noise is independent from the agents' type (and as a result, best response). In particular, $\boldsymbol{\beta}^T\boldsymbol{\epsilon} \sim N(0,\sigma^2)$ because $\boldsymbol{\epsilon} \sim N(0,\sigma^2 I_d)$. We have that

$$\begin{aligned}
P_{\boldsymbol{\beta},s}(r) &= P(\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta},s,\nu) + \boldsymbol{\beta}^T\boldsymbol{\epsilon} \leq r) && \nu,\boldsymbol{\epsilon} \text{ are random variables} \\
&= \int_{\mathcal{X}\times\mathcal{G}} P(\boldsymbol{\beta}^T\boldsymbol{\epsilon} \leq r - \boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta},s,\nu))dF && \epsilon \text{ is random variable} \\
&= \int_{\mathcal{X}\times\mathcal{G}} G(r - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta},s,\nu))dF.
\end{aligned}$$

Thus, $P_{\boldsymbol{\beta},s}(r)$ has the form given in (D.1). Under our conditions, the best response for each agent type exists and is unique via Lemma 1, so $P_{\boldsymbol{\beta},s}(r)$ is a well-defined function.

First, we establish that $P_{\boldsymbol{\beta},s}$ is strictly increasing because we know that G is strictly increasing, and the sum of strictly increasing functions is also strictly increasing.

Second, we establish that $P_{\boldsymbol{\beta},s}$ is continuously differentiable in r because from we have that G is continuously differentiable. $P_{\boldsymbol{\beta},s}$ is continuously differentiable in $\boldsymbol{\beta}, s$ because G is continuously differentiable and the best response mappings $\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ are continuously differentiable (Lemma 2).

The combination of the above two properties is sufficient for showing that $P_{\boldsymbol{\beta},s}$ has a continuous inverse distribution function.

G.12 Proof of Lemma 43

First, we compute $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}$ via implicit differentiation. We note that our expression for $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}$ consists of a convex combination of terms of the form $\boldsymbol{\beta}^T \nabla_s \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$. Finally, we can bound each term in the convex combination and bound $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}$.

From Lemma 42, we have that $P_{\boldsymbol{\beta},s}$ has an inverse distribution function, so $q(P_{\boldsymbol{\beta},s})$ is uniquely defined. Thus, $P_{\boldsymbol{\beta},s}(q(P_{\boldsymbol{\beta},s})) = q$ implicitly defines $q(P_{\boldsymbol{\beta},s})$. Using the expression for $P_{\boldsymbol{\beta},s}(r)$ from (D.1), we have

$$\int_{\mathcal{X}\times\mathcal{G}} G(q(P_{\boldsymbol{\beta},s}) - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)) dF = q \tag{G.16}$$

From Lemma 5, we have that $q(P_{\boldsymbol{\beta},s})$ is differentiable in s. So, we can differentiate both sides of (G.16) with respect to s.

$$\begin{aligned}
&\frac{\partial}{\partial s} \int_{\mathcal{X}\times\mathcal{G}} G(q(P_{\boldsymbol{\beta},s}) - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)) dF \\
&= \int_{\mathcal{X}\times\mathcal{G}} \frac{\partial}{\partial s}\Big(G(q(P_{\boldsymbol{\beta},s}) - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu))\Big) dF \\
&= \int_{\mathcal{X}\times\mathcal{G}} G'(q(P_{\boldsymbol{\beta},s}) - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)) \cdot \Big(\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s} - \boldsymbol{\beta}^T \nabla_s \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)\Big) dF \\
&= 0.
\end{aligned}$$

Rearranging the last two lines to solve for $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}$ yields

$$\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s} = \int_{\mathcal{X}\times\mathcal{G}} \boldsymbol{\beta}^T \nabla_s \mathbf{x}^*(\boldsymbol{\beta}, s, \nu) \cdot \frac{G'(q(P_{\boldsymbol{\beta},s}) - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu))}{\int_{\mathcal{X}\times\mathcal{G}} G'(q(P_{\boldsymbol{\beta},s}) - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)) dF} dF.$$

We have that G has a strictly increasing CDF, so $G' > 0$. As a result, we can define

$$w(\boldsymbol{\beta}, s, \nu) = \frac{G'(q(P_{\boldsymbol{\beta},s}) - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu))}{\int_{\mathcal{X}\times\mathcal{G}} G'(q(P_{\boldsymbol{\beta},s}) - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)) dF}$$

where $0 \le w(\boldsymbol{\beta}, s, \nu) \le 1$ and $\int w(\boldsymbol{\beta}, s, \nu) dF = 1$. As a result, $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}$ is a convex combination of $\boldsymbol{\beta}^T \nabla_s \mathbf{x}(\boldsymbol{\beta}, s, \nu)$ terms:

$$\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s} = \int_{\mathcal{X}\times\mathcal{G}} \boldsymbol{\beta}^T \nabla_s \mathbf{x}(\boldsymbol{\beta}, s, \nu) \cdot w(\boldsymbol{\beta}, s, \nu) dF.$$

We can upper bound each term $\boldsymbol{\beta}^T \nabla_s \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$. When $\sigma^2 > \frac{1}{\alpha_* \sqrt{2\pi e}}$ Lemma 39 gives us that for any agent type $\nu \in \mathrm{supp}(F)$, the function $h(s; \boldsymbol{\beta}, \nu) = s - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ is strictly increasing. Since $h(s; \boldsymbol{\beta}, \nu)$ is strictly increasing and differentiable, we have that

$$\frac{dh}{ds} = 1 - \boldsymbol{\beta}^T \nabla_s \mathbf{x}^*(\boldsymbol{\beta}, s, \nu) > 0.$$

As a result, each term satisfies $\boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu) < 1$. Since $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}$ is a convex combination of such terms, we also have that $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s} < 1$.

When $\sigma^2 > \frac{2}{\alpha_* \sqrt{2\pi e}}$, Lemma 3 gives us that for any agent type $\nu \in \text{supp}(F)$, the function $\boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ is a contraction in s, so $|\boldsymbol{\beta}^T \nabla_s \mathbf{x}^*| < 1$. As a result, since $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}$ is a convex combination of such terms $\boldsymbol{\beta}^T \nabla_s \mathbf{x}^*$, then $|\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}| < 1$.

G.13 Proof of Lemma 44

The cdf of P must be

$$P(r) = \int_{\mathcal{X} \times \mathcal{G}} G(r - \boldsymbol{\beta}^T \boldsymbol{\eta}) dF. \tag{G.17}$$

We have that

$$\begin{aligned}
\lim_{s \to \infty} P_{\boldsymbol{\beta},s}(r) &= \lim_{s \to \infty} \int_{\mathcal{X} \times \mathcal{G}} G(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)) dF \\
&= \int_{\mathcal{X} \times \mathcal{G}} G(r - \boldsymbol{\beta}^T \boldsymbol{\eta}) dF \\
&= P(r).
\end{aligned}$$

The first line follows from the definition of $P_{\boldsymbol{\beta},s}$ from (D.1). The second line follows from Lemma 40. The third line follows from (G.17). An identical proof can be used to show that

$$\lim_{s \to -\infty} P_{\boldsymbol{\beta},s}(r) = P(r).$$

Since $P_{\boldsymbol{\beta},s}(r) \to P(r)$ pointwise in r, $P_{\boldsymbol{\beta},s}$ are continuous and invertible and have continuous inverses (Lemma 42), and P is continuous, invertible, and has a continuous inverse, then we also have that

$$\lim_{s \to \infty} q(P_{\boldsymbol{\beta},s}) = q(P), \quad \lim_{s \to -\infty} q(P_{\boldsymbol{\beta},s}) = q(P).$$

G.14 Proof of Lemma 45

First, we verify that z^t is a valid random variable. We note that

$$\begin{aligned}
P(z^t = \epsilon_g) + \sum_{k=1}^{\infty} P(z^t = S_k) &= p_n(\epsilon_g) + \sum_{k=1}^{\infty} \frac{(1 - p_n(\epsilon_g)}{2^k} \\
&= p_n(\epsilon_g) + \frac{1}{1 - \frac{1}{2}} \cdot \frac{1 - p_n(\epsilon_g))}{2} \\
&= p_n(\epsilon_g) + (1 - p_n(\epsilon_g)) \\
&= 1.
\end{aligned}$$

Now, we show that z^t stochastically dominates $|q(P^n_{\boldsymbol{\beta},s}) - q(P_{\boldsymbol{\beta},s}))|$. So, for $b \in \mathbb{R}$, we show that

$$P(|q(P^n_{\boldsymbol{\beta},s}) - q(P_{\boldsymbol{\beta},s}))| \geq b) \leq P(z^t \geq b), \tag{G.18}$$

which is equivalent to the condition that z^t stochastically dominates $|q(P^n_{\boldsymbol{\beta},s}) - q(P_{\boldsymbol{\beta},s})|$.

From Lemma 10 we realize that for $b \in \mathbb{R}$,

$$P(|q(P^n_{\boldsymbol{\beta},s}) - q(P_{\boldsymbol{\beta},s}))| \geq b) \leq 1 - p_n(b).$$

In addition, we have that

$$P(z^t \geq b) = \begin{cases} 1 & \text{if } b \leq \epsilon_g \\ 1 - p_n(\epsilon_g) & \text{if } \epsilon_g < b \leq S_1 \\ \frac{1 - p_n(\epsilon_g)}{2^{k-1}} & \text{if } S_{k-1} < b \leq S_k, k \geq 2. \end{cases}$$

We show that (G.18) holds for the three cases 1) $b \le \epsilon_g$, 2) $\epsilon_g < b \le S_1$, and 3) $S_{k-1} < b \le S_k$ for $k \ge 2$. When $b \le \epsilon_g$, we have that

$$P(|q(P^n_{\boldsymbol{\beta},s}) - q(P_{\boldsymbol{\beta},s}))| \ge b) \le 1 - p_n(b) \le 1 = P(z^t \ge b).$$

When $\epsilon_g < b \le S_1$, we have that $p_n(b) \ge p_n(\epsilon_g)$ because $p_n(y)$ is increasing in y. So, we note that $1 - p_n(b) \le 1 - p_n(\epsilon_g)$. This yields

$$P(|q(P^n_{\boldsymbol{\beta},s}) - q(P_{\boldsymbol{\beta},s}))| \ge b) \le 1 - p_n(b) \le 1 - p_n(\epsilon_g) = P(z^t \ge b).$$

To prove the result in the case where $S_{k-1} < b \le S_k, k \ge 2$, we first show that the definition of S_k in (E.1) implies that

$$1 - p_n(S_{k-1}) = \frac{1 - p_n(\epsilon_g)}{2^{k-1}}, \tag{G.19}$$

as follows. First, we consider the definition of S_{k-1} below

$$S_{k-1} = \sqrt{-\frac{1}{2nD^2} \cdot \log\Big(\frac{1 - p_n(\epsilon_g)}{2^k}\Big)},$$

and we square both sides:

$$S^2_{k-1} = -\frac{1}{2nD^2} \cdot \log\Big(\frac{1 - p_n(\epsilon_g)}{2^k}\Big).$$

Multiplying by $-2nD^2$ and exponentiating both sides yields

$$\exp(-2nD^2 S^2_{k-1}) = \frac{1 - p_n(\epsilon_g)}{2^k}. \tag{G.20}$$

Finally, we realize that (G.19) holds because

$$\begin{aligned} 1 - p_n(S_{k-1}) &= 2\exp(-2nD^2 S^2_{k-1}) \\ &= 2 \cdot \frac{1 - p_n(\epsilon_g)}{2^k} \\ &= \frac{1 - p_n(\epsilon_g)}{2^{k-1}}, \end{aligned}$$

and the second line follows by (G.20). Using (G.19), we observe that

$$\begin{aligned} P(|q(P^n_{\boldsymbol{\beta},s}) - q(P_{\boldsymbol{\beta},s}))| \ge b) &= 1 - p_n(b) \le 1 - p_n(S_{k-1}) \\ &= \frac{1 - p_n(\epsilon_g)}{2^{k-1}} \\ &= P(z^t \ge S_k) \\ &= P(z^t \ge b). \end{aligned}$$

Thus, we conclude that (G.18) holds, yielding the desired result.

G.15 Proof of Lemma 46

We observe that

$$\begin{aligned} |\hat{s}^t_n - s^*| &= |\hat{s}^t_n - q(P_{\boldsymbol{\beta},\hat{s}^{t-1}_n}) + q(P_{\boldsymbol{\beta},\hat{s}^{t-1}_n}) - s^*| && \text{(G.21)} \\ &= |q(P^n_{\boldsymbol{\beta},\hat{s}^{t-1}_n}) - q(P_{\boldsymbol{\beta},\hat{s}^{t-1}_n}) + q(P_{\boldsymbol{\beta},\hat{s}^{t-1}_n}) - s^*| && \text{(G.22)} \\ &\le |q(P^n_{\boldsymbol{\beta},\hat{s}^{t-1}_n}) - q(P_{\boldsymbol{\beta},\hat{s}^{t-1}_n})| + |q(P_{\boldsymbol{\beta},\hat{s}^{t-1}_n}) - s^*| && \text{(G.23)} \\ &\le |q(P^n_{\boldsymbol{\beta},\hat{s}^{t-1}_n}) - q(P_{\boldsymbol{\beta},\hat{s}^{t-1}_n})| + \kappa|\hat{s}^{t-1}_n - s^*| && \text{(G.24)} \\ &\preceq_{\mathrm{SD}} z^t + \kappa|\hat{s}^{t-1}_n - s^*|. && \text{(G.25)} \end{aligned}$$

We have (G.22) because $\hat{s}_n^t$ is generated via (4.1). (G.24) holds because $q(P_{\boldsymbol{\beta},s})$ is a contraction mapping in s and s^* is the unique fixed point of $q(P_{\boldsymbol{\beta},s})$. (G.25) follows from Lemma 45.

Using recursion, we find that

$$|\hat{s}_n^t - s^*| \preceq_{\text{SD}} \sum_{i=0}^{k} z^{t-i}\kappa^i + \kappa^k |\hat{s}_n^{t-k} - s^*|.$$

G.16 Proof of Lemma 47

Let $\{z^t\}_{t\geq 1}$ be a sequence of random variables where

$$z^t = \begin{cases} \epsilon_g & \text{w.p. } p_n(\epsilon_g) \\ S_k & \text{w.p. } \frac{1-p_n(\epsilon_g)}{2^k}. \end{cases},$$

where $\epsilon_g = \frac{\epsilon(1-\kappa)}{2}$ and $p_n(\epsilon_g)$ is the bound from Lemma 10.

Since $s^* \in \mathcal{S}$, we have that

$$\begin{aligned} |\hat{s}_n^t - s^*| &\leq |q(P^n_{\boldsymbol{\beta},\hat{s}_n^{t-1}}) - s^*| && \text{(G.26)} \\ &\leq |q(P^n_{\boldsymbol{\beta},\hat{s}_{A,n}^{t-1}}) - q(P_{\boldsymbol{\beta},\hat{s}_n^{t-1}})| + |q(P_{\boldsymbol{\beta},\hat{s}_n^{t-1}}) - s^*| && \text{(G.27)} \\ &\preceq_{\text{SD}} z^t + \kappa|\hat{s}_n^{t-1} - s^*|. && \text{(G.28)} \end{aligned}$$

where (G.26) holds because the truncation is contraction map to the equilibrium threshold s^* and (G.28) holds because $q(P_{\boldsymbol{\beta},s})$ is a contraction in s and s^* is equilibrium threshold. So, as in Lemma 46, we can show that

$$|\hat{s}_n^t - s^*| \preceq_{\text{SD}} \sum_{i=0}^{k} z^{t-i}\kappa^i + \kappa^k |\hat{s}_n^{t-k} - s^*|.$$

Let $S = |\hat{s}_0^n - s^*|$. Thus, an identical argument as the proof of Theorem 11 can be used to show that if

$$t \geq \left\lceil \frac{\log(\frac{\epsilon}{2S})}{\log \kappa} \right\rceil, \quad n \geq \frac{1}{2M_{\epsilon_g}^2}\log(\frac{4t}{\delta}),$$

we have that

$$P(|\hat{s}_n^t - s^*| \geq \epsilon) \leq \delta.$$

Using an identical argument as the proof of Corollary 12, it can be shown that for any sequence $\{t_n\}$ such that $t_n \uparrow \infty$ as $n \to \infty$ and $t_n \prec \exp(n)$, $\hat{s}_n^{t_n} \xrightarrow{p} s^*$, as desired.

G.17 Proof of Lemma 48

Let H be the joint distribution of ν and $\boldsymbol{\epsilon}$. We can use the following abbreviations

$$\begin{aligned} \mathbf{w} &:= (\nu, \boldsymbol{\epsilon}) \\ \boldsymbol{\theta} &:= (\boldsymbol{\beta}, s, r) \\ \tilde{\pi}(\mathbf{w}, \boldsymbol{\theta}) &:= \tilde{\pi}(\nu, \boldsymbol{\epsilon}, \boldsymbol{\beta}, s, r) \\ \tilde{\ell}(\mathbf{w}, \boldsymbol{\theta}) &:= \tilde{\ell}(\nu, \boldsymbol{\epsilon}, \boldsymbol{\beta}, s, r) \\ \tilde{k}(\mathbf{w}, \boldsymbol{\theta}) &:= \tilde{k}(\nu, \boldsymbol{\epsilon}, \boldsymbol{\beta}, s, r). \end{aligned}$$

The conditions on $a(\mathbf{w};\boldsymbol{\theta})$ in Lemma 26 include that

1. $\boldsymbol{\theta} \in \Theta$, where Θ is compact.
2. $a(\mathbf{w};\boldsymbol{\theta})$ is continuous with probability 1 for each $\boldsymbol{\theta} \in \Theta$.

3. $|a(\mathbf{w};\boldsymbol{\theta})| \leq d(\mathbf{w})$ and $\mathbb{E}_{\mathbf{w}\sim H}\left[d(\mathbf{w})\right]$.

First, for all of $\tilde{\ell}, \tilde{\pi}, \tilde{k}$, we have that the parameter space $\mathcal{B} \times \mathcal{S} \times \mathcal{S}$ is compact.

Second, we can verify that for fixed parameters, $\tilde{\ell}, \tilde{\pi}$, and $\tilde{k}$ are continuous with probability 1. By Assumption 5, we have that each for $\pi \in \{0,1\}$, $\ell(\pi,\nu)$ is continuous, so the only discontinuity of $\tilde{\ell}$ occurs at the threshold when π flips from 0 to 1. Similarly, $\tilde{\pi}$ is an indicator function, so its only discontinuity occurs at threshold. Thus, $\tilde{\ell}(\cdot,\boldsymbol{\theta})$ and $\tilde{\pi}(\cdot,\boldsymbol{\theta})$ are discontinuous on the set

$$A = \{(\nu,\epsilon) : \boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta}, s, \nu) = r\}$$

but are otherwise continuous. The probability that $(\nu,\boldsymbol{\epsilon}) \sim H$ satisfies the condition of set A is equal to the probability that a score $z \sim P_{\boldsymbol{\beta},s}$ takes value exactly r. We note that a singleton subset $\{r\}$ will have measure 0, so the probability that a score takes value r is 0. Thus, A must also have measure 0. Since $\tilde{\ell}$ and $\tilde{\pi}$ are continuous except on a set of measure 0, $\tilde{\ell}$ and $\tilde{\pi}$ are continuous with probability 1. We realize that $\tilde{k}$ is continuous except for on the following set

$$A' = \{(\nu,\boldsymbol{\epsilon}) : \boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta}, s, \nu) = r + \frac{h}{2}\} \cup \{(\nu,\boldsymbol{\epsilon}) : \boldsymbol{\beta}^T\mathbf{x}(\boldsymbol{\beta}, s, \nu) = r - \frac{h}{2}\}.$$

The probability that $(\nu,\epsilon) \sim H$ satisfies the condition of set A' is equal to the probability that a score $z \sim P_{\boldsymbol{\beta},s}$ takes value exactly $r + \frac{h}{2}$ or value $r - \frac{h}{2}$. Since the sets $\{r + \frac{h}{2}\}$ and $\{r - \frac{h}{2}\}$ have measure zero and a countable union of measure zero sets has measure zero, then A' has measure zero. Thus, $\tilde{k}$ is continuous with probability 1.

Third, we note that $\tilde{\ell}, \tilde{\pi}$, and $\tilde{k}$ are dominated. For $\tilde{\ell}$, Assumption 5 gives us that $\ell(\pi,\nu)$ is bounded, so any constant function $d(\nu) = c$ for $c \geq \sup_{\nu\in\mathcal{X}\times\mathcal{G},\pi\in\{0,1\}} \ell(\nu,\pi)$ dominates $\tilde{\ell}$. Since $\tilde{\pi}$ and $\tilde{k}$ are indicators, they only takes values $\{0, 1\}$, so any constant function $d(\nu) = c$ where $c > 1$ satisfies the required condition. Thus, $\tilde{\ell}$, $\tilde{k}$, and $\tilde{\pi}$ satisfy the conditions of Lemma 26.

G.18 Proof of Lemma 49

In this proof, we first verify that the agent with type $\nu'_{b,\boldsymbol{\zeta},\zeta}$ has $\nu'_{b,\boldsymbol{\zeta},\zeta} \in \mathcal{X} \times \mathcal{G}$ and has a strongly-convex cost function $c_{\nu'}$. Second, we verify the value of the best response for this agent and show that it lies in $\mathrm{Int}(\mathcal{X})$. Lastly, we show that this agent's raw score (without noise) matches that of the agent with type ν under the perturbed model.

By the following lemma, we realize that with $\boldsymbol{\eta}'_{b,\boldsymbol{\zeta},\zeta}$ as defined in (F.2), $\nu'_{b,\boldsymbol{\zeta},\zeta} = (\boldsymbol{\eta}'_{b,\boldsymbol{\zeta},\zeta}, \boldsymbol{\gamma}) \in \mathcal{X} \times \mathcal{G}$ is a valid type as long as the perturbation magnitude b is sufficiently small.

Lemma 60. *Let $\boldsymbol{\zeta} \in \{-1,1\}^d$ and $\zeta \in \{-1,1\}$ denote perturbations. Let b be the magnitude of the perturbations. Let $\boldsymbol{\beta} \in \mathcal{B}$. If $\mathbf{y} \in$ Interior$(\mathcal{X})$ and $\mathbf{x} \in \mathcal{X}$, then for any b sufficiently small and*

$$\mathbf{y}' = \mathbf{y} + \boldsymbol{\beta} \cdot b \cdot (\boldsymbol{\zeta}^T\mathbf{x} - \boldsymbol{\zeta}),$$

we have that $\mathbf{y}' \in Int(\mathcal{X})$. Proof in Appendix G.29.

With the above lemma, we define b_1 so that $\boldsymbol{\eta}'_{b,\boldsymbol{\zeta},\zeta} \in \mathcal{X}$ for $b < b_1$.

We verify that $c_{\nu'}$ satisfies Assumption 1. We note that $c_{\nu'}$ is twice continuously differentiable because it is the sum of twice continuously differentiable functions. Second, we show that $c_{\nu'}$ is strongly convex. Since c_ν satisfies Assumption 1, then c_ν is α_ν−strongly convex for $\alpha_\nu > 0$ and twice continuously differentiable. In addition, $G'(s-r)\boldsymbol{\beta}^T\mathbf{y}$ is differentiable and convex in $\mathbf{y}$. By the strong convexity of c_ν and the convexity of $G'(s-r)\boldsymbol{\beta}^T\mathbf{y}$, we have that

$$\begin{aligned}
c_{\nu'}(\mathbf{y}) =& c_\nu(\mathbf{y}) + G'(s-r)\boldsymbol{\beta}^T\mathbf{y} \\
\geq& (c_\nu(\mathbf{y}_0) + \nabla c_\nu(\mathbf{y}_0)^T(\mathbf{y} - y_0) + \frac{\alpha_\nu}{2}||\mathbf{y} - \mathbf{y}_0||_2^2) \\
&+ (G'(s-r)\boldsymbol{\beta}^T(\mathbf{y}_0) + G'(s-r)\boldsymbol{\beta}^T(\mathbf{y} - y_0)) \\
=& (c_\nu(\mathbf{y}_0) + G'(s-r)\boldsymbol{\beta}^T(\mathbf{y}_0)) + (\nabla c_\nu(\mathbf{y}_0)^T + G'(s-r)\boldsymbol{\beta}^T)(\mathbf{y} - \mathbf{y}_0)) + \frac{\alpha_\nu}{2}||\mathbf{y} - \mathbf{y}_0||_2^2 \\
=& c_{\nu'}(\mathbf{y}_0) + \nabla c_{\nu'}(\mathbf{y}_0)^T(\mathbf{y} - \mathbf{y}_0) + \frac{\alpha_\nu}{2}||\mathbf{y} - \mathbf{y}_0||_2^2.
\end{aligned}$$

So, $c_{\nu'}$ is $\alpha_{\nu'}$-strongly convex where $\alpha_{\nu'} = \alpha_\nu$, satisfying Assumption 1.

Let

$$\mathbf{x}_2 = \mathbf{x}_1 + b \cdot \boldsymbol{\beta}(\boldsymbol{\zeta}^T \mathbf{x}_1 - \zeta).$$

Note that by Lemma 60, for sufficiently small b, we have that $\mathbf{x}_2 \in \text{Int}(\mathcal{X})$. Suppose that $\mathbf{x}_2 \in \text{Int}(\mathcal{X})$ for $b < b_2$. We will show two useful facts about $\mathbf{x}_2$ that will enable us to show that the best response of the agent with type $\nu'_{b,\boldsymbol{\zeta},\zeta}$ to the model $\boldsymbol{\beta}$ and threshold s is given by $\mathbf{x}_2$. For the first fact, we see that $\mathbf{x}_2 - \boldsymbol{\eta}'_{b,\boldsymbol{\zeta},\zeta} = \mathbf{x}_1 - \boldsymbol{\eta}$.

$$\begin{aligned}
\mathbf{x}_2 - \boldsymbol{\eta}'_{b,\boldsymbol{\zeta},\zeta} &= \boldsymbol{\beta}(b\boldsymbol{\zeta}^T \mathbf{x}_1 - b\zeta) + \mathbf{x}_1 - \boldsymbol{\eta}'_{b,\boldsymbol{\zeta},\zeta} \\
&= \boldsymbol{\beta}(b\boldsymbol{\zeta}^T \mathbf{x}_1 - b\zeta) + \mathbf{x}_1 - \boldsymbol{\eta} - \boldsymbol{\beta} \cdot b \cdot (\boldsymbol{\zeta}^T \mathbf{x}_1 - \zeta) \\
&= \mathbf{x}_1 - \boldsymbol{\eta}.
\end{aligned}$$

For the second fact, we have that $\boldsymbol{\beta}^T \mathbf{x}_2 = r$.

$$\begin{aligned}
\boldsymbol{\beta}^T \mathbf{x}_2 &= \boldsymbol{\beta}^T \left(\boldsymbol{\beta}(b\boldsymbol{\zeta}^T \mathbf{x}_1 - b\zeta) + \mathbf{x}_1 \right) &\text{(G.29)} \\
&= (\boldsymbol{\beta}^T \boldsymbol{\beta}) \cdot (b\boldsymbol{\zeta}^T \mathbf{x}_1 - b\zeta) + \boldsymbol{\beta}^T \mathbf{x}_1 &\text{(G.30)} \\
&= (\boldsymbol{\beta} + b\boldsymbol{\zeta})^T \mathbf{x}_1 - b\zeta &\text{(G.31)} \\
&= r. &\text{(G.32)}
\end{aligned}$$

Now, we show that $\mathbf{x}_2 = \mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta})$. Since Assumption 1 holds, by Lemma 36 it is sufficient to check $\nabla_{\mathbf{x}} \mathbb{E}_{\boldsymbol{\epsilon}} \left[u(\mathbf{x}_2; \boldsymbol{\beta}, s, \nu'_b) \right] = 0$ to verify that $\mathbf{x}_2$ is the best response:

$$\begin{aligned}
\nabla_{\mathbf{x}} \mathbb{E}_{\boldsymbol{\epsilon}} \left[u(\mathbf{x}_2; \boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) \right] &= -\nabla c_{\nu'}(\mathbf{x}_2 - \boldsymbol{\eta}'_{b,\boldsymbol{\zeta},\zeta}) + G(s - \boldsymbol{\beta}^T \mathbf{x}_2) \boldsymbol{\beta}^T &\text{(G.33)} \\
&= -\nabla c_{\nu'}(\mathbf{x}_1 - \boldsymbol{\eta}) + G(s - r) \boldsymbol{\beta}^T &\text{(G.34)} \\
&= -\nabla c_{\nu}(\mathbf{x}_1 - \boldsymbol{\eta}) + G(s - r) b\boldsymbol{\zeta}^T + G(s - r) \boldsymbol{\beta}^T. &\text{(G.35)}
\end{aligned}$$

To further simplify the above equation, we have that $\mathbf{x}_1 = \mathbf{x}^*(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu)$ and $\mathbf{x}_1 \in \text{Int}(\mathcal{X})$. By Lemma 36, this implies that $\nabla_{\mathbf{x}} \mathbb{E}_{\boldsymbol{\epsilon}} \left[u(\mathbf{x}_1; \boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) \right] = 0$. This gives that

$$\begin{aligned}
\nabla_{\mathbf{x}} \mathbb{E}_{\boldsymbol{\epsilon}} \left[u(\mathbf{x}_1; \boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) \right] &= -\nabla c_\nu(\mathbf{x}_1 - \boldsymbol{\eta}) + G(s + b\zeta - (\boldsymbol{\beta} + b\zeta)^T \mathbf{x}_1)(\boldsymbol{\beta} + b\boldsymbol{\zeta})^T \\
&= -\nabla c_\nu(\mathbf{x}_1 - \boldsymbol{\eta}) + G(s - r)(\boldsymbol{\beta} + b\boldsymbol{\zeta})^T \\
&= 0.
\end{aligned}$$

So, we have that

$$\nabla c_\nu(\mathbf{x}_1 - \boldsymbol{\eta}) = G(s - r)(\boldsymbol{\beta} + b\boldsymbol{\zeta})^T.$$

Substituting this result into (G.35) yields

$$\begin{aligned}
\nabla_{\mathbf{x}} \mathbb{E}_{\boldsymbol{\epsilon}} \left[u(\mathbf{x}_2; \boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) \right] &= -\nabla c_\nu(\mathbf{x}_1 - \boldsymbol{\eta}) + G(s - r) b\boldsymbol{\zeta}^T + G(s - r) \boldsymbol{\beta}^T \\
&= -G(s - r)(\boldsymbol{\beta} + b\zeta)^T + G(s - r) b\boldsymbol{\zeta}^T + G(s - r) \boldsymbol{\beta}^T \\
&= 0.
\end{aligned}$$

We note that if $b < \min(b_1, b_2)$, then we have that $\boldsymbol{\eta}'_{b,\boldsymbol{\zeta},\zeta}, \mathbf{x}_2 \in \text{Int}(\mathcal{X})$. Under such conditions, we conclude that $\mathbf{x}_2 = \mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta})$. The score obtained by the agent with type $\nu'_{b,\boldsymbol{\zeta},\zeta}$ and cost function $c_{\nu'}$ under the model $\boldsymbol{\beta}$ and threshold s is $\boldsymbol{\beta}^T \mathbf{x}_2$. As we showed earlier in (G.32), this quantity is equal to r. Thus, for sufficiently small perturbations, the agent with type ν under perturbations obtains the same raw score as the agent with type $\nu'_{b,\boldsymbol{\zeta},\zeta}$ in the unperturbed setting.

G.19 Proof of Lemma 50

Note that by Assumption 2, F has finitely many types with positive probability. Let f be the probability mass function of F, so $f(\nu)$ gives the probability of that an agent has type ν.

We construct the probability mass function $\tilde{f}_b$ of the distribution $\tilde{F}_b$. Let $\nu \sim F$ and c_ν be its cost function. Under Assumptions 3 and 4, we can compute $T(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta)$, as defined in Lemma 49, for each perturbation $\boldsymbol{\zeta} \sim \{-1, 1\}$ and $\zeta \sim \{-1, 1\}$ and type $\nu \sim F$. Let

$$T(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta) = (\nu'_{b,\boldsymbol{\zeta},\zeta}, c_{\nu'}).$$

We assign $\tilde{f}_b(\nu'_{b,\boldsymbol{\zeta},\zeta}) = \frac{1}{2^{d+1}} f(\nu)$. Since there are finitely many types that occur with positive probability in F and finitely many perturbations (2^{d+1} possible perturbations), there exists $b > 0$ such that this transformation is possible simultaneously for all types in the support of F and all perturbations. Note that this is a valid probability mass function because $\sum_{\nu \sim \text{supp}(F)} f(\nu) = 1$, so

$$\sum_{\nu'_{b,\boldsymbol{\zeta},\zeta} \in \text{supp}(\tilde{F}_b)} \tilde{f}_b(\nu'_{b,\boldsymbol{\zeta},\zeta}) = \sum_{\nu \sim \text{supp}(F)} \sum_{\substack{\boldsymbol{\zeta} \in \{-1,1\}^d \\ \zeta \in \{-1,1\}}} \frac{1}{2^{d+1}} f(\nu) = 1.$$

In addition, note that the transformation yields

$$\boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) = (\boldsymbol{\beta} + b\boldsymbol{\zeta})^T \mathbf{x}^*(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) - b\zeta.$$

The transformation given in Lemma 49 also provides other desirable properties. For instance, $\nu'_{b,\boldsymbol{\zeta},\zeta} \in \mathcal{X} \times \mathcal{G}$, which means that the support of $\tilde{F}_b$ is contained in $\mathcal{X} \times \mathcal{G}$. The cost function of the $c_{\nu'}$ satisfies Assumption 1 with $\alpha_{\nu'} = \alpha_\nu$, which means that $\alpha_*(\tilde{F}_b) = \alpha_*(F)$. Lastly, for $\nu'_{b,\boldsymbol{\zeta},\zeta} \sim \tilde{F}_b$, the best responses of the agents lie in $\text{Int}(\mathcal{X})$.

Additionally, we have that

$$\begin{aligned}
P_{\boldsymbol{\beta},s,b}(r) &= \frac{1}{2^{d+1}} \sum_{\substack{\boldsymbol{\zeta} \in \{-1,1\}^d \\ \zeta \in \{-1,1\}}} \int G\Big(r - (\boldsymbol{\beta} + b\boldsymbol{\zeta})^T \mathbf{x}^*(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) - b\zeta\Big) dF \\
&= \frac{1}{2^{d+1}} \sum_{\substack{\boldsymbol{\zeta} \in \{-1,1\}^d \\ \zeta \in \{-1,1\}}} \sum_{\nu \in \text{supp}(F)} G\Big(r - (\boldsymbol{\beta} + b\boldsymbol{\zeta})^T \mathbf{x}^*(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) - b\zeta\Big) f(\nu) \\
&= \sum_{\substack{\boldsymbol{\zeta} \in \{-1,1\}^d \\ \zeta \in \{-1,1\}}} \sum_{\nu \in \text{supp}(F)} G\Big(r - (\boldsymbol{\beta} + b\boldsymbol{\zeta})^T \mathbf{x}^*(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) - b\zeta\Big) \frac{f(\nu)}{2^{d+1}} \\
&= \sum_{\substack{\boldsymbol{\zeta} \in \{-1,1\}^d \\ \zeta \in \{-1,1\}}} \sum_{\nu \in \text{supp}(F)} G\Big(r - (\boldsymbol{\beta} + b\boldsymbol{\zeta})^T \mathbf{x}^*(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) - b\zeta\Big) \tilde{f}_b(\nu'_{b,\boldsymbol{\zeta},\zeta}) \\
&= \sum_{\nu'_{b,\boldsymbol{\zeta},\zeta} \in \text{supp}(\tilde{F}_b)} G(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta})) \tilde{f}_b(\nu'_{b,\boldsymbol{\zeta},\zeta}) \\
&= \int G(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)) d\tilde{F}_b.
\end{aligned}$$

The final line matches the form of the score distribution's CDF given in Lemma 42 assuming that the agent types are distributed according to $\tilde{F}_b$.

G.20 Proof of Lemma 51

We define the sequence of functions $\{h_b(s)\}$ where $h_b : \mathcal{S} \to \mathcal{S}$. Let $h_b(s) := s - q(P_{\boldsymbol{\beta},s,b})$ and $h(s) := s - q(P_{\boldsymbol{\beta},s})$.

We aim to apply Lemma 31 to this sequence of functions. We realize that the requirements on $h(s)$ are given by our results from Section 3. Theorem 4 and Theorem 6 give us that $h(s)$ has a unique root, which is the unique fixed point of $q(P_{\boldsymbol{\beta},s})$ called $s(\boldsymbol{\beta})$. Also, we note that $h_b(s)$ and $h(s)$ are defined on the compact set $\mathcal{S}$. It remains to check that

1. Each $h_b(s)$ is continuous,
2. Each $h_b(s)$ has a unique root, which is the fixed point of $q(P_{\boldsymbol{\beta},s,b})$ called $s(\boldsymbol{\beta}, b)$,
3. As $b \to 0$, $h_b(s) \to h(s)$ is uniformly.

To verify the first two properties from the above list, we apply the transformation provided in Lemma 50 to $P_{\boldsymbol{\beta},s,b}$. This transformation enables us to apply the results from Section 3 directly to expressions involving $P_{\boldsymbol{\beta},s,b}$.

Since the transformation maintains all of our assumptions and

$$\sigma^2 > \frac{1}{\alpha_*(F)\sqrt{2\pi e}} = \frac{1}{\alpha_*(\tilde{F}_b)\sqrt{2\pi e}},$$

we can apply Lemma 5 to see that $q(P_{\boldsymbol{\beta},s,b})$ is continuous in s. This gives the continuity of $h_b(s)$. In addition, we can apply Theorem 4 and Theorem 6 to find that $q(P_{\boldsymbol{\beta},s,b})$ has a unique fixed point in $\mathcal{S}$. We can call the fixed point $s(\boldsymbol{\beta}, b)$, and $s(\boldsymbol{\beta}, b)$ is also the unique root of $h_b(s)$.

Finally, we must check the third point, which is uniform convergence of $h_b(s)$ to $h(s)$. We aim to apply Lemma 29. First, we note that the continuity of $h(s)$ is given by Lemma 5. Second, we check that each $h_b(s)$ is monotonically increasing. Under the transformation from Lemma 50, we can apply Lemma 43 to observe that under our conditions, $\frac{\partial q(P_{\boldsymbol{\beta},s,b})}{\partial s} < 1$, so $h_b(s)$ is strictly increasing. Third, we show that $h_b(s)$ converges pointwise to $h(s)$ as follows.

To show $h_b(s) \to h(s)$ pointwise, we show $q(P_{\boldsymbol{\beta},s,b}) \to q(P_{\boldsymbol{\beta},s})$ pointwise. Note that by Lemma 42, $P_{\boldsymbol{\beta},s}$ is strictly increasing, so we can let a lower bound on its density be d for $s \in \mathcal{S}$, i.e.

$$d = \inf_{r \in \mathcal{S}} p_{\boldsymbol{\beta},s}(r).$$

Then, we realize that

$$|q(P_{\boldsymbol{\beta},s,b}) - q(P_{\boldsymbol{\beta},s})| \leq \frac{1}{d} \cdot \sup_{r \in \mathcal{S}} |P_{\boldsymbol{\beta},s,b}(r) - P_{\boldsymbol{\beta},s}(r)|.$$

The following lemma gives us the required uniform convergence in r.

Lemma 61. *Under Assumptions 1, 2, 3, and 4, if $\sigma^2 > \frac{1}{\alpha_*(F)\sqrt{2\pi e}}$ then $P_{\boldsymbol{\beta},s,b}(r) \to P_{\boldsymbol{\beta},s}(r)$ uniformly in r as $b \to 0$.Proof in Appendix G.30.*

So, we have that $q(P_{\boldsymbol{\beta},s,b}) \to q(P_{\boldsymbol{\beta},s})$ pointwise in s. This implies $h_b(s) \to h(s)$ pointwise. Thus, we have that $h_b(s)$ and $h(s)$ satisfy the conditions of Lemma 29, which implies that $h_b(s) \to h(s)$ uniformly.

Thus, the conditions of Lemma 31 are satisfied, so we have that $s(\boldsymbol{\beta}, b) \to s(\boldsymbol{\beta})$ as $b \to 0$.

G.21 Proof of Lemma 52

For sufficiently small b, we can apply Lemma 50 to show that $P_{\boldsymbol{\beta},s,b}$ is equal to the score distribution generated when agents with type $\nu'_{b,\boldsymbol{\zeta},\zeta} \sim \tilde{F}_b$ and cost functions $c_{\nu'}$ best respond to a model $\boldsymbol{\beta}$ and threshold s. The conditions assumed when types are distributed $\nu \sim F$ and cost functions are c_ν also hold when types are distributed $\nu'_{b,\boldsymbol{\zeta},\zeta} \sim \tilde{F}_b$ and cost functions are $c_{\nu'}$. In particular, $\alpha_*(\tilde{F}_b) = \alpha_*(F)$, so we have that $\sigma^2 > \frac{2}{\alpha_*(\tilde{F}_b)\cdot\sqrt{2\pi e}}$.

As a result, the results from Section 3 and Section 4 can be used to study $q(P_{\boldsymbol{\beta},s,b})$ and the stochastic fixed point iteration process given by (5.4). First, we have that $\sigma^2 > \frac{2}{\alpha_*(\tilde{F}_b)\cdot\sqrt{2\pi e}}$, so we have that $q(P_{\boldsymbol{\beta},s,b})$ is a contraction in s by Corollary 8. Furthermore, the conditions of Lemma 47 are satisfied by the assumed conditions, the results of Lemma 50, and the fact that $q(P_{\boldsymbol{\beta},s,b})$ is a contraction. So, we have that

$$\hat{s}_{b,n}^{t_n} \xrightarrow{p} s(\boldsymbol{\beta}, b),$$

where $s(\boldsymbol{\beta}, b)$ is the unique fixed point of $q(P_{\boldsymbol{\beta},s,b})$. In addition, since $\{t_n\}$ is a sequence such that $t_n \uparrow \infty$ as $n \to \infty$ and $t_n \prec \exp(n)$, we certainly have that $\{t_n + 1\}$ is a sequence such that $t_n + 1 \uparrow \infty$ as $n \to \infty$ and $t_n + 1 \prec \exp(n)$, so again by Lemma 47, we have that

$$\hat{s}_{b,n}^{t_n+1} \xrightarrow{p} s(\boldsymbol{\beta}, b).$$

G.22 Proof of Corollary 53

The proof of this result is analogous to Theorem 14.

G.23 Proof of Lemma 54

To simplify notation, we use the following abbreviations. Let $s(\boldsymbol{\beta})$ be the unique fixed point of $q(P_{\boldsymbol{\beta},s})$.

$$\begin{aligned}\tilde{\pi}_i(\boldsymbol{\beta}, s, r) &:= \pi(\mathbf{x}^*(\boldsymbol{\beta}, s, \nu_i) + \boldsymbol{\epsilon}_i; \boldsymbol{\beta}, r)\\ \tilde{\pi}_n(\boldsymbol{\beta}, s, r) &:= \frac{1}{n}\sum_{i=1}^{n}\tilde{\pi}_i(\boldsymbol{\beta}, s, r).\\ \tilde{\Pi}(\boldsymbol{\beta}, s, r) &:= \mathbb{E}_{\nu,\boldsymbol{\epsilon}}\left[\tilde{\pi}_i(\boldsymbol{\beta}, s, r)\right]\end{aligned}$$

We note that $\tilde{\pi}_i(\boldsymbol{\beta}, s, r) = \pi_i(\boldsymbol{\beta} + b_n\boldsymbol{\zeta}_i, s + b_n\zeta_i, r)$.

The regression coefficient obtained by running OLS of $\boldsymbol{\pi}(\boldsymbol{\beta}, \hat{s}^{t_n}_{b_n,n}, \hat{s}^{t_n}_{b_n,n})$ on $\mathbf{Z}$ is denoted by $\Gamma^n_{\boldsymbol{\pi},\boldsymbol{\beta}}(\boldsymbol{\beta}, \hat{s}^{t_n}_{b_n,n}, \hat{s}^{t_n}_{b_n,n})$. The regression coefficient must have the following form.

$$\Gamma^n_{\boldsymbol{\pi},\boldsymbol{\beta}}(\boldsymbol{\beta}, \hat{s}^{t_n}_{b_n,n}, \hat{s}^{t_n}_{b_n,n}) = (\mathbf{S}^n_{zz})^{-1}\mathbf{s}^n_{zy}, \text{ where } \mathbf{S}^n_{zz} := \frac{1}{b_n^2 n}\mathbf{Z}^T_{\boldsymbol{\beta}}\mathbf{Z}_{\boldsymbol{\beta}}, \quad \mathbf{s}^n_{zy} := \frac{1}{b_n^2 n}\mathbf{Z}^T_{\boldsymbol{\beta}}\boldsymbol{\pi}. \tag{G.36}$$

In this proof, we establish convergence in probability of the two terms above separately. The bulk of the proof is the first step, which entails showing that

$$\mathbf{s}^n_{zy} \xrightarrow{p} \frac{\partial \Pi}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s(\boldsymbol{\beta}); s(\boldsymbol{\beta})).$$

Due to $\tilde{\boldsymbol{\pi}}'$s dependence on the stochastic process $\{\hat{s}^{t_n}_{b_n,n}\}$, the main workhorse of this result is Lemma 28. To apply this lemma, we must establish stochastic equicontinuity for the collection of stochastic processes $\{\tilde{\pi}_n(\boldsymbol{\beta}, s, r)\}$. Second, through a straightforward application of the Weak Law of Large Numbers, we show that

$$\mathbf{S}^n_{zz} \xrightarrow{p} \mathbf{I}_d.$$

Finally, we use Slutsky's Theorem to establish the convergence the regression coefficient.

We proceed with the first step of establishing convergence of $\mathbf{s}_{zy}$. We have that

$$\begin{aligned}\mathbf{s}^n_{zy} &= \frac{1}{b_n^2 n}\mathbf{Z}^T\boldsymbol{\pi}(\boldsymbol{\beta}, \hat{s}^{t_n}_{b_n,n}, \hat{s}^{t_n}_{b_n,n})\\ &= \frac{1}{b_n^2 n}\sum_{i=1}^{n} b_n\boldsymbol{\zeta}_i\boldsymbol{\pi}_i(\boldsymbol{\beta}, \hat{s}^{t_n}_{b_n,n}, \hat{s}^{t_n}_{b_n,n})\\ &= \frac{1}{b_n}\cdot\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\zeta}_i\boldsymbol{\pi}_i(\boldsymbol{\beta}, \hat{s}^{t_n}_{b_n,n}, \hat{s}^{t_n}_{b_n,n})\end{aligned}$$

We fix j and $b_n = b$ where $b > 0$ and is small enough to satisfy the hypothesis of Lemma 52. For each $\boldsymbol{\zeta} \in \{-1, 1\}^d$ and $\zeta \in \{-1, 1\}$, let

$$n_{\boldsymbol{\zeta},\zeta} = \sum_{i=1}^{n}\mathbb{I}(\boldsymbol{\zeta}_i = \boldsymbol{\zeta}, \zeta_i = \zeta).$$

Let $z(\boldsymbol{\zeta})$ map a perturbation $\boldsymbol{\zeta} \in \{-1, 1\}^d$ to the identical vector $\boldsymbol{\zeta}$, except with j-th entry set to 0. So, if the j-th entry of $\boldsymbol{\zeta}$ is 1, then $\boldsymbol{\zeta} = \mathbf{e}_j + z(\boldsymbol{\zeta})$. If the j-th entry of $\boldsymbol{\zeta}$ is -1, then $\boldsymbol{\zeta} = -\mathbf{e}_j + z(\boldsymbol{\zeta})$. So, we have that

$$\begin{aligned}\boldsymbol{\pi}_i(\boldsymbol{\beta}, \hat{s}^{t_n}_{b,n}, \hat{s}^{t_n}_{b,n}) &= \tilde{\pi}_i(\boldsymbol{\beta} + b\boldsymbol{\zeta}_i, \hat{s}^{t_n}_{b,n} + b\zeta_i, \hat{s}^{t_n}_{b,n})\\ &= \tilde{\pi}_i(\boldsymbol{\beta} + b\boldsymbol{\zeta}_{i,j}\mathbf{e}_j + b\cdot z(\boldsymbol{\zeta}_i), \hat{s}^{t_n}_{b,n} + b\zeta_i, \hat{s}^{t_n}_{b,n}).\end{aligned}$$

As a result, we have that

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\zeta}_{i,j}\boldsymbol{\pi}_i(\boldsymbol{\beta},\hat{s}_{b,n}^{t_n},\hat{s}_{b,n}^{t_n}) \tag{G.37}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\zeta}_{i,j}\cdot\tilde{\pi}_i(\boldsymbol{\beta}+b\boldsymbol{\zeta}_{i,j}\mathbf{e}_j+b\cdot z(\boldsymbol{\zeta}_i),\hat{s}_{b,n}^{t_n}+b\zeta_i,\hat{s}_{b,n}^{t_n}) \tag{G.38}$$

$$=\sum_{\substack{\boldsymbol{\zeta}\in\{-1,1\}^d \text{ s.t. } \boldsymbol{\zeta}_j=1\\ \zeta\in\{-1,1\}}}\frac{n_{\boldsymbol{\zeta},\zeta}}{n}\sum_{i=1}^{n_{\boldsymbol{\zeta},\zeta}}\tilde{\pi}_i(\boldsymbol{\beta}+b\mathbf{e}_j+b\cdot z(\boldsymbol{\zeta}),\hat{s}_{b,n}^{t_n}+b\zeta,\hat{s}_{b,n}^{t_n}) \tag{G.39}$$

$$-\sum_{\substack{\boldsymbol{\zeta}\in\{-1,1\}^d \text{ s.t. } \boldsymbol{\zeta}_j=-1\\ \zeta\in\{-1,1\}}}\frac{n_{\boldsymbol{\zeta},\zeta}}{n}\sum_{i=1}^{n_{\boldsymbol{\zeta},\zeta}}\tilde{\pi}_i(\boldsymbol{\beta}-b\mathbf{e}_j+b\cdot z(\boldsymbol{\zeta}),\hat{s}_{b,n}^{t_n}+b\zeta,\hat{s}_{b,n}^{t_n}) \tag{G.40}$$

To establish convergence properties of terms in the double sums in (G.39) and (G.40), we must establish stochastic equicontinuity of the collection of stochastic processes $\{\tilde{\pi}_n(\boldsymbol{\beta},s,r)\}$ indexed by $(s,r)\in\mathcal{S}\times\mathcal{S}$. Because $\mathcal{S}\times\mathcal{S}$ is compact and $\tilde{\Pi}(\boldsymbol{\beta},s;r)$ is continuous in (s,r), then we can show that $\{\tilde{\pi}_n(\boldsymbol{\beta},s,r)\}$ by showing that $\tilde{\pi}_n(\boldsymbol{\beta},s,r)$ converges uniformly in probability to $\tilde{\Pi}(\boldsymbol{\beta},s;r)$ (Lemma 27). We can use Lemma 26 to show the necessary uniform convergence result.

By Lemma 48, we have that $\tilde{\pi}$ satisfies the conditions of Lemma 26. Thus, we can apply Lemma 26 to establish uniform convergence in probability of $\tilde{\pi}_n(\boldsymbol{\beta},s,r)$ with respect to (s,r). As a consequence, the collection of stochastic processes $\{\tilde{\pi}_n(\boldsymbol{\beta},s,r)\}$ is stochastically equicontinuous. In particular, $\tilde{\pi}_n(\boldsymbol{\beta},s,r)$ is stochastically equicontinuous at $(s(\boldsymbol{\beta},b),s(\boldsymbol{\beta},b))$, where $s(\boldsymbol{\beta},b)$ is the unique fixed point of $q(P_{\boldsymbol{\beta},s,b})$ (see Lemma 51). By Lemma 52, we have that

$$\hat{s}_{b,n}^{t_n}\xrightarrow{p}s(\boldsymbol{\beta},b).$$

Now, we can apply Lemma 28 to establish convergence in probability for the terms in (G.39) and (G.40). As an example, for a perturbation $\boldsymbol{\zeta}\in\{-1,1\}^d$ with j-th entry equal to 1 and arbitrary $\zeta\in\{-1,1\}$, Lemma 28 gives that

$$\begin{aligned}&\tilde{\pi}_{n_{\boldsymbol{\zeta},\zeta}}(\boldsymbol{\beta}+b\mathbf{e}_j+b\cdot z(\boldsymbol{\zeta}),\hat{s}_{b,n}^{t_n}+b\zeta,\hat{s}_{b,n}^{t_n})\\ &\xrightarrow{p}\tilde{\pi}_{n_{\boldsymbol{\zeta},\zeta}}(\boldsymbol{\beta}+b\mathbf{e}_j+b\cdot z(\boldsymbol{\zeta}),s(\boldsymbol{\beta},b)+b\zeta,s(\boldsymbol{\beta},b)),\end{aligned}$$

and by the Weak Law of Large Numbers, we have that

$$\begin{aligned}&\tilde{\pi}_{n_{\boldsymbol{\zeta},\zeta}}(\boldsymbol{\beta}+b\mathbf{e}_j+b\cdot z(\boldsymbol{\zeta}),s(\boldsymbol{\beta},b)+b\zeta,s(\boldsymbol{\beta},b))\\ &\xrightarrow{p}\tilde{\Pi}(\boldsymbol{\beta}+b\mathbf{e}_j+b\cdot z(\boldsymbol{\zeta}),s(\boldsymbol{\beta},b)+b\zeta,s(\boldsymbol{\beta},b)).\end{aligned}$$

Analogous results for the remaining terms in (G.39) and (G.40). Also,

$$\frac{n_{\boldsymbol{\zeta},\zeta}}{n}\xrightarrow{p}\frac{1}{2^{d+1}},\quad \boldsymbol{\zeta}\in\{-1,1\}^d,\zeta\in\{-1,1\}.$$

By Slutsky's Theorem, when any j and b fixed, we have

$$\mathbf{s}_{zy,j}^{n}\xrightarrow{p}\sum_{\substack{\boldsymbol{\zeta}\in\{-1,1\}^d \text{ s.t. } \boldsymbol{\zeta}_j=1\\ \zeta\in\{-1,1\}}}\frac{\tilde{\Pi}(\boldsymbol{\beta}+b\mathbf{e}_j+b\cdot z(\boldsymbol{\zeta}),s(\boldsymbol{\beta},b)+b\zeta,s(\boldsymbol{\beta},b))}{2^{d+1}\cdot b} \tag{G.41}$$

$$-\sum_{\substack{\boldsymbol{\zeta}\in\{-1,1\}^d \text{ s.t. } \boldsymbol{\zeta}_j=-1\\ \zeta\in\{-1,1\}}}\frac{\tilde{\Pi}(\boldsymbol{\beta}-b\mathbf{e}_j+b\cdot z(\boldsymbol{\zeta}),s(\boldsymbol{\beta},b)+b\zeta,s(\boldsymbol{\beta},b))}{2^{d+1}\cdot b}. \tag{G.42}$$

Let R_b denote the expression on the right side of the above equation. If there is a sequence $\{b_n\}$ such that $b_n \to 0$, then by Lemma 51, $s(\boldsymbol{\beta}, b_n) \to s(\boldsymbol{\beta})$, where $s(\boldsymbol{\beta})$ is the unique fixed point of $q(P_{\boldsymbol{\beta},s})$. Since $\tilde{\Pi} = \Pi$ and by continuity of Π,

$$R_{b_n} \to \frac{\partial \Pi}{\partial \boldsymbol{\beta}_j}(\boldsymbol{\beta}, s(\boldsymbol{\beta}); s(\boldsymbol{\beta})).$$

Using the definition of convergence in probability, we show that there exists such a sequence $\{b_n\}$. From (G.41) and (G.42), we have that for each $\epsilon, \delta > 0$ and $b > 0$ and sufficiently small, there exists $n(\epsilon, \delta, b)$ such that for $n \geq n(\epsilon, \delta, b)$

$$P(|\mathbf{s}^n_{zy,j} - R_b| \leq \epsilon) \geq 1 - \delta.$$

So, we can fix $\delta > 0$. For $k = 1, 2, \ldots$, let $N(k) = n(\frac{1}{k}, \delta, \frac{1}{k})$. Then, we can define a sequence such that $b_n = \epsilon_n = \frac{1}{k}$ for all $N(k) \leq n \leq N(k+1)$. So, we have that $\epsilon_n \to 0$ and $b_n \to 0$. Thus, we have that

$$\mathbf{s}^n_{zy,j} \xrightarrow{p} \frac{\partial \Pi}{\partial \boldsymbol{\beta}_j}(\boldsymbol{\beta}, s(\boldsymbol{\beta}); s(\boldsymbol{\beta})).$$

Considering all indices j,

$$\mathbf{s}^n_{zy} \xrightarrow{p} \frac{\partial \Pi}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s(\boldsymbol{\beta}); s(\boldsymbol{\beta})).$$

It remains to establish convergence in probability for $\mathbf{S}_{zz}$. We have that

$$\begin{aligned}
\mathbf{S}^n_{zz} &= \frac{1}{b_n^2 n}\mathbf{Z}^T\mathbf{Z} \\
&= \frac{1}{b_n^2 n}\sum_{i=1}^n (b_n \boldsymbol{\zeta}_i)^T (b_n \boldsymbol{\zeta}_i). \\
&= \frac{1}{n}\sum_{i=1}^n \boldsymbol{\zeta}_i^T \boldsymbol{\zeta}_i.
\end{aligned}$$

We note that

$$\mathbb{E}_{\boldsymbol{\zeta}_i \sim R^d}\left[\boldsymbol{\zeta}_{i,j}\boldsymbol{\zeta}_{i,k}\right] = \begin{cases} 1 \text{ if } j = k \\ 0 \text{ if } j \neq k \end{cases}$$

because $\boldsymbol{\zeta}_i$ is a vector of independent Rademacher random variables. So, $\mathbb{E}\left[\boldsymbol{\zeta}_i^T\boldsymbol{\zeta}_i\right] = \mathbf{I}_d$. By the Weak Law of Large Numbers, we have that

$$\mathbf{S}^n_{zz} \xrightarrow{p} \mathbf{I}_d.$$

Finally, we can use Slutsky's Theorem to show that

$$\hat{\Gamma}^n_{\boldsymbol{\pi},\boldsymbol{\beta}}(\boldsymbol{\beta}, \hat{s}^{t_n}_n; \hat{s}^{t_n}_n) = (\mathbf{S}^n_{zz})^{-1}\mathbf{s}^n_{zy} \xrightarrow{p} (I_d)^{-1}\frac{\partial \Pi}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s(\boldsymbol{\beta}); s(\boldsymbol{\beta})) = \frac{\partial \Pi}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s(\boldsymbol{\beta}); s(\boldsymbol{\beta})).$$

G.24 Proof of Corollary 55

The proof of this result is analogous to Lemma 54.

G.25 Proof of Lemma 56

We study the convergence of the kernel density estimate $p^n_{\boldsymbol{\beta},\hat{s}^{t_n}_n, b_n}(\hat{s}^{t_n}_{b_n,n})$. Let $p^n_{\boldsymbol{\beta},s,b}(r)$ is a kernel density estimate of density of $P_{\boldsymbol{\beta},s,b}$ at a point r. Let $\boldsymbol{\beta}_i = \boldsymbol{\beta} + b\boldsymbol{\zeta}_i$, $s_i = s + b\zeta_i$, where $\boldsymbol{\zeta} \sim R^d$ and $\zeta \sim R$. We can write the explicit form of $p^n_{\boldsymbol{\beta},s,b}(r)$ as follows

$$p^n_{\boldsymbol{\beta},s,b}(r) = \frac{1}{h_n}\sum_{i=1}^n k\Big(\frac{r - \boldsymbol{\beta}_i^T \mathbf{x}(\boldsymbol{\beta}_i, s_i, \nu_i) + b\zeta_i}{h_n}\Big).$$

For sufficiently small b, we can apply Lemma 50 to map the types $\nu \sim F$ and cost functions c_ν to types $\nu'_{b,\boldsymbol{\zeta},\zeta} \sim \tilde{F}_b$ and cost functions $c_{\nu'}$, so that when the agent with type $\nu'_{b,\boldsymbol{\zeta},\zeta}$ best responds to the unperturbed model and threshold, they obtain the same raw score (without noise) as the agent with type ν who responds to a perturbed model and threshold. So, we can write

$$p^n_{\boldsymbol{\beta},s,b}(r) = \frac{1}{h_n}\sum_{i=1}^n k\Big(\frac{r - \boldsymbol{\beta}^T \mathbf{x}(\boldsymbol{\beta}, s, \nu'_i)}{h_n}\Big) = \frac{1}{h_n}\sum_{i=1}^n k\Big(\frac{r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_i) - \boldsymbol{\beta}^T\boldsymbol{\epsilon}_i}{h_n}\Big).$$

Let H denote the joint distribution of $\nu'_{b,\boldsymbol{\zeta},\zeta} \sim \tilde{F}_b$ and $\boldsymbol{\epsilon} \sim N(0, \sigma^2 I_d)$.

$$\begin{aligned}
\mathbf{w}_b &:= (\nu'_{b,\boldsymbol{\zeta},\zeta}, \boldsymbol{\epsilon})\\
\tilde{k}(\mathbf{w}_b, \boldsymbol{\beta}, s, r; h) &:= k\Big(\frac{r - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) - \boldsymbol{\beta}^T\boldsymbol{\epsilon}}{h}\Big)\\
\tilde{k}_n(\boldsymbol{\beta}, s, r; h) &:= \frac{1}{n}\sum_{i=1}^n \tilde{k}(\mathbf{w}_b, \boldsymbol{\beta}, s, r, ; h)\\
K(\boldsymbol{\beta}, s, r; h) &:= \mathbb{E}_{\mathbf{w}\sim H}\left[\tilde{k}(\mathbf{w}_b; \boldsymbol{\beta}, s, r; h)\right].
\end{aligned}$$

We can write

$$p^n_{\boldsymbol{\beta},\hat{s}^{t_n}_{b_n,n},b}(\hat{s}^{t_n}_{b_n,n}) = \frac{1}{h_n}\tilde{k}_n(\boldsymbol{\beta}, \hat{s}^{t_n}_{b_n,n}, \hat{s}^{t_n}_{b_n,n}; h_n).$$

Due to the density estimate's dependence on the stochastic process $\hat{s}^{t_n}_{b_n,n}$, we first must establish the stochastic equicontinuity of the collection of stochastic processes $\{\tilde{k}_n(\boldsymbol{\beta}, s, r)\}$ indexed by $(s, r) \in \mathcal{S}\times\mathcal{S}$. We show stochastic equicontinuity via uniform convergence in probability (Lemma 26). The remainder of the proof follows by the Weak Law of Large Numbers and taking standard limits.

We fix $h_n = h$. Since $\tilde{k}$ satisfies the conditions of Lemma 26, we can apply Lemma 26 to realize that $\tilde{k}_n(\boldsymbol{\beta}, s, r; h)$ converges uniformly in probability to $K(\boldsymbol{\beta}, s, r; h)$ with respect to $(s, r) \in \mathcal{S}\times\mathcal{S}$. As a result, the collection of stochastic processes $\{\tilde{k}_n(\boldsymbol{\beta}, s, r; h)\}$ indexed by $(s, r) \in \mathcal{S}\times\mathcal{S}$ are stochastically equicontinuous. In particular, $\{\tilde{k}_n(\boldsymbol{\beta}, s, r; h)\}$ is stochastically equicontinuous at $(s(\boldsymbol{\beta}, b), s(\boldsymbol{\beta}, b))$. By Lemma 52, we have that

$$\hat{s}^{t_n}_{b,n} \xrightarrow{p} s(\boldsymbol{\beta}, b),$$

where $s(\boldsymbol{\beta}, b)$ is the unique fixed point of $q(P_{\boldsymbol{\beta},s,b})$. we can apply Lemma 28 to see that

$$k_n(\boldsymbol{\beta}, \hat{s}^{t_n}_{b,n}, \hat{s}^{t_n}_{b,n}; h) - k_n(\boldsymbol{\beta}, s(\boldsymbol{\beta}, b), s(\boldsymbol{\beta}, b); h) \xrightarrow{p} 0.$$

Furthermore, by the Weak Law of Large Numbers, we have that

$$k_n(\boldsymbol{\beta}, s(\boldsymbol{\beta}, b), s(\boldsymbol{\beta}, b); h) \xrightarrow{p} K(\boldsymbol{\beta}, s(\boldsymbol{\beta}, b), s(\boldsymbol{\beta}, b); h).$$

Given our definition of the kernel function k and for fixed h, we have that

$$p^n_{\boldsymbol{\beta},\hat{s}^{t_n}_{b,n},b}(\hat{s}^{t_n}_{b_n,n}) \xrightarrow{p} \frac{K(\boldsymbol{\beta}, s(\boldsymbol{\beta}, b), s(\boldsymbol{\beta}, b); h)}{h} = \frac{P_{\boldsymbol{\beta},s(\boldsymbol{\beta},b),b}(s(\boldsymbol{\beta}, b) + \frac{h}{2}) - P_{\boldsymbol{\beta},s(\boldsymbol{\beta},b),b}(s(\boldsymbol{\beta}, b) - \frac{h}{2})}{h}.$$

Given that our sequence $h_n \to 0$ and $nh_n \to \infty$ and k satisfies the assumptions of Theorem 33, we can apply Theorem 33 to see that for each fixed b, we obtain a consistent density estimate.

$$\begin{aligned}
p^n_{\boldsymbol{\beta},\hat{s}^{t_n}_{b,n},b}(\hat{s}^{t_n}_{b,n}) &\xrightarrow{p} \lim_{h_n\to 0}\frac{P_{\boldsymbol{\beta},s(\boldsymbol{\beta},b),b}(s(\boldsymbol{\beta}, b) + \frac{h_n}{2}) - P_{\boldsymbol{\beta},s(\boldsymbol{\beta},b),b}(s(\boldsymbol{\beta}, b) - \frac{h_n}{2})}{h_n} && \text{(G.43)}\\
&= p_{\boldsymbol{\beta},s(\boldsymbol{\beta},b),b}(s(\boldsymbol{\beta}, b)). && \text{(G.44)}
\end{aligned}$$

Let R_b denote the right side of the above equation. Suppose there exists a sequence such that $b_n \to 0$. By Lemma 51, this gives us that $s(\boldsymbol{\beta}, b_n) \to s^*$, where s^* is the unique fixed point of $q(P_{\boldsymbol{\beta},s})$. We can show that

$R_{b_n} \to p_{\boldsymbol{\beta},s^*}(s^*)$ as follows.

$$\begin{aligned} |R_{b_n} - p_{\boldsymbol{\beta},s^*}(s^*)| \leq & |p_{\boldsymbol{\beta},s(\boldsymbol{\beta},b_n),b_n}(s(\boldsymbol{\beta},b_n)) - p_{\boldsymbol{\beta},s(\boldsymbol{\beta},b_n)}(s(\boldsymbol{\beta},b_n))| \\ & + |p_{\boldsymbol{\beta},s(\boldsymbol{\beta},b_n)}(s(\boldsymbol{\beta},b_n)) - p_{\boldsymbol{\beta},s^*}(s^*)| \\ \leq & \sup_{s,r\in\mathcal{S}} |p_{\boldsymbol{\beta},s,b_n}(r) - p_{\boldsymbol{\beta},s}(r)| \\ & + |p_{\boldsymbol{\beta},s(\boldsymbol{\beta},b_n)}(s(\boldsymbol{\beta},b_n)) - p_{\boldsymbol{\beta},s^*}(s^*)|. \end{aligned}$$

Since $p_{\boldsymbol{\beta},s}(r)$ is continuous in s and r (Lemma 42), there exists N such that for $n \geq N$, the second term is less than ϵ. To bound the first term, we require the following lemma.

Lemma 62. *Under Assumptions 1, 2, 3, and 4, if $\sigma^2 > \frac{2}{\alpha_*(F)\cdot\sqrt{2\pi e}}$, then $p_{\boldsymbol{\beta},s,b}(r) \to p_{\boldsymbol{\beta},s}(r)$ uniformly in s and in r as $b \to 0$. Proof in Appendix G.31.*

Due to the uniform convergence result, we have that if there exists a sequence $\{b_n\}$ such that $b_n \to 0$, then

$$R_{b_n} \to p_{\boldsymbol{\beta},s^*}(s^*).$$

It remains to show that there exists such a sequence $\{b_n\}$ where $b_n \to 0$. Using the definition of convergence in probability, we show that there exists such a sequence $\{b_n\}$. From (G.44), we have that for each $\epsilon, \delta > 0$ and $b > 0$ and sufficiently small, there exists $n(\epsilon, \delta, b)$ such that for $n \geq n(\epsilon, \delta, b)$

$$P(|p^n_{\boldsymbol{\beta},\hat{s}^{t_n}_{b_n,n},b_n}(\hat{s}^{t_n}_{b_n,n}) - R_b| \leq \epsilon) \geq 1 - \delta.$$

So, we can fix $\delta > 0$. For $k = 1, 2, \ldots$, let $N(k) = n(\frac{1}{k}, \delta, \frac{1}{k})$. Then, we can define a sequence such that $b_n = \epsilon_n = \frac{1}{k}$ for all $N(k) \leq n \leq N(k+1)$. So, we have that $\epsilon_n \to 0$ and $b_n \to 0$. Finally, this gives that

$$p^n_{\boldsymbol{\beta},\hat{s}^{t_n}_n,b_n}(\hat{s}^{t_n}_n) \xrightarrow{p} p_{\boldsymbol{\beta},s^*}(s^*).$$

G.26 Proof of Lemma 57

First, we note that $s(\boldsymbol{\beta})$ is continuously differentiable in $\boldsymbol{\beta}$ by Corollary 9, so we can use implicit differentiation to compute the following expression for $\frac{\partial s}{\partial \boldsymbol{\beta}}$

$$\frac{\partial s}{\partial \boldsymbol{\beta}} = \frac{1}{1 - \frac{\partial q(P_{\boldsymbol{\beta},s(\boldsymbol{\beta})})}{\partial s}} \cdot \frac{\partial q(P_{\boldsymbol{\beta},s(\boldsymbol{\beta})})}{\partial \boldsymbol{\beta}}. \tag{G.45}$$

After that, we apply the lemma below to express the partial derivatives of the quantile mapping $q(P_{\boldsymbol{\beta},s})$ in terms of partial derivatives of the complementary CDF $\Pi(\boldsymbol{\beta}, s; r)$.

Lemma 63. *Let $\boldsymbol{\beta} \in \mathcal{B}$, $s \in \mathcal{S}$. Under Assumption 1, 2, 3, if $\sigma^2 > \frac{2}{\alpha_*\cdot\sqrt{2\pi e}}$ then for $\boldsymbol{\beta}^t, s^t$ sufficiently close to $\boldsymbol{\beta}, s$, the derivative of $q(P_{\boldsymbol{\beta},s})$ with respect to a one-dimensional parameter θ is given by*

$$\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial \theta} = \frac{1}{p_{\boldsymbol{\beta}^t,s^t}(r^t)} \cdot \frac{\partial \Pi}{\partial \theta}(\boldsymbol{\beta}, s; r^t),$$

where $r^t = q(P_{\boldsymbol{\beta}^t,s^t})$. Proof in Appendix G.32.

Since $s(\boldsymbol{\beta})$ is the fixed point induced by $\boldsymbol{\beta}$, we have that

$$s(\boldsymbol{\beta}) - q(P_{\boldsymbol{\beta},s(\boldsymbol{\beta})}) = 0.$$

From Corollary 9, we have that $s(\boldsymbol{\beta})$ is continuously differentiable in $\boldsymbol{\beta}$. Differentiating both sides of the above equation with respect to $\boldsymbol{\beta}$ yields

$$\frac{\partial s}{\partial \boldsymbol{\beta}} - \Big(\frac{\partial q(P_{\boldsymbol{\beta},s(\boldsymbol{\beta})})}{\partial \boldsymbol{\beta}} + \frac{\partial q(P_{\boldsymbol{\beta},s(\boldsymbol{\beta})})}{\partial s} \cdot \frac{\partial s}{\partial \boldsymbol{\beta}}\Big) = 0.$$

Rearranging the above equation yields (F.15), which shows that $\frac{\partial s}{\partial \boldsymbol{\beta}}$ in terms of $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s}$ and $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial \boldsymbol{\beta}}$. From Lemma 63, we have that for $\boldsymbol{\beta}^t, s^t$ sufficiently close to $\boldsymbol{\beta}, s$, we have that

$$\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial s} = \frac{1}{p_{\boldsymbol{\beta}^t,s^t}(r^t)} \cdot \frac{\partial \Pi}{\partial s}(\boldsymbol{\beta}, s; r^t)$$
$$\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial \boldsymbol{\beta}} = \frac{1}{p_{\boldsymbol{\beta}^t,s^t}(r^t)} \cdot \frac{\partial \Pi}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s; r^t),$$

where $r^t = q(P_{\boldsymbol{\beta}^t,s^t})$. Let $s^* = s(\boldsymbol{\beta})$. Suppose that we aim to estimate the derivative when the model parameters are $\boldsymbol{\beta}$ and the threshold is s^*. If we consider $\boldsymbol{\beta}^t = \boldsymbol{\beta}$, $s^t = s^*$, then $r^t = s^*$. So, we have that

$$\frac{\partial q(P_{\boldsymbol{\beta},s^*})}{\partial s} = -\frac{1}{p_{\boldsymbol{\beta},s^*}(s^*)} \cdot \frac{\partial \Pi}{\partial s}(\boldsymbol{\beta}, s^*; s^*) \tag{G.46}$$
$$\frac{\partial q(P_{\boldsymbol{\beta},s^*})}{\partial \boldsymbol{\beta}} = -\frac{1}{p_{\boldsymbol{\beta},s^*}(s^*)} \cdot \frac{\partial \Pi}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s^*; s^*). \tag{G.47}$$

Substituting (G.46) and (G.47) into (G.45) yields

$$\frac{\partial s}{\partial \boldsymbol{\beta}} = \frac{1}{p_{\boldsymbol{\beta},s^*}(s^*) - \frac{\partial \Pi}{\partial s}(\boldsymbol{\beta}, s^*; s^*)} \cdot \frac{\partial \Pi}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}, s^*; s^*).$$

G.27 Proof of Lemma 58

By Lemma 38, $\mathbf{H}$ is positive definite and invertible. As a result, we can apply the Sherman-Morrison Formula (Theorem 20) to
$(\mathbf{H} + G''(s - \boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta}\boldsymbol{\beta}^T)^{-1}$: let $\mathbf{A} = \mathbf{H}$, $\mathbf{u} = G''(s - \boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta}$, and $\mathbf{v} = \boldsymbol{\beta}$.

$$\begin{aligned}(\mathbf{H} + G''(s - \boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta}\boldsymbol{\beta}^T)^{-1} &= \mathbf{H}^{-1} - \frac{\mathbf{H}^{-1}(G''(s - \boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta})\boldsymbol{\beta}^T\mathbf{H}^{-1}}{1 + \boldsymbol{\beta}^T\mathbf{H}^{-1}(G''(s - \boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta})} \\ &= \mathbf{H}^{-1} - \frac{G''(s - \boldsymbol{\beta}^T\mathbf{x})\mathbf{H}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{H}^{-1}}{1 + G''(s - \boldsymbol{\beta}^T\mathbf{x})\boldsymbol{\beta}^T\mathbf{H}^{-1}\boldsymbol{\beta}}.\end{aligned}$$

G.28 Proof of Lemma 59

In the first part of the proof, we establish existence of a fixed point of $\omega(s; \boldsymbol{\beta}, \nu)$. In the second part of the proof, we show that if a fixed point exists, then it must be unique.

First, we use the IVT to show existence of a fixed point. We apply the IVT to the function $h(s; \boldsymbol{\beta}, \nu) = s - \omega(s; \boldsymbol{\beta}, \nu)$. We note that by Lemma 2 that $\omega(s)$ is continuous. It remains to show that there exists s_l such that $h(s_1) < 0$ and there exists s_2 such that $s_2 > s_1$ and $h(s_2) > 0$. Then, by the Intermediate Value Theorem, there must be $s \in [s_1, s_2]$ for which $h(s) = 0$, which gives that $\omega(s)$ has at least one fixed point.

Let $\delta > 0$. By Lemma 40, we have that there exists $S_{l,1}$ so that for all $s \leq S_l$, we have that

$$|\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) - \boldsymbol{\beta}^T\boldsymbol{\eta}| < \delta.$$

Let $S_{l,2} = \boldsymbol{\beta}^T\boldsymbol{\eta} - \delta$. Let $s_1 < \min(S_{l,1}, S_{l,2})$. Then we have that

$$\begin{aligned}h(s_1) &= s_1 - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s_1, \nu) \\ &\leq s_1 - \boldsymbol{\beta}^T\boldsymbol{\eta} + \delta \\ &< (\boldsymbol{\beta}^T\boldsymbol{\eta} - \delta) - \boldsymbol{\beta}^T\boldsymbol{\eta} + \delta. \\ &< 0.\end{aligned}$$

Second, by Lemma 40, we have that there exists $S_{h,1}$ so that for all $s \geq S_l$, we have that

$$|\boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu) - \boldsymbol{\beta}^T \boldsymbol{\eta}| < \delta.$$

Let $S_{h,2} = \boldsymbol{\beta}^T \boldsymbol{\eta} + \delta$. Let $s_2 > \max(S_{h,1}, S_{h,2})$. Then we have that

$$\begin{aligned} h(s_2) &= s_2 - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s_2, \nu) \\ &\geq s_2 - \boldsymbol{\beta}^T \boldsymbol{\eta} - \delta \\ &> 0. \end{aligned}$$

We have that $s_1 < \boldsymbol{\beta}^T \boldsymbol{\eta} - \delta < \boldsymbol{\beta}^T \boldsymbol{\eta} + \delta < s_2$. By the IVT, there must be some $s \in [s_1, s_2]$ so that $h(s) = 0$. Second, we show that if a fixed point exists, then the fixed point must be unique. By Lemma 39, $h(s; \boldsymbol{\beta}, \nu)$ is strictly increasing in s. There can be only one point at which $h(s; \boldsymbol{\beta}, \nu) = 0$. So, there is only one s such that $s - \omega(s; \boldsymbol{\beta}, \nu) = 0$. Thus, $\omega(s)$ has a unique fixed point.

G.29 Proof of Lemma 60

Since $\mathbf{y}$ is in the interior of $\mathcal{X}$, then there exists some $\epsilon > 0$ such that the open ball of radius ϵ about $\mathbf{y}$ is a subset of $\mathcal{X}$. We note that

$$\begin{aligned} |\mathbf{y}' - \mathbf{y}| &= \left| \boldsymbol{\beta} \cdot (b\boldsymbol{\zeta}^T \mathbf{x} - b\boldsymbol{\zeta}) \right| \\ &\leq ||\boldsymbol{\beta}|| \cdot \left| b\boldsymbol{\zeta}^T \mathbf{x} - b\boldsymbol{\zeta} \right| \\ &\leq |b\boldsymbol{\zeta}^T \mathbf{x} - b\zeta| \\ &\leq b|\boldsymbol{\zeta}^T \mathbf{x} - \zeta| \\ &\leq b(|\boldsymbol{\zeta}||\mathbf{x}| + |\zeta|) \\ &\leq b(\sqrt{d} \cdot \sup_{x \in \mathcal{X}} |\mathbf{x}| + 1) \end{aligned}$$

Since $\mathcal{X}$ is compact, we can say that the supremum in the above equation is achieved on $\mathcal{X}$ and we can call its value m. So, if $b < \frac{\epsilon}{(m\sqrt{d}+1)}$, then $\mathbf{y}' \in \text{Int}(\mathcal{X})$.

G.30 Proof of Lemma 61

We first show that $P_{\boldsymbol{\beta},s,b}(r) \to P_{\boldsymbol{\beta},s}(r)$ uniformly in r as $b \to 0$. We aim to apply Lemma 29. First, note that the continuity of $P_{\boldsymbol{\beta},s}$ in r is given by Lemma 42. We recall that

$$P_{\boldsymbol{\beta},s,b}(r) = \frac{1}{2^{d+1}} \sum_{\substack{\boldsymbol{\zeta} \in \{-1,1\}^d \\ \zeta \in \{-1,1\}}} \int G\Big(r - (\boldsymbol{\beta} + b\boldsymbol{\zeta})^T \mathbf{x}^*(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) + b\zeta\Big) dF.$$

G is strictly increasing, so $P_{\boldsymbol{\beta},s,b}(r)$ is strictly increasing because the sum of strictly increasing functions is also strictly increasing. By continuity of $\mathbf{x}$ in $\boldsymbol{\beta}$ and s (Lemma 2), we have that $P_{\boldsymbol{\beta},s,b}(r) \to P_{\boldsymbol{\beta},s}(r)$ pointwise in r. By Lemma 29, as $b \to 0$, we have that

$$\sup_{r \in \mathcal{S}} |P_{\boldsymbol{\beta},s,b}(r) - P_{\boldsymbol{\beta},s}(r)| \to 0.$$

G.31 Proof of Lemma 62

We show that $p_{\boldsymbol{\beta},s,b}(r) \to p_{\boldsymbol{\beta},s}(r)$ uniformly in s and r as $b \to 0$. We prove the claim in two steps. First, we rewrite $p_{\boldsymbol{\beta},s}(r)$ and $p_{\boldsymbol{\beta},s,b}(r)$ as a finite sum of terms that align by type and perturbation. Second, we can show uniform convergence for pairs of terms in the sums, which gives that the aggregate quantity $p_{\boldsymbol{\beta},s,b}(r) \to p_{\boldsymbol{\beta},s}(r)$ uniformly.

First, we rewrite $p_{\boldsymbol{\beta},s}$ as follows

$$p_{\boldsymbol{\beta},s}(r) = \int_{\mathcal{X}\times\mathcal{G}} G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu))dF \tag{G.48}$$

$$= \sum_{\nu\in\text{supp}(F)} \sum_{\substack{\boldsymbol{\zeta}\in\{-1,1\}^d \\ \zeta\in\{-1,1\}}} G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)) \frac{f(\nu)}{2^{d+1}}. \tag{G.49}$$

To rewrite $p_{\boldsymbol{\beta},s,b}$, recall that for sufficiently small b, we can use Lemma 50 to express $P_{\boldsymbol{\beta},s,b}$ as the score distribution induced by agents with types $\nu'_{b,\boldsymbol{\zeta},\zeta} \sim \tilde{F}_b$ and cost functions $c_{\nu'}$ who best respond to a model $\boldsymbol{\beta}$ and threshold s. The type and cost function is given by transformation T from Lemma 60. Recall that $T : (\nu, c_\nu, \boldsymbol{\zeta}, \zeta, b) \to (\nu'_{b,\boldsymbol{\zeta},\zeta}, c_{\nu'})$. Let T_1 be defined as

$$T_1 : (\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta) \to \nu'_{b,\boldsymbol{\zeta},\zeta}.$$

Since our assumed conditions also transfer to $\tilde{F}_b$, we have that that $P_{\boldsymbol{\beta},s,b}(r)$ is continuously differentiable in r with density $p_{\boldsymbol{\beta},s,b}$ (Lemma 42). Using the function T_1, we have that

$$p_{\boldsymbol{\beta},s,b}(r) = \int_{\mathcal{X}\times\mathcal{G}} G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu))d\tilde{F}_b \tag{G.50}$$

$$= \sum_{\nu'_{b,\boldsymbol{\zeta},\zeta}\in\text{supp}(\tilde{F}_b)} G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta})) \cdot \tilde{f}_b(\nu'_{b,\boldsymbol{\zeta},\zeta}) \tag{G.51}$$

$$= \sum_{\nu\sim\text{supp}(F)} \sum_{\substack{\boldsymbol{\zeta}\in\{-1,1\}^d \\ \zeta\in\{-1,1\}}} G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, T_1(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta))) \cdot \tilde{f}_b(T_1(\nu, c_\nu; \boldsymbol{\zeta}, \zeta, b)) \tag{G.52}$$

$$= \sum_{\nu\sim\text{supp}(F)} \sum_{\substack{\boldsymbol{\zeta}\in\{-1,1\}^d \\ \zeta\in\{-1,1\}}} G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, T_1(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta))) \cdot \frac{f(\nu)}{2^{d+1}}. \tag{G.53}$$

The last line follows from Lemma 50, where we have that

$$\tilde{f}_b(T_1(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta)) = \frac{f(\nu)}{2^{d+1}} \quad \nu \sim F.$$

Therefore, the terms of $p_{\boldsymbol{\beta},s}$ in (G.49) align with the terms of $p_{\boldsymbol{\beta},s,b}$ in (G.53) by type and perturbation. We have that

$$|p_{\boldsymbol{\beta},s,b}(r) - p_{\boldsymbol{\beta},s}(r)|$$
$$= \Big| \sum_{\nu\in\text{supp}(F)} \sum_{\substack{\boldsymbol{\zeta}\in\{-1,1\}^d \\ \zeta\in\{-1,1\}}} (G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, T_1(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta))) - G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu))) \cdot \frac{f(\nu)}{2^{d+1}} \Big|$$
$$\leq \sum_{\substack{\nu\in\text{supp}(F) \\ \boldsymbol{\zeta}\in\{-1,1\}^d \\ \zeta\in\{-1,1\}}} \Big| G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, T_1(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta))) - G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)) \Big| \cdot \frac{f(\nu)}{2^{d+1}}.$$

Since the sum in the above inequality is finite, we can show $p_{\boldsymbol{\beta},s,b}(r) \to p_{\boldsymbol{\beta},s}(r)$ uniformly in s, r if we can show that for every type ν and perturbation $(\boldsymbol{\zeta}, \zeta)$, we have that

$$\sup_{(s,r)\in\mathcal{S}\times\mathcal{S}} |G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, T_1(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta))) - G'(r - \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu))| \to 0.$$

Now, we can use the following lemma to show uniform convergence (in s and r) of the arguments to G' in the above expression.

Lemma 64. *Suppose Assumption 1 and 4 hold. Let $\nu \sim F$ and c_ν be a cost function. Let $\nu'_{b,\boldsymbol{\zeta},\zeta}, c_{\nu'} = T(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta)$, where T is as defined in Lemma 49 for any $\boldsymbol{\zeta} \in \{-1,1\}^d, \zeta \in \{-1,1\}$, and $b > 0$ and sufficiently small. As $b \to 0$, $\boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) \to \boldsymbol{\beta}^T \mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ uniformly in s. Proof in Appendix G.33.*

We observe that

$$\begin{aligned}
&\sup_{(s,r)\in\mathcal{S}\times\mathcal{S}} |(r - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, T_1(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta))) - (r - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu))| \\
&= \sup_{(s,r)\in\mathcal{S}\times\mathcal{S}} |\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, T_1(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta)) - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu))| \\
&= \sup_{s\in\mathcal{S}} |\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, T_1(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta)) - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)| \to 0,
\end{aligned}$$

where the uniform convergence in the last line follows from Lemma 64. Since the argument to G' in (G.53) converges uniformly in s and r, the argument to G' is uniformly bounded. So, we can restrict the domain of G' to an closed interval on which it is uniformly continuous. As a result, we also have that

$$\sup_{(s,r)\in\mathcal{S}\times\mathcal{S}} |G'(r - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, T_1(\nu, c_\nu; b, \boldsymbol{\zeta}, \zeta))) - G'(r - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu))| \to 0,$$

which concludes the proof.

G.32 Proof of Lemma 63

For simplicity, we can write

$$r^t = q(P_{\boldsymbol{\beta}^t,s^t}) = P^{-1}_{\boldsymbol{\beta}^t,s^t}(q), \quad r = q(P_{\boldsymbol{\beta},s}) = P^{-1}_{\boldsymbol{\beta},s}(q).$$

From Lemma 5, we have the $q(P_{\boldsymbol{\beta},s})$ is continuous in $\boldsymbol{\beta}, s$. In addition, we note that the density of the scores $p_{\boldsymbol{\beta},s}(y)$ is continuous with respect to $\boldsymbol{\beta}, s, y$ (Lemma 42). By the continuity of the density of the scores and the quantile mapping, we can choose $\boldsymbol{\beta}^t, s^t$ sufficiently close to $\boldsymbol{\beta}, s$ such that $|r - r^t| < \epsilon$ and $|p_{\boldsymbol{\beta},s}(r^t) - p_{\boldsymbol{\beta}^t,s^t}(r^t)| < \epsilon$.

From Lemma 42, we have that $P_{\boldsymbol{\beta},s}$ and $P_{\boldsymbol{\beta}^t,s^t}$ have unique inverses. So, the quantile mapping is uniquely defined, which means

$$P_{\boldsymbol{\beta}^t,s^t}(r^t) = q, \quad P_{\boldsymbol{\beta},s}(r) = q.$$

As a result, we have that $P_{\boldsymbol{\beta}^t,s^t}(r^t) = P_{\boldsymbol{\beta},s}(r)$. Without loss of generality, suppose that $r > r^t$,

$$\begin{aligned}
P_{\boldsymbol{\beta}^t,s^t}(r^t) - P_{\boldsymbol{\beta},s}(r^t) &= P_{\boldsymbol{\beta},s}(r) - P_{\boldsymbol{\beta},s}(r) \\
&= \int_{-\infty}^{r} p_{\boldsymbol{\beta},s}(y)dy - \int_{-\infty}^{t} p_{\boldsymbol{\beta},s}(y)dy \\
&= \int_{r^t}^{r} p_{\boldsymbol{\beta},s}(y)dy \\
&= (r - r^t)p_{\boldsymbol{\beta},s}(r^t) + o(|r^t - r|) \\
&= (r - r^t)p_{\boldsymbol{\beta}^t,s^t}(r^t) + o((r - r^t)|p_{\boldsymbol{\beta},s}(r^t) - p_{\boldsymbol{\beta}^t,s^t}(r^t)|) + o(|r - r^t|) \\
&= (r - r^t)p_{\boldsymbol{\beta}^t,s^t}(r^t) + o(|r - r^t| \cdot |p_{\boldsymbol{\beta},s}(r^t) - p_{\boldsymbol{\beta}^t,s^t}(r^t)|) + o(|r - r^t|) \\
&= (r - r^t)p_{\boldsymbol{\beta}^t,s^t}(r^t) + o(\epsilon^2) + o(\epsilon) \\
&= (q(P_{\boldsymbol{\beta},s}) - q(P_{\boldsymbol{\beta}^t,s^t}))p_{\boldsymbol{\beta}^t,s^t}(r^t) + o(\epsilon^2) + o(\epsilon)
\end{aligned}$$

We can differentiate both sides of the above equation with respect to a one-dimensional parameter θ.

$$-\frac{\partial P_{\boldsymbol{\beta},s}(r^t)}{\partial \theta} = \frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial \theta} \cdot p_{\boldsymbol{\beta}^t,s^t}(r^t).$$

From the Definition of $\Pi(\boldsymbol{\beta}, s; r)$ in (5.3), we observe that

$$\frac{\partial P_{\boldsymbol{\beta},s}(r)}{\partial\theta} = -\frac{\partial\Pi}{\partial\theta}(\boldsymbol{\beta}, s; r).$$

Solving for $\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial\theta}$, we find that

$$\frac{\partial q(P_{\boldsymbol{\beta},s})}{\partial\theta} = \frac{1}{p_{\boldsymbol{\beta}^t,s^t}(r^t)} \cdot \frac{\partial\Pi}{\partial\theta}(\boldsymbol{\beta}, s; r^t).$$

G.33 Proof of Lemma 64

Consider b sufficiently small so that the transformation in Lemma 49 is possible. Let $\eta'_{b,\boldsymbol{\zeta},\zeta}$ be as defined in (F.2). Let $h_b : \mathcal{S} \to \mathbb{R}$, where $h_b(s) := s - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta})$ and $h(s) := s - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$. It is sufficient to show that $h_b(s) \to h(s)$ uniformly in s because

$$\begin{aligned}\sup_{s\in\mathcal{S}} |\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)| &= \sup_{s\in\mathcal{S}} |s - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) - s + \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)| \\ &= \sup_{s\in\mathcal{S}} |h_b(s) - h(s)|.\end{aligned}$$

We aim to apply Lemma 29 to show $h_b \to h$ uniformly. We have that $\mathcal{S}$ compact. By Lemma 2, we have that $h(s)$ is continuous. By Lemma 39, we have that each h_b strictly increasing in s. In addition, we have the following pointwise convergence

$$\begin{aligned}\lim_{b\to 0} h_b(s) &= \lim_{b\to 0} s - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) && \text{(G.54)}\\ &= \lim_{b\to 0} s - \boldsymbol{\beta}^T\Big(\mathbf{x}(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) + b \cdot \boldsymbol{\beta}(\boldsymbol{\zeta}^T\mathbf{x}^*(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) - \zeta)\Big) && \text{(G.55)}\\ &= \lim_{b\to 0} s - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) - b \cdot (\boldsymbol{\zeta}^T\mathbf{x}^*(\boldsymbol{\beta} + b\boldsymbol{\zeta}, s + b\zeta, \nu) - \zeta) && \text{(G.56)}\\ &= s - \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu) && \text{(G.57)}\\ &= h(s). && \text{(G.58)}\end{aligned}$$

(G.55) follows from Lemma 49, which gives an explicit expression for $\mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta})$. (G.57) follows from continuity of the best response mapping in $\boldsymbol{\beta}, s$ (Lemma 2). Thus, $h_b \to h$ uniformly, so we have that $\boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu'_{b,\boldsymbol{\zeta},\zeta}) \to \boldsymbol{\beta}^T\mathbf{x}^*(\boldsymbol{\beta}, s, \nu)$ uniformly in s.