Zero-Sum Stochastic Stackelberg Games

Denizalp Goktas
Department of Computer Science
Brown University
Providence, RI 02906, USA
denizalp_goktas@brown.edu

Jiayi Zhao
Department of Computer Science
Pomona College
Pomona, CA, USA
jzae2019@mymail.pomona.edu

Amy Greenwald
Brown University
Providence, RI 02906, USA
amy_greenwald@brown.edu

Abstract

Min-max optimization problems (i.e., zero-sum games) have been used to model problems in a variety of fields in recent years, from machine learning to economics. The literature to date has mostly focused on static zero-sum games, assuming independent strategy sets. In this paper, we study a form of dynamic zero-sum games, called stochastic games, with dependent strategy sets. Just as zero-sum games with dependent strategy sets can be interpreted as zero-sum Stackelberg games, stochastic zero-sum games with dependent strategy sets can be interpreted as zero-sum stochastic Stackelberg games. We prove the existence of an optimal solution in zero-sum stochastic Stackelberg games (i.e., a recursive Stackelberg equilibrium), provide necessary and sufficient conditions for a solution to be optimal, and show that a recursive Stackelberg equilibrium can be computed in polynomial time via value iteration. Finally, we show that stochastic Stackelberg games can model the problem of pricing and allocating goods across agents and time; more specifically, we propose a stochastic Stackelberg game whose solutions correspond to a recursive competitive equilibrium in a stochastic Fisher market. We close with a series of experiments which confirm our theoretical results and show how value iteration performs in practice.

Min-max optimization has paved the way for recent progress in a variety of fields, from machine learning to economics. These applications require computing solutions to **the min-max operator**, i.e., solving a **constrained min-max optimization problem**, also known as zero-sum games (with independent strategy sets): i.e., $\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}} f(\boldsymbol{x},\boldsymbol{y})$, where the objective function $f : \mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ is continuous, and the strategy sets $\mathcal{X}\subset\mathbb{R}^n$ and $\mathcal{Y}\subset\mathbb{R}^m$ are nonempty and compact. When f is convex-concave, the seminal minimax theorem (Neumann, 1928; Sion et al., 1958) holds, and such a problem can be interpreted as a simultaneous-move zero-sum game between an outer player $\boldsymbol{x}$ and an inner player $\boldsymbol{y}$, with the solutions $(\boldsymbol{x}^*,\boldsymbol{y}^*)\in\mathcal{X}\times\mathcal{Y}$ of the min-max operator corresponding to a Nash equilibrium. More generally, one can consider **zero-sum stochastic games** (with independent strategy sets) $\mathcal{X}\subset\mathbb{R}^n$ and $\mathcal{Y}\subset\mathbb{R}^m$, nonempty and compact, and a state-dependent payoff function $r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$, for all $\boldsymbol{s}\in\mathcal{S}$, where the players seek to optimize their cumulative (discounted) payoffs, in expectation. When $r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$ is bounded, continuous, and concave-convex in $(\boldsymbol{x},\boldsymbol{y})$, for all $\boldsymbol{s}\in\mathcal{S}$, these games are guaranteed to have a unique solution (Shapley, 1953),[1] while the **optimal policies**, i.e., the per-state collection of solutions to the min-max operator can be interpreted as a **recursive Nash equilibrium** of a zero-sum stochastic game, and can be computed in polynomial time by iterative application of the min-max operator (Shapley, 1953). Zero-sum *stochastic* games generalize zero-sum games from a single state to multiple states, and have found even more applications in a variety of fields (Jaśkiewicz & Nowak, 2018).

[1] Although Shapley's original results concern payoffs which are bilinear in the outer and inner players' actions, they extend directly to payoffs which are convex-concave in the players' actions.

Recently, Goktas & Greenwald (2021) studied the computation of a **generalized min-max operator** i.e., solving a **constrained min-max game with dependent strategy sets**: i.e., $\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:h(\boldsymbol{x},\boldsymbol{y})\geq\mathbf{0}} r(\boldsymbol{x},\boldsymbol{y})$ where, in addition to the aforementioned assumptions, $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is continuous. Goktas & Greenwald have shown that even more problems of interest can be captured by solutions to the generalized min-max operator. However, a minimax theorem is not guaranteed to hold in this setting; as a result, these problems are best interpreted as zero-sum sequential, i.e., Stackelberg (Von Stackelberg, 1934), games, and their solutions, as Stackelberg equilibria. One can likewise consider **zero-sum stochastic games with dependent feasible strategy sets** $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^m$, nonempty and compact and a state-dependent payoff function $r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$, as well as a state dependent constraint function $\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$, for all $\boldsymbol{s} \in \mathcal{S}$, where the players seek to optimize their cumulative (discounted) payoffs, in expectation, while satisfying the constraint $\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) \geq \mathbf{0}$ at each state $\boldsymbol{s} \in \mathcal{S}$. Such a problem is best interpreted as a zero-sum stochastic Stackelberg game, for which the appropriate solution concept is the **recursive Stackelberg equilibrium (recSE)**. Although very little is known about the properties of such problems, they generalize both min-max games with dependent strategy sets and zero-sum stochastic games (with independent strategy sets), and, as we show, have novel applications.

In this paper, we prove the existence of recSE in zero-sum stochastic Stackelberg games, provide necessary and sufficient conditions for a solution to be a recSE, and show that a recSE can be computed in (weakly) polynomial time via value iteration. We further show that stochastic Stackelberg games can be used to solve problems of pricing and allocating goods across agents and time. In particular, we introduce **stochastic Fisher markets**, a stochastic generalization of the Fisher market (Brainard et al., 2000), and a special case of Friesen's (1979) financial market model, which itself is a stochastic generalization of the Arrow & Debreu model of a competitive economy (1954). We then prove the existence of recursive competitive equilibrium (Mehra & Prescott, 1977) in this model, under the assumption that consumers have continuous and homogeneous utility functions, by characterizing the competitive equilibria of any stochastic Fisher market as the Stackelberg equilibria of the corresponding stochastic Stackelberg game. Finally, we use value iteration to solve various stochastic Fisher markets, highlighting the issues that value iteration might face as a consequence of the smoothness properties of the utility functions.

Related Work Algorithms for min-max optimization problems (i.e., zero-sum games) with independent strategy sets have been extensively studied; for a summary see Goktas & Greenwald (2021), Section G. Goktas & Greenwald (2021) and (2022) studied zero-sum games with dependent strategy sets, proposing polynomial-time nested gradient descent ascent (GDA) and simultaneous GDA algorithms for such problems.

The computation of Stackelberg equilibrium in dynamic Stackelberg games has been studied in several interesting settings, but always with independent strategy sets. Bensoussan et al. (2015) study continuous-time stochastic Stackelberg games with continuous action spaces, and prove existence of a solution in their setting. Vasal (2020) and Vorobeychik & Singh (2012) study discrete-time stochastic Stackelberg games with discrete action and state spaces, and provide algorithms to solve such games. DeMiguel & Xu (2009) consider a stochastic Stackelberg game-like market model with n leaders and m followers; they prove the existence of a Stackelberg equilibrium in their model, and provide (without theoretical guarantees) algorithms that converge to such an equilibrium in experiments. Dynamic Stackelberg games (Li & Sethi, 2017) have been applied to a wide range of problems, including security (Vasal, 2020; Vorobeychik & Singh, 2012), insurance provision (Chen & Shen, 2018; Yuan et al., 2021), advertising (He et al., 2008), robust agent design (Rismiller et al., 2020), allocating goods across time intertemporal pricing (Oksendal et al., 2013).

The study of algorithms that compute competitive equilibria in Fisher markets was initiated by Devanur et al. (2002), who provided a polynomial-time method for solving these markets assuming linear utilities. More recently, there have been efforts to study markets in dynamic settings (Cheung et al., 2019; Gao et al., 2021; Goktas & Greenwald, 2021), in which the goal is to either track the changing equilibrium of a changing market, or minimize some regret-like quantity for the market. The models considered in these earlier works differ from ours as they do not have stochastic structure and do not invoke a dynamic solution concept.

1 Preliminaries

Notation We use caligraphic uppercase letters to denote sets (e.g., $\mathcal{X}$), bold uppercase letters to denote matrices (e.g., $\boldsymbol{X}$), bold lowercase letters to denote vectors (e.g., $\boldsymbol{p}$), bold uppercase letters to denote vector-valued random variables (e.g., $\boldsymbol{\Gamma}$), lowercase letters to denote scalar quantities, (e.g., x), and uppercase letters to denote scalar-valued random variables (e.g., X). We denote the ith row vector of a matrix (e.g., $\boldsymbol{X}$) by the corresponding bold lowercase letter with subscript i (e.g., $\boldsymbol{x}_i$). Similarly, we denote the jth entry of a vector (e.g., $\boldsymbol{p}$ or $\boldsymbol{x}_i$) by the corresponding Roman lowercase letter with subscript j (e.g., p_j or x_{ij}). We denote functions by a letter: e.g., f if the function is scalar valued, and $\boldsymbol{f}$ if the function is vector valued. We denote the vector of ones of size n by $\mathbf{1}_n$. We denote the set of integers $\{1, \dots, n\}$ by $[n]$, the set of natural numbers by $\mathbb{N}$, the set of real numbers by $\mathbb{R}$. We denote the postive and strictly positive elements of a set by a $+$ and $++$ subscript respectively, e.g., $\mathbb{R}_+$ and $\mathbb{R}_{++}$. We denote the orthogonal projection operator onto a set C by Π_C, i.e., $\Pi_C(\boldsymbol{x}) = \arg\min_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|^2$. We denote by $\Delta_n = \{\boldsymbol{x} \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i = 1\}$, and by $\Delta(A)$ the set of probability measures on the set A.

A **stochastic Stackelberg game** $(\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r_{\boldsymbol{x}}, r_{\boldsymbol{y}}, \boldsymbol{g}, p, \gamma)$ is a dynamic two-player sequential, i.e., Stackelberg, game where one player we call the outer-player (resp. inner-player) moves through a set of states $\mathcal{S}$ picking an action to play from their continuous set of actions $\mathcal{X} \subset \mathbb{R}^n$ (resp. $\mathcal{Y} \subset \mathbb{R}^m$). Players start the game at an initial state determined by an initial state distribution $\mu^{(0)} : \mathcal{S} \to [0, 1]$ s.t. for all states $\boldsymbol{s} \in \mathcal{S}$, $\mu^{(0)}(\boldsymbol{s}) \geq 0$ denotes the probability of the game being initialized at state $\boldsymbol{s}$. At each state $\boldsymbol{s} \in \mathcal{S}$ the action $\boldsymbol{x} \in \mathcal{X}$ chosen by the outer player determines the set of **feasible** actions $\{\boldsymbol{y} \in \mathcal{Y} \mid \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \mathbf{0}\}$ that the inner player can in turn play. Once the outer and inner players make their moves, they receive payoffs $r_{\boldsymbol{x}} : \boldsymbol{s} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and $r_{\boldsymbol{y}} : \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, respectively, and the game either ends with probability $1 - \gamma$, where $\gamma \in (0, 1)$ is called the **discount factor**, or transitions to a new state $\boldsymbol{s}' \in \mathcal{S}$, according to a **transition** probability function $p : \mathcal{S} \times \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \to [0, 1]$ s.t. $p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \in [0, 1]$ denotes the probability of transitioning to state $\boldsymbol{s}' \in \mathcal{S}$ from state $\boldsymbol{s} \in \mathcal{S}$ when a strategy profile $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y}$ is chosen by the players.

In this paper, we focus on stochastic Stackelberg **zero-sum** games $\mathcal{G}^{(0)} = (\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r, \boldsymbol{g}, p, \gamma)$, in which the outer player's loss is the inner player's gain, i.e., $r_{\boldsymbol{x}} = -r_{\boldsymbol{y}}$. We say that a zero-sum stochastic Stackelberg game is **convex-concave** if, for all $\boldsymbol{s} \in \mathcal{S}$, $r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \dots, g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ are convex-concave in $(\boldsymbol{x}, \boldsymbol{y})$. A zero-sum stochastic Stackelberg game reduces to zero-sum stochastic game (Shapley, 1953) if for all state-action tuples $(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$, $\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq 0$. As we will show, it suffices to focus on deterministic policies in which a **policy** for the outer (resp. inner) player is a mapping from states to actions $\boldsymbol{\pi}_{\boldsymbol{x}} : \mathcal{S} \to \mathcal{X}$ (resp. $\boldsymbol{\pi}_{\boldsymbol{y}} : \mathcal{S} \to \mathcal{Y}$). The **outcome** of a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$ is a policy profile $(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$ consisting of policies for the outer and inner players, respectively. An outcome $(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$ is said to be **feasible** if, for all states $\boldsymbol{s} \in \mathcal{S}$, $\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}(\boldsymbol{s})) \geq 0$. For simplicity, we introduce a function $\boldsymbol{G} : \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}} \to \mathbb{R}^{|\mathcal{S}| \times d}$ such that $\boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) = (\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}(\boldsymbol{s})))_{\boldsymbol{s} \in \mathcal{S}}$, and define feasible outcomes as those $(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$ s.t. $\boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) \geq \mathbf{0}$. For the rest of this paper, we assume:

Assumption 1.1. *1. For all states $\boldsymbol{s} \in \mathcal{S}$, the functions $r(\boldsymbol{s}, \cdot, \cdot), g_1(\boldsymbol{s}, \cdot, \cdot), \dots, g_d(\boldsymbol{s}, \cdot, \cdot)$ are continuous in $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y}$ with payoffs r bounded, i.e., $\|r\|_\infty \leq \alpha < \infty$, for some $\alpha \in \mathbb{R}_+$, and 2 $\mathcal{X}, \mathcal{Y}$ are non-empty and compact.*

Given a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$, the **state-value function**, $v : \mathcal{S} \times \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}} \to \mathbb{R}$, and the **action-value function**, $q : \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}} \to \mathbb{R}$ are respectively defined as: $v(\boldsymbol{s}; \boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) = \mathbb{E}^{\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}}{}_{\boldsymbol{S}^{(t+1)} \sim p(\cdot \mid \boldsymbol{S}^{(t)}, \boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)})} \left[\sum_{t=0}^{\infty} (1 - \gamma) \gamma^t r(\boldsymbol{S}^{(t)}, \boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)}) \mid \boldsymbol{S}^{(0)} = \boldsymbol{s}\right]$; $q(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) = \mathbb{E}^{\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}}{}_{\boldsymbol{S}^{(t+1)} \sim p(\cdot \mid \boldsymbol{S}^{(t)}, \boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)})} \left[\sum_{t=0}^{\infty} (1 - \gamma) \gamma^t r(\boldsymbol{S}^{(t)}, \boldsymbol{X}^{(t)}, \boldsymbol{Y}^{(t)}) \mid \boldsymbol{S}^{(0)} = \boldsymbol{s}\right.$ $\left.\boldsymbol{X}^{(0)} = \boldsymbol{x}, \boldsymbol{Y}^{(0)} = \boldsymbol{y}\right]$.

For clarity, we write expectations conditional on $\boldsymbol{X}^{(t)} = \boldsymbol{\pi}_{\boldsymbol{x}}(\boldsymbol{S}^{(t)})$ and $\boldsymbol{Y}^{(t)} = \boldsymbol{\pi}_{\boldsymbol{y}}(\boldsymbol{S}^{(t)})$ as $\mathbb{E}^{\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}}$, and denote the state- and action-value functions by $v^{\boldsymbol{\pi}_{\boldsymbol{x}} \boldsymbol{\pi}_{\boldsymbol{y}}}(\boldsymbol{s})$, and $q^{\boldsymbol{\pi}_{\boldsymbol{x}} \boldsymbol{\pi}_{\boldsymbol{y}}}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$, respectively. Additionally, we let $\mathcal{V} = [-\alpha, \alpha]^{\mathcal{S}}$ be the space of all state-value functions of the form $v : \mathcal{S} \to [-\alpha, \alpha]$, and we let $\mathcal{Q} = [-\alpha, \alpha]^{\mathcal{S} \times \mathcal{X} \times \mathcal{Y}}$ be the space of all action-value functions of the form $q : \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \to [-\alpha, \alpha]$. Note that by Assumption 1.1 the range of

the state- and action-value functions is $[-\alpha, \alpha]$. The cumulative payoff function of the game $u : \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}} \to \mathbb{R}$ is the total expected loss (resp. gain) of the outer (resp. inner) player is then given by $u(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) = \mathbb{E}_{\boldsymbol{s}\sim\mu^{(0)}(\boldsymbol{s})}\left[v^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}(\boldsymbol{s})\right]$.

Definition 1.2. *A feasible policy profile $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$ is said to be a **recursive Stackelberg equilibrium (recSE)** of a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$ iff $\max_{\boldsymbol{\pi_y}\in\mathcal{Y}^{\mathcal{S}}:\boldsymbol{G}(\boldsymbol{\pi}^*_{\boldsymbol{x}},\boldsymbol{\pi_y})\geq\mathbf{0}} u\left(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi_y}\right) \leq u\left(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}}\right) \leq \min_{\boldsymbol{\pi_x}\in\mathcal{X}^{\mathcal{S}}} \max_{\boldsymbol{\pi_y}\in\mathcal{Y}^{\mathcal{S}}:\boldsymbol{G}(\boldsymbol{\pi_x},\boldsymbol{\pi_y})\geq\mathbf{0}} u\left(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}\right)$.*

Remark 1.3. *This is a strong definition of a recSE, since we require the constraints $\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \mathbf{0}$ to be satisfied at all states $\boldsymbol{s} \in \mathcal{S}$, not only states which are reached with strictly positive probability.*

Mathematical Preliminaries A probability measure $q_1 \in \Delta(\mathcal{S})$ **convex stochastically dominates (CSD)** $q_2 \in \Delta(\mathcal{S})$ if $\int_{\mathcal{S}} v(s)q_1(s)ds \geq \int_{\mathcal{S}} v(s)q_2(s)ds$ for all continuous, bounded, and convex functions v on S. A transition function p is termed **CSD convex** in $\boldsymbol{x}$ if for all $\lambda \in (0,1)$, $\boldsymbol{y} \in \mathcal{Y}$ and any $(\boldsymbol{s}', \boldsymbol{x}'), (\boldsymbol{s}^\dagger, \boldsymbol{x}^\dagger) \in \mathcal{S} \times \mathcal{X}$, with $(\boldsymbol{s}, \boldsymbol{x}) = \lambda(\boldsymbol{s}', \boldsymbol{x}') + (1-\lambda)(\boldsymbol{s}^\dagger, \boldsymbol{x}^\dagger)$, we have $\lambda p(\cdot \mid \boldsymbol{s}', \boldsymbol{x}', \boldsymbol{y}) + (1-\lambda)p(\cdot \mid \boldsymbol{s}^\dagger, \boldsymbol{x}^\dagger, \boldsymbol{y})$ CSD $p(\cdot \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$. A transition function p is termed **CSD concave** in $\boldsymbol{y}$ if for all $\lambda \in (0,1)$ and any $(\boldsymbol{s}', \boldsymbol{y}'), (\boldsymbol{s}^\dagger, \boldsymbol{y}^\dagger) \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$, with $(\boldsymbol{s}, \boldsymbol{y}) = \lambda(\boldsymbol{s}', \boldsymbol{y}') + (1-\lambda)(\boldsymbol{s}^\dagger, \boldsymbol{y}^\dagger)$, we have $p(\cdot \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ CSD $\lambda p(\cdot \mid \boldsymbol{s}', \boldsymbol{x}, \boldsymbol{y}') + (1-\lambda)p(\cdot \mid \boldsymbol{s}^\dagger, \boldsymbol{x}, \boldsymbol{y}^\dagger)$. A mapping $L : \mathcal{A} \to \mathcal{B}$ is said to be a **contraction mapping** (resp. **non-expansion**) w.r.t. norm $\|\cdot\|$ iff for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{A}$, and for $k \in [0,1)$ (resp. $k = 1$) such that $\|L(\boldsymbol{x}) - L(\boldsymbol{y})\| \leq k\|\boldsymbol{x} - \boldsymbol{y}\|$.

2 Properties of Recursive Stackelberg equilibrium

We first state a necessary condition that the state-value function induced by a policy profile needs to satisfy in order for that policy profile to be a recSE. It is not directly clear if there exists a policy for any zero-sum stochastic Stackelberg game which satisfies this necessary condition since there is no way to guarantee that such an inequality will hold at all states.[2]

Lemma 2.1. *Consider a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$. A policy profile $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}}) \in \mathcal{S}^{\mathcal{X}} \times \mathcal{S}^{\mathcal{Y}}$ is a recSE if, for all $\boldsymbol{s} \in \mathcal{S}$, we have that: $\max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}),\boldsymbol{y})\geq\mathbf{0}} q^{\boldsymbol{\pi}^*_{\boldsymbol{x}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{y}) \leq q^{\boldsymbol{\pi}^*_{\boldsymbol{x}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{x}), \boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{y})) \leq \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\mathbf{0}} q^{\boldsymbol{\pi}^*_{\boldsymbol{x}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$. Equivalently, a policy profile $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}})$ is a recSE if $(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}))$ is a Stackelberg equilibrium: i.e., a solution to $\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\mathbf{0}} q^{\boldsymbol{\pi}^*_{\boldsymbol{x}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ at each $\boldsymbol{s} \in \mathcal{S}$.*

We define an operator $C : \mathcal{V} \to \mathcal{V}$ associated with a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$ whose fixed points satisfy the condition given in Lemma 2.1, and hence correspond to the value function associated with a recSE of $\mathcal{G}^{(0)}$. We then show that this operator is a contraction mapping, thereby establishing the existence of such a fixed point. This result generalizes a result first shown by Shapley (1953) for zero-sum stochastic games, i.e., zero-sum stochastic Stackelberg games in which $\boldsymbol{G}(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \mathbf{0}$, for all $(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$. Define $C : \mathcal{V} \to \mathcal{V}$ for a stochastic Stackelberg game $\mathcal{G}^{(0)}$ as the operator $(Cv)(\boldsymbol{s}) = \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\mathbf{0}} \mathbb{E}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}\left[r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma v(\boldsymbol{S}')\right]$.

Theorem 2.2. *$(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}})$ is a recSE of $\mathcal{G}^{(0)}$ of $v^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}$ iff it induces a value function which is a fixed point of C: i.e., $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}})$ is a Stackelberg equilbrium iff, for all $\boldsymbol{s} \in \mathcal{S}$, $\left(Cv^{\boldsymbol{\pi}^*_{\boldsymbol{x}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}\right)(\boldsymbol{s}) = v^{\boldsymbol{\pi}^*_{\boldsymbol{x}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s})$.*

The following technical lemma is crucial to proving that C is a contraction mapping. It tells us that the generalized min-max operator is non-expansive; in other words, the generalized min-max operator is 1-Lipschitz w.r.t. the sup-norm.

Lemma 2.3. *Suppose that $f, h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, $\boldsymbol{g} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$ are continuous functions, and $\mathcal{X}, \mathcal{Y}$ are compact sets, we then have: $\left|\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq\mathbf{0}} f(\boldsymbol{x}, \boldsymbol{y}) - \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq\mathbf{0}} h(\boldsymbol{x}, \boldsymbol{y})\right| \leq \max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}\times\mathcal{Y}} |f(\boldsymbol{x}, \boldsymbol{y}) - h(\boldsymbol{x}, \boldsymbol{y})|$.*

With the above lemma in hand, we can now prove that C is a contraction mapping.

[2] All omitted results and proofs can be found in the appendix.

Theorem 2.4. *Consider the operator C associated with a stochastic Stackelberg game $\mathcal{G}^{(0)}$. If Assumption 1.1 holds, then C is a contraction mapping w.r.t. to the sup norm $\|.\|_\infty$ with constant γ.*

Given some initial state-value function $v^{(0)} \in \mathcal{V}$, we define the **value iteration** process as $v^{(t+1)} = Cv^{(t)}$, for all $t \in \mathbb{N}_+$ (Algorithm 1). One way to interpret $v^{(t)}$ is as the function that returns the value $v^{(t)}(\boldsymbol{s})$ of each state $\boldsymbol{s} \in \mathcal{S}$ in the t stage zero-sum Stackelberg game starting at the last stage t and continuing until stage 0, with terminal payoffs given by $v^{(0)}$. The following theorem, which is a consequence of Theorems 2.2 and 2.4, not only states the existence of a recSE but also provides us with a means of computing a recSE via value iteration.

Theorem 2.5. *Consider a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$. If Assumption 1.1 holds, then $\mathcal{G}^{(0)}$ has at least one recSE $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ with unique value function $v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}$. Further, $v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}$ can be computed by iteratively applying C to any initial state-value function $v^{(0)} \in \mathcal{V}$: $\lim_{t\to\infty} v^{(t)} = v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}$.*

Remark 2.6. *Unlike Shapley's existence theorem for recursive Nash equilibria in zero-sum stochastic games, the above theorem does not require that the payoff function be convex-concave. The only conditions needed are continuity of the payoffs and constraints, and bounded payoffs. This makes the recSE a potentially useful solution concept, even for non-convex-non-concave stochastic games.*

Note that the proof of Theorem 2.5 shows that there exists a policy profile which satisfies the conditions of Lemma 2.1. Since this recSE definition is independent of the initial state distribution, we can infer that the recSE of any zero-sum Stackelberg game $\mathcal{G}^{(0)} = (\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r, \boldsymbol{g}, p, \gamma)$ is independent of the initial state distribution $\mu^{(0)}$. Hence, from now on, we denote a zero-sum Stackelberg game by $\mathcal{G}$.

It seems that a stronger equilibrium concept for zero-sum stochastic Stackelberg games would be one that requires the strategy profile be a recSE of each subgame, aptly called **subgame perfect Stackelberg equilibrium**. A feasible policy profile $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*) \in \mathcal{S}^{\mathcal{X}} \times \mathcal{S}^{\mathcal{Y}}$ is a subgame perfect Stackelberg equilibrium of a zero-sum stochastic Stackelberg game $\mathcal{G}$ if, for all $\boldsymbol{s} \in \mathcal{S}, t \in \mathbb{N}, \mu^{(t+1)}(\boldsymbol{s}) = \sum_{\boldsymbol{s}' \in \mathcal{S}} p(\boldsymbol{s} \mid \boldsymbol{s}', \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}'), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{s}'))\mu^{(t)}(\boldsymbol{s}')$, $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ is a recSE of $(\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(t)}, r, \boldsymbol{g}, p, \gamma)$. By definition, the set of subgame perfect Stackelberg equilibria of a zero-sum stochastic Stackelberg game is a subset of the set of recSE of the game. Since the set of recSE is independent of the initial state distribution $\mu^{(0)}$, the reverse inclusion also holds. We thus have the following corollary:

Corollary 2.7. *The set of recSE and subgame perfect Stackelberg equilibria coincide in zero-sum stochastic Stackelberg games.*

Theorem 2.5 tells us that value iteration converges to the value function associated with a recSE. Additionally, under Assumption 1.1, recSE is computable in (weakly) polynomial time.[3]

Theorem 2.8. *[Convergence of Value Iteration] Suppose value iteration is run on input $\mathcal{G}$. If Assumption 1.1 holds, and if we initialize $v^{(0)}(\boldsymbol{s}) = 0$, for all $\boldsymbol{s} \in \mathcal{S}$, then for $k \geq \frac{1}{1-\gamma} \log \frac{2\alpha}{\epsilon(1-\gamma)}$, we have $v^{(k)}(\boldsymbol{s}) - v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}) \leq \epsilon$.*

3 Subdifferential Envelope Theorems and Optimality Conditions for Recursive Stackelberg Equilibrium

In this section, we derive optimality conditions for recursive Stackelberg equilibria. In particular, we provide necessary conditions for a policy profile to be a recSE of any zero-sum stochastic Stackelberg game, and we provide sufficient conditions for a policy profile to be a recursive Stackelberg equilibrium of a zero-sum convex-concave stochastic Stackelberg game. Using these results, we prove in the next section that recursive market equilibrium (Mehra & Prescott, 1977) is an instance of recSE in a large class of stochastic markets.

[3]This convergence is only weakly polynomial time, since the computation of the generalized min-max operator applied to an arbitrary continuous function is an NP-hard problem; it is at least as hard as non-convex optimization. If, however, one restricts themselves to convex-concave stochastic Stackelberg games, Stackelberg equilibrium is computable in polynomial time, making the problem much more tractable.

The Benveniste-Scheinkman theorem characterizes the derivative of the optimal value function associated with a recursive optimization problem w.r.t. its parameters, when it is differentiable (Benveniste & Scheinkman, 1979). Our proofs of the necessary and sufficient optimality conditions rely on a novel subdifferential generalization (Theorem C.3, Appendix C) of this theorem, which applies even when the optimal value function is not differentiable. A consequence of our subdifferential version of the Benveniste-Scheinkman theorem is that we can easily derive the first-order necessary conditions for a policy profile to be a recSE of any stochastic Stackelberg game $\mathcal{G}$ satisfying Assumption 1.1, under standard regularity conditions.

Theorem 3.1. *Consider a zero-sum stochastic Stackelberg game $\mathcal{G}$, where $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid q_1(\boldsymbol{x}) \leq 0, \dots, q_p(\boldsymbol{x}) \leq 0\}$ and $\mathcal{Y} = \{\boldsymbol{y} \in \mathbb{R}^m \mid r_1(\boldsymbol{y}) \geq 0, \dots, r_l(\boldsymbol{y}) \geq 0\}$. Let $\mathcal{L}_{\boldsymbol{s},\boldsymbol{x}}(\boldsymbol{y}, \boldsymbol{\lambda}) = r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma \mathbb{E}_{\boldsymbol{S}' \sim p(\cdot \mid \boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}[v(\boldsymbol{S}', \boldsymbol{x})] + \sum_{k=1}^{d} \lambda_k g_k(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$. Suppose that Assumption 1.1 holds, and that 1. for all $\boldsymbol{s} \in \mathcal{S}, \boldsymbol{y} \in \mathcal{Y}$, $r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$, $g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \dots, g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ are concave in $\boldsymbol{x}$, 2. $\nabla_{\boldsymbol{x}} r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \nabla_{\boldsymbol{x}} g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \dots, \nabla_{\boldsymbol{x}} g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$, $\nabla_{\boldsymbol{y}} r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \nabla_{\boldsymbol{y}} g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \dots, \nabla_{\boldsymbol{y}} g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ exist, for all $\boldsymbol{s} \in \mathcal{S}, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \in \mathcal{Y}$, 4. $p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ is continuous CSD convex and differentiable in $(\boldsymbol{x}, \boldsymbol{y})$, and 5. Slater's condition holds, i.e., $\forall \boldsymbol{s} \in \mathcal{S}, \boldsymbol{x} \in \mathcal{X}, \exists \widehat{\boldsymbol{y}} \in \mathcal{Y}$ s.t. $g_k(\boldsymbol{s}, \boldsymbol{x}, \widehat{\boldsymbol{y}}) > 0$, for all $k = 1, \dots, d$ and $r_j(\widehat{\boldsymbol{y}}) > 0$, for all $j = 1, \dots, l$, and $\exists \boldsymbol{x} \in \mathbb{R}^n$ s.t. $q_k(\boldsymbol{x}) < 0$ for all $k = 1 \dots, p$. Then, there exists $\boldsymbol{\mu}^* : \mathcal{S} \to \mathbb{R}^p_+$, $\boldsymbol{\lambda}^* : \mathcal{S} \times \mathcal{X} \to \mathbb{R}^d_+$, and $\boldsymbol{\nu}^* : \mathcal{S} \times \mathcal{X} \to \mathbb{R}^l_+$ s.t. a policy profile $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$ is a recSE of $\mathcal{G}$ only if it satisfies the following conditions, for all $\boldsymbol{s} \in \mathcal{S}$:*

$$\nabla_{\boldsymbol{x}} \mathcal{L}_{\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})}(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}), \boldsymbol{\lambda}^*(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})))) + \sum_{k=1}^{p} \mu^*_k(\boldsymbol{s}) \nabla_{\boldsymbol{x}} q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) = 0 \tag{1}$$

$$\nabla_{\boldsymbol{y}} \mathcal{L}_{\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})}(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}), \boldsymbol{\lambda}^*(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})))) + \sum_{k=1}^{l} \nu^*_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \nabla_{\boldsymbol{x}} r_k(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \tag{2}$$

$$\mu^*_k(\boldsymbol{s}) q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) = 0 \qquad q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \leq 0 \quad \forall k \in [p] \tag{3}$$

$$g_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) \geq 0 \qquad \lambda^*_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) g_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \quad \forall k \in [d] \tag{4}$$

$$\nu^*_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \nabla_{\boldsymbol{x}} r_k(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \qquad r_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \geq 0 \quad \forall k \in [l] \tag{5}$$

Note, that in Theorem 3.1, when one additionally assumes concavity of $\max_{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) \geq \boldsymbol{0}} \left\{ r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma \mathbb{E}_{\boldsymbol{S}' \sim p(\cdot \mid \boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}[v(\boldsymbol{S}')] \right\}$ in $\boldsymbol{x}$, Equations (74) to (78) become necessary *and sufficient* optimality conditions. For completeness, the reader can find the necessary and sufficient optimality conditions for convex-concave stochastic Stackelberg games under standard regularity conditions in Theorem C.4 (Appendix C). The proof follows exactly as that of Theorem 2.2.

4 Recursive Market Equilibrium

We now introduce an application of zero-sum stochastic Stackelberg games, which generalizes a well known market model, the Fisher market (Brainard et al., 2000), to a dynamic setting in which buyers not only participate in markets across time, but their wealth persists. A **(static) Fisher market** consists of n buyers and m divisible goods (Brainard et al., 2000). Each buyer $i \in [n]$ is endowed with a budget $b_i \in \mathcal{B}_i \subset \mathbb{R}_+$ and a utility function $u_i : \mathbb{R}^m_+ \times \mathcal{T}_i \to \mathbb{R}$, which is parameterized by a type $t_i \in \mathcal{T}_i$ that defines a preference relation over the consumption space $\mathbb{R}^m_+$. Each good is characterized by a supply $q_j \in \mathcal{Q}_j \subset \mathbb{R}_+$. A **stochastic Fisher market** is a dynamic market in which each state corresponds to a static Fisher market: i.e., each state $\boldsymbol{s} \in \mathcal{S}$ is characterized by a tuple $(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q})$. In each state, the buyers choose their allocations $\boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)^T \in \mathbb{R}^{n \times m}_+$ and the market determines prices, after which the market comes to an end with probability $(1 - \gamma)$, or it moves into a new state with probability $p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{X})$.[4] A **stochastic Fisher market with savings** is a stochastic Fisher market in which, at each state, each buyer i, in addition to choosing their allocation, can set aside some **savings** $\beta_i \in \mathbb{R}_+$ to spend at some future state, and transitions depend on their savings choices as well: i.e., $p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{X}, \boldsymbol{\beta})$.

[4]Note that, as is standard in the literature, we assume that prices do not determine the next state since market prices are set by a "fictional auctioneer." There is no market participant who determined prices!

A stochastic Fisher market with savings is denoted by $(\mathcal{S}, \mathcal{U}, \boldsymbol{b}^{(0)}, p, \gamma)$. Given such a market, a **recursive competitive equilibrium (recCE)** (Mehra & Prescott, 1977) is a tuple $(\boldsymbol{X}^*, \boldsymbol{\beta}^*, \boldsymbol{p}^*) \in \mathbb{R}_+^{n\times m\times\mathcal{S}} \times \mathbb{R}_+^{n\times\mathcal{S}} \times \mathbb{R}_+^{m\times\mathcal{S}}$, which consists of an allocation, savings, and price system s.t. 1) the buyers are expected utility maximizing, constrained by their savings and spending constraints, i.e., for all buyers $i \in [n]$, $(\boldsymbol{x}_i^*, \beta_i^*)$ is the optimal policy that, for all states $(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) \in \mathcal{S}$, solves the Bellman equation $\nu_i(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) =$

$$\max_{(\boldsymbol{x}_i,\beta_i)\in\mathbb{R}_+^{m+1}:\boldsymbol{x}_i\cdot\boldsymbol{p}^*(\boldsymbol{t},\boldsymbol{b},\boldsymbol{q})+\beta_i\le b_i} \left\{ u_i\left(\boldsymbol{x}_i, t_i\right) + \gamma \mathop{\mathbb{E}}_{(\boldsymbol{t}',\boldsymbol{b}',\boldsymbol{q}')\sim p(\cdot|\boldsymbol{t},\boldsymbol{b},\boldsymbol{q},(\boldsymbol{x}_i,\boldsymbol{X}^*_{-i}(\boldsymbol{s})),(\beta_i,\boldsymbol{\beta}^*_{-i}(\boldsymbol{s})))} \left[\nu_i(\boldsymbol{t}', \boldsymbol{b}' + \beta_i, \boldsymbol{q}')\right] \right\},$$

where $\boldsymbol{X}^*_{-i}$, $\boldsymbol{\beta}^*_{-i}$ denote the allocation and saving systems excluding buyer i, and 2) the market clears in each state so that unallocated goods in each state are priced at 0, i.e., for all $j \in [m]$ and $\boldsymbol{s} \in \mathcal{S}$, $p_j^*(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) > 0 \implies \sum_{i\in[n]} x_{ij}^*(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) = 1$ and $p_j^*(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) \ge 0 \implies \sum_{i\in[n]} x_{ij}^*(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) \le 1$.

The following theorem shows that the recSE of a stochastic Fisher market with savings are in fact recursive competitive equilibria. We note that if one ignores the savings terms, then the recSE are recursive competitive equilibria of the same market without savings.

Theorem 4.1. *A stochastic Fisher market with savings $(\mathcal{S}, \mathcal{U}, \boldsymbol{b}^{(0)}, p, \gamma)$ in which $\mathcal{U}$ is a vector of continuous and homogeneous utility functions has at least one recCE. Additionally, the recSE $(\boldsymbol{p}^*, \boldsymbol{X}^*, \boldsymbol{\beta}^*)$ that solves the following Bellman equation corresponds to the recCE of $(\mathcal{S}, \mathcal{U}, \boldsymbol{b}^{(0)}, p, \gamma)$:*

$$v(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) = \min_{\boldsymbol{p}\in\mathbb{R}_+^m} \max_{(\boldsymbol{X},\boldsymbol{\beta})\in\mathbb{R}_+^{n\times(m+1)}:\boldsymbol{X}\boldsymbol{p}+\boldsymbol{\beta}\le\boldsymbol{b}} \sum_{j\in[m]} q_j p_j + \sum_{i\in[n]} (b_i - \beta_i)\log(u_i(\boldsymbol{x}_i, t_i)) + \gamma \mathop{\mathbb{E}}_{(\boldsymbol{t}',\boldsymbol{b}',\boldsymbol{q}')\sim p(\cdot|\boldsymbol{t},\boldsymbol{b},\boldsymbol{q},\boldsymbol{X},\boldsymbol{\beta})} \left[v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}, \boldsymbol{q}')\right] \quad (6)$$

Remark 4.2. *This result could not be obtained by modifying the Lagrangian formulation, i.e., the simultaneous-move game form, of the Eisenberg-Gale program, because the saving problem is a convex-non-concave problem, and existence of recursive Nash equilibrium in deterministic policies requires convex-concavity of payoffs (Jaśkiewicz & Nowak, 2018).*

5 Experiments

In order to better understand the iteration complexity of value iteration, and to understand how it performs in a continuous state space, we computed the recursive Stackelberg equilibria of various stochastic Fisher market with savings. We created markets with three different classes of utility functions, each of which endowed the state-value function with different smoothness properties.[5] Let $\boldsymbol{\theta}_i \in \mathbb{R}^m$ be a vector of parameters that describes the utility function of buyer $i \in [n]$. We considered the following (standard) utility function classes: 1. linear: $u_i(\boldsymbol{x}_i) = \sum_{j\in[m]} \theta_{ij} x_{ij}$; 2. Cobb-Douglas: $u_i(\boldsymbol{x}_i) = \prod_{j\in[m]} x_{ij}^{\theta_{ij}}$; and 3. Leontief: $u_i(\boldsymbol{x}_i) = \min_{j\in[m]} \left\{\frac{x_{ij}}{\theta_{ij}}\right\}$.

Since the state space is continuous, the value function is also continuous in stochastic Fisher markets. As a result, we used fitted value iteration, finding a fit via linear regression (e.g., Boyan & Moore (1994)). To solve the generalized min-max operator at each step of value iteration, we used two methods: 1. **nested gradient descent ascent (GDA)** (Goktas & Greenwald (2021); Algorithm 2), which is not guaranteed to converge to a global optimum since the zero-sum Stackelberg game for stochastic Fisher markets is convex-non-concave; and 2. **max-oracle gradient descent** (Goktas & Greenwald (2021); Algorithm 3), where we used simulated annealing (Bertsimas & Tsitsiklis, 1993), a metaheuristic which aims to find a global optimum, as the max-oracle. Although simulated annealing is not guaranteed to converge to a global optimum, we observed that it outperformed nested GDA, more often finding a global maximum for the inner player.

To check whether the value function computed was optimal, we measured the exploitability of the market, meaning the distance between the recursive competitive equilibrium computed and the actual competitive equilibrium. To do so required that we check two conditions: 1) if each buyer's expected utility is maximized at the computed allocation and saving system at the price system output by the algorithm, and 2) if the market always clears. For both settings, given the value

[5] Our code can be found here, and details of our experimental setup can be found in Appendix E.

function computed by value iteration, we extracted the greedy policy and unrolled it across time to obtain the greedy actions $(\boldsymbol{X}^{(t)}, \boldsymbol{\beta}^{(t)}, \boldsymbol{p}^{(t)})$ at each state $\boldsymbol{s}^{(t)}$. We then computed the cumulative utility of the allocation and saving systems computed by the algorithms, i.e., for all $i \in [n]$, $\sum_{t=0}^{T} \gamma^t u_i(\boldsymbol{x}_i^{(t)})$, and compared this value to the expected maximum utility u_i^*, obtained by solving the consumption/saving problem for individual buyers given the price systems computed by our algorithms. We report the normalized distance between these two values: e.g., in the case of two buyers, we report $\frac{||(u_1,u_2)-(u_1^\star,u_2^\star)||}{||(u_1^*,u_2^*)||}$, Finally, we measured the excess demand, which we took as the distance to market clearance, i.e., $\frac{1}{T}\sum_{t=1}^{T}||\sum_{i\in[n]} \boldsymbol{x}_i^{(t)} - \boldsymbol{q}^{(t)}||$.

The exploitability of the recursive competitive equilibrium computed by both nested GDA and max-oracle gradient descent in all market types is shown in Figure 2, while Figure 1 depicts the average value of the value function across all states as it varies with time. We observe that value iteration converges for both nested GDA and max-oracle gradient descent in Cobb-Douglas markets. Max-oracle gradient descent achieves a higher value, because the oracle does a better job at finding a global maximum for the inner player. In linear markets, only the max-oracle gradient descent algorithm seems to converge. This is most likely due to the non-differentiability of the payoffs in linear Fisher markets, making it difficult for nested GDA to find a global optimum. In Leontief markets, neither algorithm seems to affect the value function very much, which might lead one to conclude that value iteration converges. In terms of exploitability, both algorithms seem to perform more or less as they do in the linear markets, which suggests that the initial value function is close to the optimal value function. Surprisingly, even though value iteration converges in both settings for Cobb-Douglas markets, the allocation and savings computed seem to be far from optimal. This is most likely due to the fact that we used linear regression to approximate the value function at each state, even though the value function associated with each individual buyer's consumption/saving problem, and likewise the market's value function, is strictly concave in Cobb-Douglas markets. We note that the buyers' value functions are linear in linear and Leontief markets, and hence compatible with our fitted linear regression, which might explain the greater exploitability of Cobb-Douglas markets even though value iteration seems to converge.

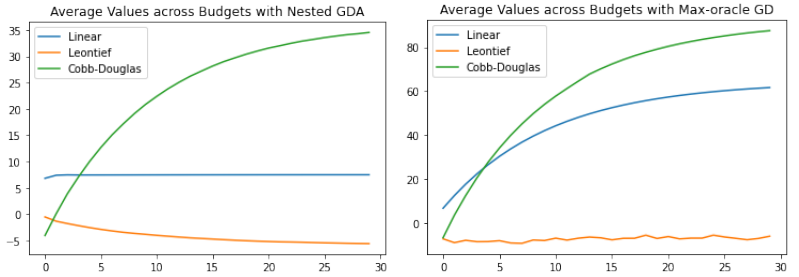

Figure 1: The average value of the value function under Nested GDA (left) and Max-oracle gradient descent with simulated annealing (right).

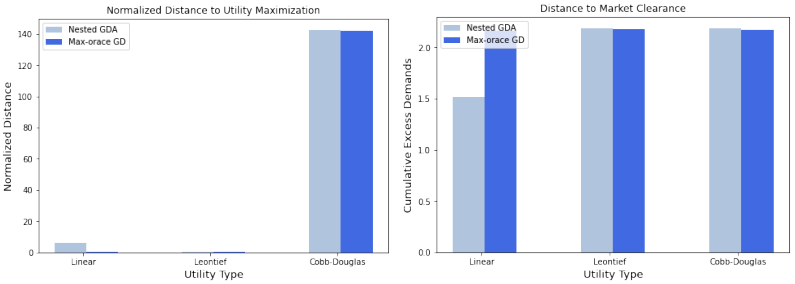

Figure 2: Distance to utility maximization (left) and market clearance (right) of the computed recCE.

6 Conclusion

In this paper, we proved the existence of recursive Stackelberg equilibria in zero-sum stochastic Stackelberg games, provided necessary and sufficient conditions for a policy profile to be a recursive Stackelberg equilibrium, and showed that a Stackelberg equilibrium can be computed

in (weakly) polynomial time via value iteration. Further, we provided a characterization of recursive competitive equilibria in stochastic Fisher markets with savings via zero-sum stochastic Stackelberg games. Finally, we used value iteration to solve for recursive competitive equilibria in stochastic Fisher markets. Future work seeking to improve our solution methods for stochastic Fisher markets should aim to replace the linear regression step in our fitted value iteration method with concave regression, since the value function is guaranteed to be concave in budgets for stochastic Fisher markets. Additionally, it is conceivable that policy-network-based deep reinforcement learning methods may be able to resolve the difficulties that our method has in solving for global optima in stochastic Fisher markets.

References

Mohammad Alkousa, Darina Dvinskikh, Fedor Stonyakin, Alexander Gasnikov, and Dmitry Kovalev. Accelerated methods for composite non-bilinear saddle point problem, 2020.

Kenneth Arrow and Gerard Debreu. Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, pp. 265–290, 1954.

Alp E. Atakan. Stochastic convexity in dynamic programming. *Economic Theory*, 22(2):447–455, 2003. ISSN 09382259, 14320479. URL http://www.jstor.org/stable/25055693.

Stefan Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fund. math*, 3(1):133–181, 1922.

Richard Bellman. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences of the United States of America*, 38(8):716, 1952.

Alain Bensoussan, Shaokuan Chen, and Suresh P Sethi. The maximum principle for global solutions of stochastic stackelberg differential games. *SIAM Journal on Control and Optimization*, 53(4): 1956–1981, 2015.

L. M. Benveniste and J. A. Scheinkman. On the differentiability of the value function in dynamic models of economics. *Econometrica*, 47(3):727–732, 1979. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1910417.

Dimitris Bertsimas and John Tsitsiklis. Simulated annealing. *Statistical science*, 8(1):10–15, 1993.

Justin Boyan and Andrew Moore. Generalization in reinforcement learning: Safely approximating the value function. *Advances in neural information processing systems*, 7, 1994.

William C Brainard, Herbert E Scarf, et al. *How to compute equilibrium prices in 1891*. Citeseer, 2000.

Lv Chen and Yang Shen. On a new paradigm of optimal reinsurance: a stochastic stackelberg differential game between an insurer and a reinsurer. *ASTIN Bulletin: The Journal of the IAA*, 48 (2):905–960, 2018.

Yun Kuen Cheung, Martin Hoefer, and Paresh Nakhe. Tracing equilibrium in dynamic markets via distributed adaptation. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 1225–1233, 2019.

Victor DeMiguel and Huifu Xu. A stochastic multiple-leader stackelberg model: analysis, computation, and application. *Operations Research*, 57(5):1220–1235, 2009.

N. R. Devanur, C. H. Papadimitriou, A. Saberi, and V. V. Vazirani. Market equilibrium via a primal-dual-type algorithm. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pp. 389–395, 2002. doi: 10.1109/SFCS.2002.1181963.

Harley Flanders. Differentiation under the integral sign. *The American Mathematical Monthly*, 80 (6):615–627, 1973. ISSN 00029890, 19300972. URL http://www.jstor.org/stable/2319163.

Peter H. Friesen. The arrow-debreu model extended to financial markets. *Econometrica*, 47(3): 689–707, 1979. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1910415.

Yuan Gao, Christian Kroer, and Alex Peysakhovich. Online market equilibrium with application to fair division. *Advances in Neural Information Processing Systems*, 34, 2021.

Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, and Simon Lacoste-Julien. A variational inequality perspective on generative adversarial networks, 2020.

Denizalp Goktas and Amy Greenwald. Convex-concave min-max stackelberg games. *Advances in Neural Information Processing Systems*, 34, 2021.

Denizalp Goktas and Amy Greenwald. Robust no-regret learning in min-max Stackelberg games, 2022.

Erfan Yazdandoost Hamedani and Necdet Serhat Aybat. A primal-dual algorithm for general convex-concave saddle point problems. *arXiv preprint arXiv:1803.01401*, 2, 2018.

Xiuli He, Ashutosh Prasad, and Suresh P Sethi. Cooperative advertising and pricing in a dynamic stochastic supply chain: Feedback stackelberg strategies. In *PICMET'08-2008 Portland International Conference on Management of Engineering & Technology*, pp. 1634–1649. IEEE, 2008.

Mingyi Hong, Junyu Zhang, , Shuzhong Zhang, et al. On lower iteration complexity bounds for the saddle point problems, 2020.

Adam Ibrahim, Waiss Azizian, Gauthier Gidel, and Ioannis Mitliagkas. Lower bounds and conditioning of differentiable games. *arXiv preprint arXiv:1906.07300*, 2019.

Anna Jaśkiewicz and Andrzej S Nowak. Zero-sum stochastic games. *Handbook of dynamic game theory*, pp. 1–64, 2018.

Chi Jin, Praneeth Netrapalli, and Michael I. Jordan. What is local optimality in nonconvex-nonconcave minimax optimization?, 2020.

Anatoli Juditsky, Arkadi Nemirovski, et al. First order methods for nonsmooth convex large-scale optimization, ii: utilizing problems structure. *Optimization for Machine Learning*, 30(9):149–183, 2011.

HW Kuhn and AW Tucker. Proceedings of 2nd berkeley symposium, 1951.

Tao Li and Suresh P Sethi. A review of dynamic stackelberg game models. *Discrete & Continuous Dynamical Systems-B*, 22(1):125, 2017.

Tianyi Lin, Chi Jin, and Michael Jordan. On gradient descent ascent for nonconvex-concave minimax problems. In *International Conference on Machine Learning*, pp. 6083–6093. PMLR, 2020a.

Tianyi Lin, Chi Jin, and Michael I Jordan. Near-optimal algorithms for minimax optimization. In *Conference on Learning Theory*, pp. 2738–2779. PMLR, 2020b.

Songtao Lu, Ioannis Tsaknakis, and Mingyi Hong. Block alternating optimization for non-convex min-max problems: algorithms and applications in signal processing and communications. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4754–4758. IEEE, 2019.

R Mehra and EC Prescott. Recursive competitive equilibria and capital asset pricing. *Essays in Financial Economics*, 1977.

Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. Convergence rate of $\mathcal{O}(1/k)$ for optimistic gradient and extra-gradient methods in smooth convex-concave saddle point problems, 2020.

Arkadi Nemirovski. Prox-method with rate of convergence o (1/t) for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

Yurii Nesterov. Dual extrapolation and its applications to solving variational inequalities and related problems. *Mathematical Programming*, 109(2):319–344, 2007.

Yurii Nesterov and Laura Scrimali. Solving strongly monotone variational and quasi-variational inequalities. *Discrete & Continuous Dynamical Systems*, 31(4):1383–1396, 2011.

J v Neumann. Zur theorie der gesellschaftsspiele. *Mathematische annalen*, 100(1):295–320, 1928.

Maher Nouiehed, Maziar Sanjabi, Tianjian Huang, Jason D Lee, and Meisam Razaviyayn. Solving a class of non-convex min-max games using iterative first order methods. *arXiv preprint arXiv:1902.08297*, 2019.

Bernt Oksendal, Leif Sandal, and Jan Uboe. Stochastic stackelberg equilibria with applications to time-dependent newsvendor models. *Journal of Economic Dynamics and Control*, 37(7):1284–1299, 2013.

Dmitrii M Ostrovskii, Andrew Lowy, and Meisam Razaviyayn. Efficient search of first-order nash equilibria in nonconvex-concave smooth min-max problems. *arXiv preprint arXiv:2002.07919*, 2020.

Yuyuan Ouyang and Yangyang Xu. Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems, 2018.

Hassan Rafique, Mingrui Liu, Qihang Lin, and Tianbao Yang. Non-convex min-max optimization: Provable algorithms and applications in machine learning, 2019.

Zhao Renbo. Optimal algorithms for stochastic three-composite convex-concave saddle point problems, 2019.

Sean C Rismiller, Jonathan Cagan, and Christopher McComb. Stochastic stackelberg games for agent-driven robust design. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 84003, pp. V11AT11A039. American Society of Mechanical Engineers, 2020.

Maziar Sanjabi, Jimmy Ba, Meisam Razaviyayn, and Jason D. Lee. On the convergence and robustness of training gans with regularized optimal transport. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 7091–7101, Red Hook, NY, USA, 2018. Curran Associates Inc.

Lloyd S Shapley. Stochastic games. *Proceedings of the national academy of sciences*, 39(10):1095–1100, 1953.

Maurice Sion et al. On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176, 1958.

Kiran Koshy Thekumparampil, Prateek Jain, Praneeth Netrapalli, and Sewoong Oh. Efficient algorithms for smooth minimax optimization, 2019.

Paul Tseng. On linear convergence of iterative methods for the variational inequality problem. *Journal of Computational and Applied Mathematics*, 60(1):237–252, 1995. ISSN 0377-0427. doi: https://doi.org/10.1016/0377-0427(94)00094-H. URL https://www.sciencedirect.com/science/article/pii/037704279400094H. Proceedings of the International Meeting on Linear/Nonlinear Iterative Methods and Verification of Solution.

Paul Tseng. On accelerated proximal gradient methods for convex-concave optimization. *submitted to SIAM Journal on Optimization*, 1, 2008.

Deepanshu Vasal. Stochastic stackelberg games, 2020.

Heinrich Von Stackelberg. *Marktform und gleichgewicht*. J. springer, 1934.

Yevgeniy Vorobeychik and Satinder Singh. Computing stackelberg equilibria in discounted stochastic games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, pp. 1478–1484, 2012.

Yu Yuan, Zhibin Liang, and Xia Han. Robust reinsurance contract with asymmetric information in a stochastic stackelberg differential game. *Scandinavian Actuarial Journal*, pp. 1–28, 2021.

Renbo Zhao. A primal dual smoothing framework for max-structured nonconvex optimization, 2020.

A Pseudo-Code for Algorithms

Algorithm 1 Value Iteration (with Min-Max Oracle)

Inputs: $\mathcal{S}, \mathcal{X}, \mathcal{Y}, r, \boldsymbol{g}, p, \gamma, T$
Outputs: $v^{(T)}$

Initialize $v^{(0)}$ arbitrarily, e.g. $v^{(0)} = 0$
for $t = 1, \ldots, T$ **do**
  for $\boldsymbol{s} \in \mathcal{S}$ **do**
    $v^{(t+1)}(\boldsymbol{s}) = \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}} \mathbb{E}_{\boldsymbol{S}' \sim p(\cdot | \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})} \left[r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma v^{(t)}(\boldsymbol{S}') \right]$
  end for
end for
return $v^{(T)}$

Algorithm 2 Nested GDA on Action Value Function

Inputs: $v, \boldsymbol{b}, \eta_{\boldsymbol{x}}, \eta_{\boldsymbol{y}}, T_{\boldsymbol{p}}, T_{\boldsymbol{X}}$
Output: $(\boldsymbol{p}^{(t)}, \boldsymbol{X}^{(t)})_{t=1}^{T}$

for $t = 1, \ldots, T_{\boldsymbol{p}}$ **do**

  for $s = 1, \ldots, T_{\boldsymbol{X}}$ **do**
    For all $i \in [n]$,
    $\boldsymbol{x}_i^{(s)} = \Pi_{\{(\boldsymbol{x}_i, \beta_i) \in \mathbb{R}_+^m \times \mathbb{R}_+ : \boldsymbol{x}_i \boldsymbol{p}^{(t-1)} + \beta_i \leq b_i\}} \left(\boldsymbol{x}_i^{(s-1)} + \eta_{\boldsymbol{X}} \left(\frac{b_i - \beta_i^{(s-1)}}{u_i^{(t)}\left(\boldsymbol{x}_i^{(t)}\right)} \nabla_{\boldsymbol{x}_i} u_i^{(t)}\left(\boldsymbol{x}_i^{(t)}\right) \right) \right)$
    $\beta_i^{(s)} = \Pi_{\{(\boldsymbol{x}_i, \beta_i) \in \mathbb{R}_+^m \times \mathbb{R}_+ : \boldsymbol{x}_i \boldsymbol{p}^{(t-1)} + \beta_i \leq b_i\}} \left(\beta_i^{(s-1)} + \eta_{\boldsymbol{X}} \left(-\log(u_i(\boldsymbol{x}_i{}^{(s-1)})) + \gamma \frac{\partial v(\boldsymbol{b})}{\partial b_i} \right) \right)$
  end for
  $\boldsymbol{p}^{(t)} = \Pi_{\mathbb{R}_+^m} \left(\boldsymbol{p}^{(t-1)} - \eta_{\boldsymbol{p}} (\boldsymbol{1} - \sum_{i \in [n]} \boldsymbol{x}_i) \right)$
end for
return $(\boldsymbol{p}^{(t)}, \boldsymbol{X}^{(t)})_{t=1}^{T}$

Algorithm 3 Max-Oracle GDA with Simulated Annealing on Action Value Function

Inputs: $v, \boldsymbol{b}, \eta_{\boldsymbol{x}}, \eta_{\boldsymbol{y}}, T_{\boldsymbol{p}}, T_{\boldsymbol{X}}$
Output: $(\boldsymbol{p}^{(t)}, \boldsymbol{X}^{(t)})_{t=1}^{T}$

for $t = 1, \ldots, T_{\boldsymbol{p}}$ **do**
  $q(\boldsymbol{p}, \boldsymbol{X}, \boldsymbol{\beta}) = \sum_{j \in [m]} q_j p_j + \sum_{i \in [n]} (b_i - \beta_i) \log(u_i(\boldsymbol{x}_i, t_i)) + \gamma\, \mathbb{E}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}') \sim p(\cdot | \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}, \boldsymbol{X}, \boldsymbol{\beta})} \left[v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}, \boldsymbol{q}') \right]$
  For all $i \in [n]$, $\boldsymbol{x}_i^{(t)}, \beta_i^{(t)} = \text{Simulated\_Annealing}(q)$
  $\boldsymbol{x}_i^{(t)}, \beta_i^{(t)} = \Pi_{\{(\boldsymbol{x}_i, \beta_i) \in \mathbb{R}_+^m \times \mathbb{R}_+ : \boldsymbol{x}_i \boldsymbol{p}^{(t-1)} + \beta_i \leq b_i\}} \left[\left(\boldsymbol{x}_i^{(t)}, \beta_i^{(t)} \right) \right]$
end for
$\boldsymbol{p}^{(t)} = \Pi_{\mathbb{R}_+^m} \left(\boldsymbol{p}^{(t-1)} + \eta_{\boldsymbol{p}} (\boldsymbol{1} - \sum_{i \in [n]} \boldsymbol{x}_i) \right)$
return $(\boldsymbol{p}^{(t)}, \boldsymbol{X}^{(t)})_{t=1}^{T}$

Algorithm 4 Value Iteration on Stochastic Fisher Market

Initialize $v^{(0)}$ arbitrarily, e.g. $v^{(0)} = 0$
for $k = 1, \dots, T_v$ **do**
  for $\boldsymbol{s} \in \mathcal{S}$ **do**
    $v^{(k+1)}(\boldsymbol{s}) = (1-\gamma) \min_{\boldsymbol{p} \in \mathbb{R}^m_+} \max_{(\boldsymbol{X}, \boldsymbol{\beta}) \in \mathbb{R}^{n \times m}_+ \times \mathbb{R}^n_+ : \boldsymbol{X}\boldsymbol{p} + \boldsymbol{\beta} \leq \boldsymbol{b}} \sum_{j \in [m]} q_j p_j + \sum_{i \in [n]} (b_i - \beta_i) \log(u_i(\boldsymbol{x}_i)) + \gamma v^*(\boldsymbol{\beta})$
  end for
end for

B Omitted Results and Proofs Section 2

We first note the following fundamental relationship between the state-value and action-value functions which is an analog of Bellman's Theorem (Bellman, 1952) and which follows from their definitions:

Theorem B.1. *Given a stochastic min-max Stackelberg game $(\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r, \boldsymbol{g}, p, \gamma)$, for all $v \in \mathcal{V}$, $q \in \mathcal{Q}$, $\boldsymbol{\pi_x} \in \mathcal{X}^{\mathcal{S}}$, and $\boldsymbol{\pi_y} \in \mathcal{Y}^{\mathcal{S}}$, $v = v^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}$ and $q = q^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}$ iff:*

$$v(\boldsymbol{s}) = q(\boldsymbol{s}, \boldsymbol{\pi_x}(\boldsymbol{s}), \boldsymbol{\pi_y}(\boldsymbol{s})) \tag{7}$$

$$q(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) = \mathop{\mathbb{E}}_{\boldsymbol{S}' \sim p(\cdot | \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})} [r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma v(\boldsymbol{S}')] \tag{8}$$

Proof of Theorem B.1. By the definition of the state value function we have $v_i^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}} = q_i(\boldsymbol{s}, \boldsymbol{\pi_x}(\boldsymbol{s}), \boldsymbol{\pi_y}(\boldsymbol{s}))$, hence by Equation (7) we must have that $v_i = v_i^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}$. Additionally, by Equation (8) and the definition of the action-value functions this also implies that $q_i(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) = q_i^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ □

Lemma B.2. *Suppose that there exists $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*) \in \mathcal{S}^{\mathcal{X}} \times \mathcal{S}^{\mathcal{Y}}$ such that for all states $\boldsymbol{s} \in \mathcal{S}$:*

$$\max_{\boldsymbol{\pi_y} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi_y}) \geq \boldsymbol{0}} v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi_y}}(\boldsymbol{s}) \leq v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}) \leq \min_{\boldsymbol{\pi_x} \in \mathcal{S}^{\mathcal{X}}} \max_{\boldsymbol{\pi_y} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \boldsymbol{0}} v^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}(\boldsymbol{s}) \ . \tag{9}$$

Then, $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ is a recSE.

Lemma B.2. *Suppose that there exists $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*) \in \mathcal{S}^{\mathcal{X}} \times \mathcal{S}^{\mathcal{Y}}$ such that for all states $\boldsymbol{s} \in \mathcal{S}$:*

$$\max_{\boldsymbol{\pi_y} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi_y}) \geq \boldsymbol{0}} v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi_y}}(\boldsymbol{s}) \leq v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}) \leq \min_{\boldsymbol{\pi_x} \in \mathcal{S}^{\mathcal{X}}} \max_{\boldsymbol{\pi_y} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \boldsymbol{0}} v^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}(\boldsymbol{s}) \ . \tag{9}$$

Then, $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ is a recSE.

Proof of Lemma B.2. We prove the right hand side inequality first:

$$v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}) \leq \min_{\boldsymbol{\pi_x} \in \mathcal{S}^{\mathcal{X}}} \max_{\boldsymbol{\pi_y} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \boldsymbol{0}} v^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}(\boldsymbol{s}) \tag{10}$$

$$\sum_{\boldsymbol{s} \in \mathcal{S}} \mu^{(0)}(\boldsymbol{s}) v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}) \leq \sum_{\boldsymbol{s} \in \mathcal{S}} \mu^{(0)}(\boldsymbol{s}) \min_{\boldsymbol{\pi_x} \in \mathcal{S}^{\mathcal{X}}} \max_{\boldsymbol{\pi_y} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \boldsymbol{0}} v^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}(\boldsymbol{s}) \tag{11}$$

$$u\left(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*\right) \leq \sum_{\boldsymbol{s} \in \mathcal{S}} \mu^{(0)}(\boldsymbol{s}) \min_{\boldsymbol{\pi_x} \in \mathcal{S}^{\mathcal{X}}} \max_{\boldsymbol{\pi_y} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \boldsymbol{0}} v^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}(\boldsymbol{s}) \tag{12}$$

Since $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ is well-defined, it must mean that the same policy minimizes (resp. maximizes) the value of each state for the $\boldsymbol{x}$ player ($\boldsymbol{y}$-player), hence the min-max operator can be pulled out of the sum, giving us:

$$u\left(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*\right) \leq \min_{\boldsymbol{\pi_x} \in \mathcal{S}^{\mathcal{X}}} \max_{\boldsymbol{\pi_y} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \boldsymbol{0}} \sum_{\boldsymbol{s} \in \mathcal{S}} \mu^{(0)}(\boldsymbol{s}) v^{\boldsymbol{\pi_x}\boldsymbol{\pi_y}}(\boldsymbol{s}) \tag{13}$$

$$u\left(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*\right) \leq \min_{\boldsymbol{\pi_x} \in \mathcal{S}^{\mathcal{X}}} \max_{\boldsymbol{\pi_y} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}) \geq \boldsymbol{0}} u\left(\boldsymbol{\pi_x}, \boldsymbol{\pi_y}\right) \tag{14}$$

Similarly, manipulating the the left hand side inequality , we obtain:

$$\min_{\boldsymbol{\pi}_{\boldsymbol{x}} \in \mathcal{S}^{\mathcal{X}}} \max_{\boldsymbol{\pi}_{\boldsymbol{y}} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) \geq \mathbf{0}} u\left(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}\right) \leq u\left(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*\right) \tag{15}$$

□

We will show, there always exists a policy profile which satisfies the conditions of Lemma B.2. To show the existence of such a policy profile, we first re-state the condition given in Lemma B.2, in terms of action value functions.

Lemma 2.1. *Consider a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$. A policy profile $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*) \in \mathcal{S}^{\mathcal{X}} \times \mathcal{S}^{\mathcal{Y}}$ is a recSE if, for all $\boldsymbol{s} \in \mathcal{S}$, we have that: $\max_{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y}) \geq \mathbf{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y}) \leq q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{x}), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{y})) \leq \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \mathbf{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$. Equivalently, a policy profile $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ is a recSE if $(\boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{s}))$ is a Stackelberg equilibrium: i.e., a solution to $\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \mathbf{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ at each $\boldsymbol{s} \in \mathcal{S}$.*

Proof of Lemma 2.1.

$$\max_{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y}) \geq \mathbf{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y}) \leq q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{s})) \leq \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \mathbf{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \tag{16}$$

$$\max_{\boldsymbol{\pi}_{\boldsymbol{y}} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}) \geq \mathbf{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}(\boldsymbol{s})) \leq q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{s})) \leq \min_{\boldsymbol{\pi}_{\boldsymbol{x}} \in \mathcal{S}^{\mathcal{X}}} \max_{\boldsymbol{\pi}_{\boldsymbol{y}} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) \geq \mathbf{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}(\boldsymbol{s}) \tag{17}$$

$$\max_{\boldsymbol{\pi}_{\boldsymbol{y}} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}) \geq \mathbf{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}(\boldsymbol{s})) \leq q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{s})) \leq \min_{\boldsymbol{\pi}_{\boldsymbol{x}} \in \mathcal{S}^{\mathcal{X}}} \max_{\boldsymbol{\pi}_{\boldsymbol{y}} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) \geq \mathbf{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}} \boldsymbol{\pi}_{\boldsymbol{y}}}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}(\boldsymbol{s}) \tag{18}$$

$$\max_{\boldsymbol{\pi}_{\boldsymbol{y}} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}) \geq \mathbf{0}} v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}}(\boldsymbol{s}) \leq v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}) \leq \min_{\boldsymbol{\pi}_{\boldsymbol{x}} \in \mathcal{S}^{\mathcal{X}}} \max_{\boldsymbol{\pi}_{\boldsymbol{y}} \in \mathcal{S}^{\mathcal{Y}} : \boldsymbol{G}(\boldsymbol{\pi}_{\boldsymbol{x}}, \boldsymbol{\pi}_{\boldsymbol{y}}) \geq \mathbf{0}} v^{\boldsymbol{\pi}_{\boldsymbol{x}} \boldsymbol{\pi}_{\boldsymbol{y}}}(\boldsymbol{s}) \tag{19}$$

By Lemma B.2, we must then have that $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ is a recursive Stackelberg equilibrium of $(\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r, \boldsymbol{g}, p, \gamma)$.

□

Theorem 2.2. *$(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ is a recSE of $\mathcal{G}^{(0)}$ of $v^{\boldsymbol{\pi}_{\boldsymbol{x}} \boldsymbol{\pi}_{\boldsymbol{y}}}$ iff it induces a value function which is a fixed point of C: i.e., $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ is a Stackelberg equilbrium iff, for all $\boldsymbol{s} \in \mathcal{S}$, $\left(C v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}\right)(\boldsymbol{s}) = v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s})$.*

Proof of Theorem 2.2. (recursive Stackelberg equilibrium $\implies$ Fixed Point):
Suppose that $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ is a recursive Stackelberg equilibrium of $(\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r, \boldsymbol{g}, p, \gamma)$, then:

$$\left(C v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}\right)(\boldsymbol{s}) = \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \mathbf{0}} \mathbb{E}_{\boldsymbol{S}' \sim p(\cdot | \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})} \left[r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{S}') \right] \tag{20}$$

$$= \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \mathbf{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \tag{21}$$

$$= v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*} \qquad \text{(Lemma 2.1)} \tag{22}$$

(Fixed Point $\implies$ recursive Stackelberg equilibrium) Suppose that a value function $v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}$ which is induced by a policy profile $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ is a fixed point of C, we then have for all states $\boldsymbol{s} \in \mathcal{S}$:

$$v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*} = \left(C v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}\right)(\boldsymbol{s}) \tag{23}$$

$$= \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \mathbf{0}} \mathbb{E}_{\boldsymbol{S}' \sim p(\cdot | \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})} \left[r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{S}') \right] \tag{24}$$

$$= \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \mathbf{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \tag{25}$$

Hence, by Lemma 2.1, $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}})$ is recursive Stackelberg equilibrium. □

Lemma 2.3. *Suppose that $f, h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, $\boldsymbol{g} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$ are continuous functions, and $\mathcal{X}, \mathcal{Y}$ are compact sets, we then have: $\left|\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} f(\boldsymbol{x},\boldsymbol{y}) - \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} h(\boldsymbol{x},\boldsymbol{y})\right| \leq \max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}\times\mathcal{Y}} |f(\boldsymbol{x},\boldsymbol{y}) - h(\boldsymbol{x},\boldsymbol{y})|$.*

Proof of Lemma 2.3. Let $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be a Stackelberg equilibrium of $\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} f(\boldsymbol{x},\boldsymbol{y})$, and $(\boldsymbol{x}', \boldsymbol{y}')$ be a Stackelberg equilibrium of $\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} h(\boldsymbol{x},\boldsymbol{y})$. Additionally, let $\bar{y} \in \arg\max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x}',\boldsymbol{y})\geq 0} f(\boldsymbol{x}', \boldsymbol{y})$, then by the the definition of a Stackelberg equilibrium, we have $f(\boldsymbol{x}^*, \boldsymbol{y}^*) = \min_{\boldsymbol{y}\in\mathcal{Y}} \max_{\boldsymbol{x}\in\mathcal{X}:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} \leq \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x}',\boldsymbol{y})\geq 0} f(\boldsymbol{x}', \boldsymbol{y}) = f(\boldsymbol{x}', \bar{\boldsymbol{y}})$, and $h(\boldsymbol{x}', \boldsymbol{y}') = \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x}',\boldsymbol{y})\geq 0} h(\boldsymbol{x}', \boldsymbol{y}) \geq h(\boldsymbol{x}', \bar{\boldsymbol{y}})$.

Suppose that $\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} f(\boldsymbol{x},\boldsymbol{y}) \geq \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}}, h(\boldsymbol{x},\boldsymbol{y})$ this gives us:

$$
\begin{aligned}
&\left|\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} f(\boldsymbol{x},\boldsymbol{y}) - \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} h(\boldsymbol{x},\boldsymbol{y})\right| && (26)\\
&= |f(\boldsymbol{x}^*, \boldsymbol{y}^*) - h(\boldsymbol{x}', \boldsymbol{y}')| && (27)\\
&\leq |f(\boldsymbol{x}', \bar{\boldsymbol{y}}) - h(\boldsymbol{x}', \boldsymbol{y}')| && (28)\\
&\leq |f(\boldsymbol{x}', \bar{\boldsymbol{y}}) - h(\boldsymbol{x}', \bar{\boldsymbol{y}})| && (29)\\
&\leq \max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}\times\mathcal{Y}} |f(\boldsymbol{x},\boldsymbol{y}) - h(\boldsymbol{x},\boldsymbol{y})| && (30)
\end{aligned}
$$

The opposite case follows similarly by symmetry. □

Theorem 2.4. *Consider the operator C associated with a stochastic Stackelberg game $\mathcal{G}^{(0)}$. If Assumption 1.1 holds, then C is a contraction mapping w.r.t. to the sup norm $\|.\|_\infty$ with constant γ.*

Proof of Theorem 2.4. We will show that C is a contraction mapping, which then by Banach fixed point theorem establish the result. Let $v, v' \in \mathcal{V}$ be any two state value functions and $q, q' \in \mathcal{Q}$ be the respective associated action-value functions. We then have:

$$
\begin{aligned}
&\|Cv - Cv'\|_\infty && (31)\\
&\leq \max_{\boldsymbol{s}\in\mathcal{S}} \left|\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} q(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) - \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} q'(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\right| && (32)\\
&\leq \max_{\boldsymbol{s}\in\mathcal{S}} \max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}\times\mathcal{Y}} |q(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) - q'(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})| && \text{(Lemma 2.3)}\\
& && (33)\\
&\leq \max_{\boldsymbol{s}\in\mathcal{S}} \max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}\times\mathcal{Y}} \left|\mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})} [r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) + \gamma v(\boldsymbol{S}')] - \mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})} [r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) + \gamma v'(\boldsymbol{S}')]\right| && (34)\\
&\leq \max_{\boldsymbol{s}\in\mathcal{S}} \max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}\times\mathcal{Y}} \left|\mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})} [\gamma v(\boldsymbol{S}') - \gamma v'(\boldsymbol{S}')]\right| && (35)\\
&\leq \gamma \max_{\boldsymbol{s}\in\mathcal{S}} \max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}\times\mathcal{Y}} \left|\mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})} [v(\boldsymbol{S}') - v'(\boldsymbol{S}')]\right| && (36)\\
&\leq \gamma \max_{\boldsymbol{s}\in\mathcal{S}} \max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}\times\mathcal{Y}} |v(\boldsymbol{s}) - v'(\boldsymbol{s})| && (37)\\
&= \gamma \|v - v'\|_\infty && (38)
\end{aligned}
$$

Since $\gamma \in (0, 1)$, C is a contraction mapping.

□

Theorem 2.5. *Consider a zero-sum stochastic Stackelberg game $\mathcal{G}^{(0)}$. If Assumption 1.1 holds, then $\mathcal{G}^{(0)}$ has at least one recSE $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}})$ with unique value function $v^{\boldsymbol{\pi}^*_{\boldsymbol{x}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}$. Further, $v^{\boldsymbol{\pi}^*_{\boldsymbol{x}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}$ can be computed by iteratively applying C to any initial state-value function $v^{(0)} \in \mathcal{V}$: $\lim_{t\to\infty} v^{(t)} = v^{\boldsymbol{\pi}^*_{\boldsymbol{x}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}$.*

Proof of Theorem 2.5. By combining Theorem 2.4 and the Banach fixed point theorem Banach (1922), we obtain that a fixed point of C exists. Hence, by Theorem 2.2, a recursive Stackelberg equilibrium of $(\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r, \boldsymbol{g}, p, \gamma)$ exists and the value function induced by all recursive Stackelberg equilibria is the same, i.e., the optimal value function is unique. Additionally, by the second part of the Banach fixed point theorem, we must then also have $\lim_{t\to\infty} v^{(t)} \to v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}$. □

For any $q \in \mathcal{Q}$, we define a **greedy policy profile** with respect to q as a policy profile $(\boldsymbol{\pi}_{\boldsymbol{x}}^q, \boldsymbol{\pi}_{\boldsymbol{y}}^q)$ such that $\boldsymbol{\pi}_{\boldsymbol{x}}^q \in \arg\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq \boldsymbol{0}} q(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ and $\boldsymbol{\pi}_{\boldsymbol{y}}^q \in \arg\max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{x}),\boldsymbol{y})\geq \boldsymbol{0}} q(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{x}), \boldsymbol{y})$. The following lemma provides a progress bound for each iteration of value iteration which is expressed in terms of the value function associated with the greedy policy profile.

Lemma B.3. *Let $(\boldsymbol{\pi}_{\boldsymbol{x}}^*, \boldsymbol{\pi}_{\boldsymbol{y}}^*)$ be the recSEof a zero-sum Stochastic Stackelberg game $(\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r, \boldsymbol{g}, p, \gamma)$. For any $q \in \mathcal{Q}$, let $\boldsymbol{\pi}_{\boldsymbol{x}}^q$, $\boldsymbol{\pi}_{\boldsymbol{y}}^q$ denote the greedy policy with respect to q of $\boldsymbol{x}$-player and $\boldsymbol{y}$-player respectively. Then, the following bound holds:*

$$v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q \boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s}) - v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}) \leq \frac{2\|q - q^*\|_\infty}{(1-\gamma)} \tag{39}$$

Using the above progress bound, we can then obtain a convergence rate for value iteration under Assumption 1.1.

Proof of Lemma B.3. For any state $\boldsymbol{s}$, we have:

$$v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}) - v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q \boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s}) \tag{40}$$

$$= q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{s})) - q^{\boldsymbol{\pi}_{\boldsymbol{x}}^q \boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s})) \tag{41}$$

$$= q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{s})) - q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s})) + q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s})) - q^{\boldsymbol{\pi}_{\boldsymbol{x}}^q \boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s})) \tag{42}$$

$$= q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{s})) - q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s})) + \gamma \mathop{\mathbb{E}}_{\boldsymbol{s}'\sim p(\cdot|\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s}))} [v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}') - v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q \boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s}')] \tag{43}$$

$$\geq q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{s})) + \left[q(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s})) - \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y})\geq \boldsymbol{0}} q(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y}) \right] - q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s}))$$

$$+ \gamma \mathop{\mathbb{E}}_{\boldsymbol{s}'\sim p(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s}))} [v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}') - v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q \boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s}')] \tag{44}$$

$$\tag{45}$$

where the last line was obtained by the definition of a recursive Stackelberg equilibrium which ensures that $q(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s})) \leq \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y})\geq \boldsymbol{0}} q(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y})$.

Re-organizing terms, we have:

$$v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}) - v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q \boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s}) \tag{46}$$

$$\geq \left[q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^*(\boldsymbol{s})) - \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y})\geq \boldsymbol{0}} q(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y}) \right] + \left[q(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s})) - q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s})) \right]$$

$$+ \gamma \mathop{\mathbb{E}}_{\boldsymbol{s}'\sim p(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s}))} [v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}') - v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q \boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s}')] \tag{47}$$

$$= \left[\max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y})\geq \boldsymbol{0}} q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y}) - \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y})\geq \boldsymbol{0}} q(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^*(\boldsymbol{s}), \boldsymbol{y}) \right]$$

$$+ \left[q(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s})) - q^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s})) \right] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{s}'\sim p(\boldsymbol{s}, \boldsymbol{\pi}_{\boldsymbol{x}}^q(\boldsymbol{s}), \boldsymbol{\pi}_{\boldsymbol{y}}^q(\boldsymbol{s}))} [v^{\boldsymbol{\pi}_{\boldsymbol{x}}^* \boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s}') - v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q \boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s}')] \tag{48}$$

Taking the minimum of both sides over states, and re-collecting terms, we obtain:

$$\min_{\boldsymbol{s}\in\mathcal{S}}\left(v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s})-v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q\boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s})\right)\geq -2\|q-q^*\|_\infty+\gamma\min_{\boldsymbol{s}\in\boldsymbol{s}}\left[v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s})-v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q\boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s})\right] \tag{49}$$

$$(1-\gamma)\min_{\boldsymbol{s}\in\mathcal{S}}\left(v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s})-v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q\boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s})\right)\geq -2\|q-q^*\|_\infty \tag{50}$$

$$\min_{\boldsymbol{s}\in\mathcal{S}}\left(v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s})-v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q\boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s})\right)\geq -\frac{2\|q-q^*\|_\infty}{(1-\gamma)} \tag{51}$$

$$v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s})-v^{\boldsymbol{\pi}_{\boldsymbol{x}}^q\boldsymbol{\pi}_{\boldsymbol{y}}^q}(\boldsymbol{s})\geq -\frac{2\|q-q^*\|_\infty}{(1-\gamma)} \qquad \forall \boldsymbol{s}\in\mathcal{S} \tag{52}$$

□

Theorem 2.8. *[Convergence of Value Iteration] Suppose value iteration is run on input $\mathcal{G}$. If Assumption 1.1 holds, and if we initialize $v^{(0)}(\boldsymbol{s})=0$, for all $\boldsymbol{s}\in\mathcal{S}$, then for $k\geq\frac{1}{1-\gamma}\log\frac{2\alpha}{\epsilon(1-\gamma)}$, we have $v^{(k)}(\boldsymbol{s})-v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s})\leq\epsilon$.*

Proof of Theorem 2.8. First note that by Assumption 1.1, we have that $\left\|v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}\right\|_\infty\leq\alpha$. Applying the operator C repeatedly and using the fact that $v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}=Cv^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}$ from Theorem 2.2, we obtain

$$\begin{aligned}
&\|v^{(k)}-v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}\|_\infty\\
&=\|(C)^kv^{(0)}-(C)^kv^*\|_\infty\\
&\leq\gamma^k\|v^{(0)}-v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}\|_\infty\\
&=\|(C)^kq^{(0)}-(C)^kq^*\|_\infty\\
&\leq\gamma^k\|v^{(0)}-v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}\|_\infty\\
&=\gamma^k\max_{\boldsymbol{s}\in\mathcal{S}}\left|\min_{\boldsymbol{x}\in\mathcal{X}}\max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}}q^{(0)}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})-\min_{\boldsymbol{x}\in\mathcal{X}}\max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}}q^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\right|\\
&=\gamma^k\max_{\boldsymbol{s}\in\mathcal{S}}\max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}\times\mathcal{Y}}\left|q^{(0)}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})-q^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\right| \qquad \text{(Lemma 2.3)}\\
&=\gamma^k\|q^*\|_\infty\\
&\leq\gamma^k\alpha
\end{aligned}$$

Since $1-x\leq e^{-x}$ for any $x\in\mathbb{R}$, we have

$$\gamma^k=(1-(1-\gamma))^k\leq(e^{-(1-\gamma)})^k\leq e^{-(1-\gamma)k}$$

Combining this bound with Equation (39) for all $q\in\mathcal{Q}$:

$$\begin{aligned}
v^{(k)}(\boldsymbol{s})-v^{\boldsymbol{\pi}_{\boldsymbol{x}}^*\boldsymbol{\pi}_{\boldsymbol{y}}^*}(\boldsymbol{s})&\leq\frac{2\|q-q^*\|_\infty}{1-\gamma}\\
&\leq\frac{2\gamma^k\alpha}{(1-\gamma)}\\
&\leq\frac{2\alpha}{(1-\gamma)}e^{-(1-\gamma)k}
\end{aligned}$$

Thus it suffices to solve for k such that

$$\frac{2\alpha}{(1-\gamma)}e^{-(1-\gamma)k}\leq\varepsilon\ .$$

which concludes the proof. □

C Omitted Results and Proofs Section 3

Our characterization of the subdifferential of the value function associated with a Stackelberg equilibrium w.r.t. its parameters relies on a slightly generalized version of the subdifferential envelope theorem (Theorem C.1, Appendix C) of Goktas & Greenwald (2021), which characterizes the set of subdifferentials of parametrized constrained optimization problems, i.e., the set of subgradients w.r.t. $\boldsymbol{x}$ of $f^*(\boldsymbol{x}) = \max_{\boldsymbol{y} \in \mathcal{Y}: h(\boldsymbol{x}, \boldsymbol{y}) \geq 0} f(\boldsymbol{x}, \boldsymbol{y})$. In particular, we note that Goktas & Greenwald's proof goes through even without assuming the concavity of $f(\boldsymbol{x}, \boldsymbol{y}), h_1(\boldsymbol{x}, \boldsymbol{y}), \ldots, h_d(\boldsymbol{x}, \boldsymbol{y})$ in $\boldsymbol{y}$, for all $\boldsymbol{x} \in \mathcal{X}$.

Theorem C.1 (Subdifferential Envelope Theorem)**.** *Consider the function $f^*(\boldsymbol{x}) = \max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}) \geq 0} f(\boldsymbol{x}, \boldsymbol{y})$ where $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, and $\boldsymbol{h} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$. Let $\mathcal{Y}^*(\boldsymbol{x}) = \arg\max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}} f(\boldsymbol{x}, \boldsymbol{y})$. Suppose that 1. $f(\boldsymbol{x}, \boldsymbol{y}), h_1(\boldsymbol{x}, \boldsymbol{y}), \ldots, h_d(\boldsymbol{x}, \boldsymbol{y})$ are continuous in $(\boldsymbol{x}, \boldsymbol{y})$ and convex in $\boldsymbol{x}$; 2. $\nabla_{\boldsymbol{x}} f, \nabla_{\boldsymbol{x}} h_1, \ldots, \nabla_{\boldsymbol{x}} h_d$ exist for all $\boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \in \mathcal{Y}$; 3. $\mathcal{Y}$ is non-empty and compact, and 4. (Slater's condition) $\forall \boldsymbol{x} \in \mathcal{X}, \exists \widehat{\boldsymbol{y}} \in \mathcal{Y}$ s.t. $g_k(\boldsymbol{x}, \widehat{\boldsymbol{y}}) > 0$, for all $k = 1, \ldots, d$. Then, f^* is subdifferentiable and at any point $\widehat{\boldsymbol{x}} \in \mathcal{X}$, $\partial_{\boldsymbol{x}} f^*(\widehat{\boldsymbol{x}}) =$*

$$\operatorname{conv}\left(\bigcup_{\boldsymbol{y}^*(\widehat{\boldsymbol{x}}) \in \mathcal{Y}^*(\widehat{\boldsymbol{x}})} \bigcup_{\lambda_k^*(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})) \in \Lambda^*(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}}))} \left\{ \nabla_{\boldsymbol{x}} f\left(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})\right) + \sum_{k=1}^{d} \lambda_k^*(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})) \nabla_{\boldsymbol{x}} g_k\left(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})\right) \right\}\right), \tag{53}$$

where conv *is the convex hull operator and $\boldsymbol{\lambda}^*(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})) = (\lambda_1^*(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})), \ldots, \lambda_d^*(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})))^T \in \Lambda^*(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}}))$ are the optimal KKT multipliers associated with $\boldsymbol{y}^*(\widehat{\boldsymbol{x}}) \in \mathcal{Y}^*(\widehat{\boldsymbol{x}})$.*

We note the following known lemma which provides the necessary conditions for the Bellman equation associated with a recursive optimization problem to be continuous and convex in its parameters.[6]

Lemma C.2. *Consider the Bellman equation associated with a parametric recursive stochastic optimization problem, where $r : \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, with state space $\mathcal{S}$ and parameter set $\mathcal{X}$, and $\gamma \in (0, 1)$:*

$$v(\boldsymbol{s}, \boldsymbol{x}) = \max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}} \left\{ r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma \mathop{\mathbb{E}}_{\boldsymbol{S}' \sim p(\cdot \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})} \left[v(\boldsymbol{S}', \boldsymbol{x}) \right] \right\} \tag{54}$$

Suppose that Assumption 1.1 holds, and that 1. for all $\boldsymbol{s} \in \mathcal{S}, \boldsymbol{y} \in \mathcal{Y}$, $r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$, $g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \ldots, g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ are concave in $\boldsymbol{x}$, 2. $p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ is continuous (Feller property) and CSD convex in $\boldsymbol{x}$. Then, v is continuous and convex in $\boldsymbol{x}$.

Theorem C.3. *[Subdifferential Benveniste-Scheinkman Theorem] Consider the Bellman equation associated with a recursive stochastic optimization problem where $r : \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, with state space $\mathcal{S}$ and parameter set $\mathcal{X}$, and $\gamma \in (0, 1)$:*

$$v(\boldsymbol{s}, \boldsymbol{x}) = \max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}} \left\{ r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma \mathop{\mathbb{E}}_{\boldsymbol{S}' \sim p(\cdot \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})} \left[v(\boldsymbol{S}', \boldsymbol{x}) \right] \right\} \tag{55}$$

Suppose that Assumption 1.1 holds, and that 1. for all $\boldsymbol{s} \in \mathcal{S}, \boldsymbol{y} \in \mathcal{Y}$, $r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$, $g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \ldots, g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ are concave in $\boldsymbol{x}$, 2. $\nabla_{\boldsymbol{x}} r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \nabla_{\boldsymbol{x}} g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \ldots, \nabla_{\boldsymbol{x}} g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \nabla_{\boldsymbol{x}} p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ exist for all $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \in \mathcal{Y}$, 3. $p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ is continuous CSD convex, and differentiable in $\boldsymbol{x}$, and 4. Slater's condition holds for the optimization problem, i.e., $\forall \boldsymbol{x} \in \mathcal{X}, \boldsymbol{s} \in \mathcal{S}, \exists \widehat{\boldsymbol{y}} \in \mathcal{Y}$ s.t. $g_k(\boldsymbol{s}, \boldsymbol{x}, \widehat{\boldsymbol{y}}) > 0$, for all $k = 1, \ldots, d$.
Let $\mathcal{Y}^*(\boldsymbol{s}, \boldsymbol{x}) = \max_{\boldsymbol{y} \in \mathcal{Y}: \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}} \left\{ r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) + \gamma \mathbb{E}_{\boldsymbol{s}' \sim p(\cdot \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})} \left[v(\boldsymbol{s}', \boldsymbol{x}) \right] \right\}$, then v is subdifferentiable and $\partial_{\boldsymbol{x}} v(\boldsymbol{s}, \hat{\boldsymbol{x}}) =$

$$\operatorname{conv}\left(\bigcup_{\boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}) \in \mathcal{Y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})} \bigcup_{\lambda_k^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})) \in \Lambda^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}}))} \left\{ \nabla_{\boldsymbol{x}} r\left(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})\right) + \gamma \nabla_{\boldsymbol{x}} \mathop{\mathbb{E}}_{\boldsymbol{S}' \sim p(\cdot \mid \boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}))} \left[v(\boldsymbol{S}', \hat{\boldsymbol{x}}) \right] \right.\right.$$
$$\left.\left. + \sum_{k=1}^{d} \lambda_k^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})) \nabla_{\boldsymbol{x}} g_k\left(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})\right) \right\}\right). \tag{56}$$

[6] The interested reader can infer a proof from the proof of Theorem 2 of Atakan (2003)

Suppose additionally, that for all $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y}^*(\boldsymbol{s}, \boldsymbol{x}) \in \mathcal{Y}^*(\boldsymbol{s}, \boldsymbol{x})$ $\nabla_{\boldsymbol{x}} p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}^*(\boldsymbol{s}, \boldsymbol{x})) > 0$*, then* $\partial_{\boldsymbol{x}} v(\boldsymbol{s}, \hat{\boldsymbol{x}}) =$

$$\operatorname{conv}\left(\bigcup_{\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})\in\mathcal{Y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})}\ \bigcup_{\lambda_k^*(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))\in\Lambda^*(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\widehat{\boldsymbol{x}}))} \left\{\nabla_{\boldsymbol{x}} r\left(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})\right) + \gamma \mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot\mid \boldsymbol{s},\hat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))}\left[\nabla_{\boldsymbol{x}} v(\boldsymbol{S}', \hat{\boldsymbol{x}})\right]\right.$$
$$\left.+\gamma \mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot\mid \boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))}\left[v(\boldsymbol{S}', \hat{\boldsymbol{x}})\nabla_{\boldsymbol{x}} \log\left(p(\boldsymbol{S}' \mid \boldsymbol{s}, \hat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}))\right)\right] + \sum_{k=1}^{d} \lambda_k^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}))\nabla_{\boldsymbol{x}} g_k\left(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})\right)\right\}\right) . \tag{57}$$

where conv *is the convex hull operator and* $\boldsymbol{\lambda}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, , \widehat{\boldsymbol{x}})) = (\lambda_1^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})), \ldots, \lambda_d^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})))^T \in \Lambda^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}))$ *are the optimal KKT multipliers associated with* $\boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}) \in \mathcal{Y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})$.

Proof of Theorem C.3. First, note that by Lemma C.2, v is continuous and convex in $\boldsymbol{x}$, hence it is subdifferentiable. From Theorem C.1, we then obtain the first part of the theorem:

$$\partial_{\boldsymbol{x}} v(\boldsymbol{s}, \hat{\boldsymbol{x}}) = \operatorname{conv}\left(\bigcup_{\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})\in\mathcal{Y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})}\ \bigcup_{\lambda_k^*(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))\in\Lambda^*(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))} \left\{\nabla_{\boldsymbol{x}} r\left(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})\right) + \gamma \nabla_{\boldsymbol{x}} \mathop{\mathbb{E}}_{\boldsymbol{s}'\sim p(\cdot\mid \boldsymbol{s},\boldsymbol{x},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))}\left[v(\boldsymbol{S}', \boldsymbol{x})\right]\right.$$
$$\left.+\sum_{k=1}^{d} \lambda_k^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}}))\nabla_{\boldsymbol{x}} g_k\left(\boldsymbol{s}, \widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\boldsymbol{s}, \widehat{\boldsymbol{x}})\right)\right\}\right) . \tag{58}$$

By the Leibniz integral rule Flanders (1973), the gradient of the expectation can instead be expressed as an expectation of the gradient under continuity of the function whose expectation is taken, in this case v. In particular, if for all $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}, \boldsymbol{x} \in \mathcal{X}$, $\boldsymbol{y}^*(\boldsymbol{s}, \boldsymbol{x}) \in \mathcal{Y}^*(\boldsymbol{s}, \boldsymbol{x})$ $\nabla_{\boldsymbol{x}} p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}^*(\boldsymbol{s}, \boldsymbol{x})) > 0$ we have:

$$\nabla_{\boldsymbol{x}} \mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot\mid \boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}\left[v(\boldsymbol{S}', \boldsymbol{x})\right] \tag{59}$$
$$= \nabla_{\boldsymbol{x}} \int_{\boldsymbol{z}\in\mathcal{S}} p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) v(\boldsymbol{z}, \boldsymbol{x}) d\boldsymbol{z} \tag{60}$$
$$= \int_{\boldsymbol{z}\in\mathcal{S}} \nabla_{\boldsymbol{x}}[p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) v(\boldsymbol{z}, \boldsymbol{x})] d\boldsymbol{z} \quad \text{(Leibniz Integral Rule)} \tag{61}$$
$$= \int_{\boldsymbol{z}\in\mathcal{S}} \left[p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \nabla_{\boldsymbol{x}} v(\boldsymbol{z}, \boldsymbol{x}) + v(\boldsymbol{z}, \boldsymbol{x}) \nabla_{\boldsymbol{x}} p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})\right] d\boldsymbol{z} \quad \text{(Product Rule)} \tag{62}$$
$$= \int_{\boldsymbol{z}\in\mathcal{S}} \left[p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \nabla_{\boldsymbol{x}} v(\boldsymbol{z}, \boldsymbol{x}) + v(\boldsymbol{z}, \boldsymbol{x}) p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \frac{\nabla_{\boldsymbol{x}} p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})}\right] d\boldsymbol{z} \quad (p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) > 0) \tag{63}$$
$$= \int_{\boldsymbol{z}\in\mathcal{S}} \left[p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \nabla_{\boldsymbol{x}} v(\boldsymbol{z}, \boldsymbol{x}) + v(\boldsymbol{z}, \boldsymbol{x}) p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \nabla_{\boldsymbol{x}} \log p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})\right] d\boldsymbol{z} \tag{64}$$
$$= \int_{\boldsymbol{z}\in\mathcal{S}} \left[p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \nabla_{\boldsymbol{x}} v(\boldsymbol{z}, \boldsymbol{x})\right] d\boldsymbol{z} + \int_{\boldsymbol{z}\in\mathcal{S}} \left[v(\boldsymbol{z}, \boldsymbol{x}) p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \nabla_{\boldsymbol{x}} \log p(\boldsymbol{z} \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})\right] d\boldsymbol{z} \tag{65}$$
$$= \mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot\mid \boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}\left[\nabla_{\boldsymbol{x}} v(\boldsymbol{S}', \boldsymbol{x})\right] + \mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot\mid \boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}\left[v(\boldsymbol{S}', \boldsymbol{x}) \nabla_{\boldsymbol{x}} \log p(\boldsymbol{S}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})\right] \tag{66}$$

This gives us $\partial_{\boldsymbol{x}} v(\boldsymbol{s}, \hat{\boldsymbol{x}}) =$

$$\operatorname{conv}\left(\bigcup_{\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})\in\mathcal{Y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})}\bigcup_{\lambda_k^*(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))\in\Lambda^*(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\widehat{\boldsymbol{x}}))}\left\{\nabla_{\boldsymbol{x}} r\left(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})\right)+\gamma\mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))}\left[\nabla_{\boldsymbol{x}} v(\boldsymbol{S}',\hat{\boldsymbol{x}})\right]\right.\right.$$
$$\left.\left.+\gamma\mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))}\left[v(\boldsymbol{S}',\hat{\boldsymbol{x}})\nabla_{\boldsymbol{x}}\log\left(p(\boldsymbol{S}'\mid\boldsymbol{s},\hat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))\right)\right]+\sum_{k=1}^{d}\lambda_k^*(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))\nabla_{\boldsymbol{x}} g_k\left(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})\right)\right\}\right)\quad. \tag{67}$$

□

Note that in the special case that the probability transition function is representing a deterministic recursive parametrized optimization problem, $v(\boldsymbol{s},\boldsymbol{x}) = \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}}\{r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})+\gamma\left[v(\tau(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}),\boldsymbol{x})\right]\}$ i.e., $p(\boldsymbol{s}'\mid\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\in\{0,1\}$ for all $\boldsymbol{s},\boldsymbol{s}'\in\mathcal{S},\boldsymbol{x}\in\mathcal{X},\boldsymbol{y}\in\mathcal{Y}$, and $\tau:\mathcal{S}\times\mathcal{X}\times\mathcal{Y}\to\mathcal{S}$ is such that $\tau(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})=\boldsymbol{s}'$ iff $p(\boldsymbol{s}'\mid\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})=1$, the CSD convexity assumption reduces to the linearity of the deterministic state transition function τ (Proposition 1 of Atakan (2003)). In this case, the subdifferential of the Bellman equation reduces to

$$\partial_{\boldsymbol{x}} v(\boldsymbol{s},\hat{\boldsymbol{x}}) = \operatorname{conv}\left(\bigcup_{\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})\in\mathcal{Y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})}\bigcup_{\lambda_k^*(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))\in\Lambda^*(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))}\left\{\nabla_{\boldsymbol{x}} r\left(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})\right)+\gamma\nabla_{\boldsymbol{x}}\tau(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y})\nabla_{\boldsymbol{s}} v(\tau(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}),\widehat{\boldsymbol{x}})\right.\right. \tag{68}$$
$$\left.\left.+\nabla_{\boldsymbol{x}} v(\tau(\boldsymbol{s},\hat{\boldsymbol{x}}),\widehat{\boldsymbol{x}})+\sum_{k=1}^{d}\lambda_k^*(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}}))\nabla_{\boldsymbol{x}} g_k\left(\boldsymbol{s},\widehat{\boldsymbol{x}},\boldsymbol{y}^*(\boldsymbol{s},\widehat{\boldsymbol{x}})\right)\right\}\right) \tag{69}$$

Theorem 3.1. *Consider a zero-sum stochastic Stackelberg game $\mathcal{G}$, where $\mathcal{X}=\{\boldsymbol{x}\in\mathbb{R}^n\mid q_1(\boldsymbol{x})\leq 0,\ldots,q_p(\boldsymbol{x})\leq 0\}$ and $\mathcal{Y}=\{\boldsymbol{y}\in\mathbb{R}^m\mid r_1(\boldsymbol{y})\geq 0,\ldots,r_l(\boldsymbol{y})\geq 0\}$. Let $\mathcal{L}_{\boldsymbol{s},\boldsymbol{x}}(\boldsymbol{y},\boldsymbol{\lambda}) = r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})+\gamma\,\mathbb{E}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}\left[v(\boldsymbol{S}',\boldsymbol{x})\right]+\sum_{k=1}^{d}\lambda_k g_k(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$. Suppose that Assumption 1.1 holds, and that 1. for all $\boldsymbol{s}\in\mathcal{S},\boldsymbol{y}\in\mathcal{Y}$, $r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$, $g_1(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}),\ldots,g_d(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$ are concave in $\boldsymbol{x}$, 2. $\nabla_{\boldsymbol{x}} r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}),\nabla_{\boldsymbol{x}} g_1(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}),\ldots,\nabla_{\boldsymbol{x}} g_d(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$, $\nabla_{\boldsymbol{y}} r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}),\nabla_{\boldsymbol{y}} g_1(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}),\ldots,\nabla_{\boldsymbol{y}} g_d(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$ exist, for all $\boldsymbol{s}\in\mathcal{S},\boldsymbol{x}\in\mathcal{X},\boldsymbol{y}\in\mathcal{Y}$, 4. $p(\boldsymbol{s}'\mid\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$ is continuous CSD convex and differentiable in $(\boldsymbol{x},\boldsymbol{y})$, and 5. Slater's condition holds, i.e., $\forall\boldsymbol{s}\in\mathcal{S},\boldsymbol{x}\in\mathcal{X},\exists\widehat{\boldsymbol{y}}\in\mathcal{Y}$ s.t. $g_k(\boldsymbol{s},\boldsymbol{x},\widehat{\boldsymbol{y}})>0$, for all $k=1,\ldots,d$ and $r_j(\widehat{\boldsymbol{y}})>0$, for all $j=1,\ldots,l$, and $\exists\boldsymbol{x}\in\mathbb{R}^n$ s.t. $q_k(\boldsymbol{x})<0$ for all $k=1\ldots,p$. Then, there exists $\boldsymbol{\mu}^*:\mathcal{S}\to\mathbb{R}^p_+$, $\boldsymbol{\lambda}^*:\mathcal{S}\times\mathcal{X}\to\mathbb{R}^d_+$, and $\boldsymbol{\nu}^*:\mathcal{S}\times\mathcal{X}\to\mathbb{R}^l_+$ s.t. a policy profile $(\boldsymbol{\pi}^*_{\boldsymbol{x}},\boldsymbol{\pi}^*_{\boldsymbol{y}})\in\mathcal{X}^{\mathcal{S}}\times\mathcal{Y}^{\mathcal{S}}$ is a recSE of $\mathcal{G}$ only if it satisfies the following conditions, for all $\boldsymbol{s}\in\mathcal{S}$:*

$$\nabla_{\boldsymbol{x}}\mathcal{L}_{\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})}(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}),\boldsymbol{\lambda}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))) + \sum_{k=1}^{p}\mu_k^*(\boldsymbol{s})\nabla_{\boldsymbol{x}} q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))=0 \tag{1}$$

$$\nabla_{\boldsymbol{y}}\mathcal{L}_{\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})}(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}),\boldsymbol{\lambda}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))) + \sum_{k=1}^{l}\nu_k^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))\nabla_{\boldsymbol{x}} r_k(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}))=0 \tag{2}$$

$$\mu_k^*(\boldsymbol{s})q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))=0 \qquad q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))\leq 0 \quad \forall k\in[p] \tag{3}$$

$$g_k(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}),\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}))\geq 0 \qquad \lambda_k^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))g_k(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}),\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}))=0 \quad \forall k\in[d] \tag{4}$$

$$\nu_k^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))\nabla_{\boldsymbol{x}} r_k(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}))=0 \qquad r_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))\geq 0 \quad \forall k\in[l] \tag{5}$$

Proof of Theorem 3.1. By Theorem 2.2 and Theorem 2.4 we know that $(\boldsymbol{\pi}^*_{\boldsymbol{x}},\boldsymbol{\pi}^*_{\boldsymbol{y}})$ is a recursive Stackelberg equilibrium iff

$$v^{\boldsymbol{\pi}_{\boldsymbol{y}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s}) = \left(Cv^{\boldsymbol{\pi}_{\boldsymbol{y}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}\right)(\boldsymbol{s})\quad. \tag{70}$$

Note that for any policy profile $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}})$ that satisfies $v^{\boldsymbol{\pi}_{\boldsymbol{y}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s}) = \left(Cv^{\boldsymbol{\pi}_{\boldsymbol{y}}\boldsymbol{\pi}^*_{\boldsymbol{y}}}\right)(\boldsymbol{s})$ by definition of C we have that $(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}))$ is a Stackelberg equilibrium of

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} \left\{ r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) + \gamma \mathop{\mathbb{E}}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})} \left[v(\boldsymbol{S}')\right] \right\}$$

for all $\boldsymbol{s} \in \mathcal{S}$.

Fix a state $\boldsymbol{s} \in \mathcal{S}$, under the assumptions of the theorem, the conditions of Theorem C.3 are satisfied and $u^*(\boldsymbol{s},\boldsymbol{x}) = \max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} \left\{r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) + \gamma\, \mathbb{E}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}\left[v(\boldsymbol{S}')\right]\right\}$ is subdifferentiable in $\boldsymbol{x}$. Since $u^*(\boldsymbol{s},\boldsymbol{x})$ is convex in $\boldsymbol{x}$, and Slater's condition are satisfied by the assumptions of the theorem, the necesssary *and sufficient* conditions for $\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})$ to be an optimal solution to $\min_{\boldsymbol{x}\in\mathcal{X}} u^*(\boldsymbol{s},\boldsymbol{x})$ are given by the KKT conditions Kuhn & Tucker (1951) for $\min_{\boldsymbol{x}\in\mathcal{X}} u^*(\boldsymbol{s},\boldsymbol{x})$. Note that we can state the first order KKT conditions explicitly thanks to the subdifferential Benveniste-Scheinkman theorem (Theorem C.3). That is, $\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})$ is an optimal solution to $\min_{\boldsymbol{x}\in\mathcal{X}} u^*(\boldsymbol{s},\boldsymbol{x})$ if there exists $\boldsymbol{\mu}^*(\boldsymbol{s}) \in \mathbb{R}^p_+$ such that:

$$\nabla_{\boldsymbol{x}}\mathcal{L}_{\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})}(\boldsymbol{y}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})), \boldsymbol{\lambda}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}),\boldsymbol{y}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))) + \sum_{k=1}^{p} \mu^*_k(\boldsymbol{s})\nabla_{\boldsymbol{x}} q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) = 0 \tag{71}$$

$$\mu^*_k(\boldsymbol{s}) q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) = 0 \qquad \forall k \in [p] \tag{72}$$

$$q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \leq 0 \qquad \forall k \in [p] \tag{73}$$

where $\boldsymbol{y}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \in \arg\max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}),\boldsymbol{y})\geq\boldsymbol{0}} \left\{r(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}),\boldsymbol{y}) + \gamma\, \mathbb{E}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}),\boldsymbol{y})}\left[v(\boldsymbol{S}')\right]\right\}$ and
$\boldsymbol{\lambda}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}),\boldsymbol{y}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))) = (\lambda^*_1(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}),\boldsymbol{y}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))), \ldots, \lambda^*_d(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}),\boldsymbol{y}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))))^T \in \Lambda^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}),\boldsymbol{y}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})))$ are the optimal KKT multipliers associated with $\boldsymbol{y}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \in \mathcal{Y}^*(\boldsymbol{s},\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))$ which are guaranteed to exist since Slater's condition is satisfied for $\max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} \left\{r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) + \gamma\, \mathbb{E}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}\left[v(\boldsymbol{S}')\right]\right\}$.

Similarly, fix a state $\boldsymbol{s} \in \mathcal{S}$ and action for the outer player $\boldsymbol{x} \in \mathcal{X}$, since Slater's condition is satisfied for $\max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} \left\{r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) + \gamma\, \mathbb{E}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}\left[v(\boldsymbol{S}')\right]\right\}$, the necessary conditions for $\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})$ to be a Stackelberg equilibrium strategy for the inner player at state $\boldsymbol{s}$ are given by the KKT conditions for $\max_{\boldsymbol{y}\in\mathcal{Y}:\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})\geq\boldsymbol{0}} \left\{r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) + \gamma\, \mathbb{E}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}\left[v(\boldsymbol{S}')\right]\right\}$. That is, there exists $\boldsymbol{\lambda}^*(\boldsymbol{s},\boldsymbol{x}) \in \mathbb{R}^d_+$ and $\boldsymbol{\nu}^*(\boldsymbol{s},\boldsymbol{x}) \in \mathbb{R}^l_+$ such that:

$$\nabla_{\boldsymbol{y}}\mathcal{L}_{\boldsymbol{s},\boldsymbol{x}}(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}), \boldsymbol{\lambda}^*(\boldsymbol{s},\boldsymbol{x})) + \sum_{k=1}^{l} \nu^*_k(\boldsymbol{s})\nabla_{\boldsymbol{x}} r_k(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \tag{74}$$

$$g_k(\boldsymbol{s},\boldsymbol{x},\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) \geq 0 \qquad \forall k \in [d] \tag{75}$$

$$\lambda^*_k(\boldsymbol{s},\boldsymbol{x}) g_k(\boldsymbol{s},\boldsymbol{x},\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \qquad \forall k \in [d] \tag{76}$$

$$\nu^*_k(\boldsymbol{s},\boldsymbol{x})\nabla_{\boldsymbol{x}} r_k(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \qquad \forall k \in [l] \tag{77}$$

$$r_k(\boldsymbol{x}) \geq 0 \qquad \forall k \in [l] \tag{78}$$

Combining the necessary and *sufficient* conditions in Equations (71) to (73) with the necessary conditions in Equations (74) to (78), we obtain the necessary conditions for $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}})$ to be a recursive Stackelberg equilibrium.

□

Theorem C.4 (Recursive Stackelberg Equilibrium Necessary and Sufficient Optimality Conditions)**.** *Consider a zero-sum Stochastic Stackelberg game* $(\mathcal{S},\mathcal{X},\mathcal{Y},r,\boldsymbol{g},p,\gamma)$*, where* $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid q_1(\boldsymbol{x}) \leq 0, \ldots, q_p(\boldsymbol{x}) \leq 0\}$ *and* $\mathcal{Y} = \{\boldsymbol{y} \in \mathbb{R}^m \mid r_1(\boldsymbol{y}) \geq 0, \ldots, r_l(\boldsymbol{y}) \geq 0\}$*. Let* $\mathcal{L}_{\boldsymbol{s},\boldsymbol{x}}(\boldsymbol{y},\boldsymbol{\lambda}) = r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}) + \gamma\, \mathbb{E}_{\boldsymbol{S}'\sim p(\cdot|\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})}\left[v(\boldsymbol{S}',\boldsymbol{x})\right] + \sum_{k=1}^{d} \lambda_k g_k(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$*. Suppose that Assumption 1.1 holds, and that 1. for all* $\boldsymbol{s} \in \mathcal{S}, \boldsymbol{y} \in \mathcal{Y}$*,* $r(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$*,* $g_1(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y}), \ldots, g_d(\boldsymbol{s},\boldsymbol{x},\boldsymbol{y})$

are convex-concave in $(\boldsymbol{x}, \boldsymbol{y})$, *2.* $\nabla_{\boldsymbol{x}} r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \nabla_{\boldsymbol{x}} g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \ldots, \nabla_{\boldsymbol{x}} g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$, $\nabla_{\boldsymbol{y}} r(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \nabla_{\boldsymbol{y}} g_1(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}), \ldots, \nabla_{\boldsymbol{y}} g_d(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ *exist for all* $\boldsymbol{s} \in \mathcal{S}, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \in \mathcal{Y}$, *4.* $p(\boldsymbol{s}' \mid \boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y})$ *is continuous, CSD convex in* $\boldsymbol{x}$, *CSD concave in* $\boldsymbol{y}$, *and differentiable in* $(\boldsymbol{x}, \boldsymbol{y})$, *and 5. Slater's condition holds, i.e.,* $\forall \boldsymbol{s} \in \mathcal{S}, \boldsymbol{x} \in \mathcal{X}, \exists \widehat{\boldsymbol{y}} \in \mathcal{Y}$ *s.t.* $g_k(\boldsymbol{s}, \boldsymbol{x}, \widehat{\boldsymbol{y}}) > 0$, *for all* $k = 1, \ldots, d$ *and* $r_j(\widehat{\boldsymbol{y}}) > 0$ *for all* $j = 1, \ldots, l$, *and* $\exists \boldsymbol{x} \in \mathbb{R}^n$ *s.t.* $q_k(\boldsymbol{x}) < 0$ *for all* $k = 1 \ldots, p$ *. Then, there exists* $\boldsymbol{\mu}^* : \mathcal{S} \to \mathbb{R}^p_+$, $\boldsymbol{\lambda}^* : \mathcal{S} \times \mathcal{X} \to \mathbb{R}^d_+$, *and* $\boldsymbol{\nu}^* : \mathcal{S} \times \mathcal{X} \to \mathbb{R}^l_+$ *such that a policy profile* $(\boldsymbol{\pi}^*_{\boldsymbol{x}}, \boldsymbol{\pi}^*_{\boldsymbol{y}}) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$ *is a recursive Stackelberg equilibrium of* $(\mathcal{S}, \mathcal{X}, \mathcal{Y}, r, \boldsymbol{g}, p, \gamma)$ *only if it satisfies the following conditions for all* $\boldsymbol{s} \in \mathcal{S}$*:*

$$v^{\boldsymbol{\pi}^*_{\boldsymbol{y}} \boldsymbol{\pi}^*_{\boldsymbol{y}}}(\boldsymbol{s}) = \left(C v^{\boldsymbol{\pi}^*_{\boldsymbol{y}} \boldsymbol{\pi}^*_{\boldsymbol{y}}}\right)(\boldsymbol{s}) \tag{79}$$

$$\nabla_{\boldsymbol{x}} \mathcal{L}_{\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})}(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}), \boldsymbol{\lambda}^*(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))) + \sum_{k=1}^{p} \mu^*_k(\boldsymbol{s}) \nabla_{\boldsymbol{x}} q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) = 0 \tag{80}$$

$$\nabla_{\boldsymbol{y}} \mathcal{L}_{\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})}(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s}), \boldsymbol{\lambda}^*(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}))) + \sum_{k=1}^{l} \nu^*_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \nabla_{\boldsymbol{x}} r_k(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \tag{81}$$

$$\mu^*_k(\boldsymbol{s}) q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) = 0 \qquad \forall k \in [p] \tag{82}$$

$$q_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \leq 0 \qquad \forall k \in [p] \tag{83}$$

$$g_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) \geq 0 \qquad \forall k \in [d] \tag{84}$$

$$\lambda^*_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) g_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \qquad \forall k \in [d] \tag{85}$$

$$\nu^*_k(\boldsymbol{s}, \boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \nabla_{\boldsymbol{x}} r_k(\boldsymbol{\pi}^*_{\boldsymbol{y}}(\boldsymbol{s})) = 0 \qquad \forall k \in [l] \tag{86}$$

$$r_k(\boldsymbol{\pi}^*_{\boldsymbol{x}}(\boldsymbol{s})) \geq 0 \qquad \forall k \in [l] \tag{87}$$

D Omitted Results and Proofs Section 4

Before, we introduce the stochastic Stackelberg game whose recursive Stackelberg equilibria correspond to recursive competitive equilibria of an associate stochastic Fisher market, we introduce the following technical lemma, which provides the necessary and sufficient conditions for an allocation and saving system of a buyer to be expected utility maximizing.

Lemma D.1. *For any price system* $\boldsymbol{p} \in \mathbb{R}^{\mathcal{S} \times m}_+$, *a tuple* $(\boldsymbol{x}^*_i, \beta^*_i) \in \mathbb{R}^{\mathcal{S} \times n \times m}_+ \times \mathbb{R}^{n \times m}_+$ *consisting of an allocation system and saving system for a buyer* i, *given by a continuous, and homogeneous utility function* $u_i : \mathbb{R}^m_+ \times \mathcal{T} \to \mathbb{R}$ *representing a locally non-satiated preference is expected utility maximizing constrained by the saving and spending constrained, i.e., the first condition of a competitive equilibrium is satisfied only we have for all states* $\boldsymbol{s} \in \mathcal{S}$, $\boldsymbol{x}^*_i(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) \cdot \boldsymbol{p}(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) + \beta^*_i(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) \leq b_i$, *and,*

$$x^*_{ij}(\boldsymbol{s}) > 0 \implies \frac{\frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}^*_i(\boldsymbol{s}); t_i)}{p_j} = \frac{u_i(\boldsymbol{x}^*_i(\boldsymbol{s}); t_i)}{b_i - \beta^*_i(\boldsymbol{s})} \qquad \forall j \in [m] \tag{88}$$

$$\beta^*_i(\boldsymbol{s}) > 0 \implies \frac{\partial \nu_i}{\partial b_i}(\boldsymbol{s}) = \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} \left[\nu_i(\boldsymbol{t}', \boldsymbol{b}' + \beta^*_i(\boldsymbol{s}), \boldsymbol{q}')\right] \tag{89}$$

If additionally, u_i *is concave, then the above condition become also sufficient.*

Proof of Lemma D.1. Let $\mathcal{L}(\boldsymbol{x}_i, \beta_i, \lambda, \boldsymbol{\mu}; \boldsymbol{p}) = u_i(\boldsymbol{x}_i; t_i) + \gamma \mathbb{E}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}') \sim p(\cdot \mid \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}, (\boldsymbol{x}_i, \boldsymbol{X}^*_{-i}(\boldsymbol{s}), (\beta_i, \boldsymbol{\beta}^*_{-i}(\boldsymbol{s}))))} \left[\nu_i(\boldsymbol{t}', \boldsymbol{b}' + \beta_i, \boldsymbol{q}')\right] + \lambda(b_i - \boldsymbol{x}_i \boldsymbol{p}) + \sum_{j \in [m]} \mu_j x_{ij} + \mu_{m+1} \beta_i$ be the Lagrangian associated with

$$\nu_i(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) = \max_{(\boldsymbol{x}_i, \beta_i) \in \mathbb{R}^{m+1}_+ : \boldsymbol{x}_i \cdot \boldsymbol{p}^*(\boldsymbol{b}) + \beta_i \leq b_i} u_i(\boldsymbol{x}_i, t_i) + \gamma \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} \left[\nu_i(\boldsymbol{t}', \boldsymbol{b}' + \beta_i, \boldsymbol{q}')\right] \ .$$

Then, for any $b_i > 0$, Slater's condition holds and as the objective is concave (by the Weierstass M-test and the uniform limit theorem), the KKT first order necessary and sufficient optimality conditions for an allocation $\boldsymbol{x}^*_i \in \mathbb{R}^m_+$, saving $\beta^*_i \in \mathbb{R}_+$ and associated Lagrangian multipliers $\lambda^* \in \mathbb{R}_+$, $\boldsymbol{\mu}^* \in \mathbb{R}^{m+1}$ to be optimal for any prices $\boldsymbol{p} \in \mathbb{R}^m_+$ are given by the following pair of

equations:

$$\frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i) - \lambda^*(\boldsymbol{s}) p_j(\boldsymbol{s}) + \mu_j^* \doteq 0 \tag{90}$$

$$\gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} [\nu_i(\boldsymbol{t}', \boldsymbol{b}' + \beta_i^*, \boldsymbol{q}')] - \lambda_i^*(\boldsymbol{s}) + \mu_{m+1}^* \doteq 0 \tag{91}$$

Additionally, by the KKT complimentarity conditions, we have for all $j \in j$, $\mu_j^* x_{ij}^* = 0$ and $\mu_{m+1}^* \beta_i^* = 0$, which gives us:

$$x_{ij}^*(\boldsymbol{s}) > 0 \implies \lambda^* = \frac{\frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)}{p_j} \qquad \forall i \in [n], j \in [m] \tag{92}$$

$$\beta_i^*(\boldsymbol{s}) > 0 \implies \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} [\nu_i(\boldsymbol{t}', \boldsymbol{b}' + \beta_i^*, \boldsymbol{q}')] - \lambda^*(\boldsymbol{s}) = 0 \qquad \forall i \in [n], j \in [m] \tag{93}$$

Re-organizing expressions, yields:

$$x_{ij}^*(\boldsymbol{s}) > 0 \implies \lambda^* = \frac{\frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)}{p_j} \qquad \forall i \in [n], j \in [m] \tag{94}$$

$$\beta_i^*(\boldsymbol{s}) > 0 \implies \lambda^*(\boldsymbol{s}) = \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} [\nu_i(\boldsymbol{t}', \boldsymbol{b}' + \beta_i^*, \boldsymbol{q}')] \qquad \forall i \in [n] \tag{95}$$

Using the envelope theorem, we can also compute $\frac{\partial \nu_i}{\partial \beta_i}(\boldsymbol{s})$ as follows:

$$\frac{\partial \nu_i}{\partial b_i}(\boldsymbol{s}) = \lambda^*(b_i) \tag{96}$$

Hence, combining the above with Equation (94) and Equation (95), we get:

$$x_{ij}^*(\boldsymbol{s}) > 0 \implies \lambda^* = \frac{\frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)}{p_j} \qquad \forall i \in [n], j \in [m] \tag{97}$$

$$\beta_i^*(\boldsymbol{s}) > 0 \implies \frac{\partial \nu_i}{\partial b_i}(\boldsymbol{s}) = \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} [\nu_i(\boldsymbol{t}', \boldsymbol{b}' + \beta_i^*, \boldsymbol{q}')] \qquad \forall i \in [n] \tag{98}$$

Finally, going back to Equation (90), multiplying by $x_{ij}^*(\boldsymbol{s})$ and summing up across all $j \in [m]$, we obtain:

$$\sum_{j \in [m]} x_{ij}^*(\boldsymbol{s}) \frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i) - \lambda^*(\boldsymbol{s}) p_j(\boldsymbol{s}) x_{ij}^*(\boldsymbol{s}) + \mu_j^* x_{ij}^*(\boldsymbol{s}) = 0 \tag{99}$$

$$u_i(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i) - \lambda^*(\boldsymbol{s}) \sum_{j \in [m]} p_j(\boldsymbol{s}) x_{ij}^*(\boldsymbol{s}) + \mu_j^* x_{ij}^*(\boldsymbol{s}) = 0 \quad \text{(Euler's Theorem for Homogeneous Functions)} \tag{100}$$

$$u_i(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i) - \lambda^*(\boldsymbol{s})(b_i - \beta_i^*(\boldsymbol{s})) = 0 \quad \text{(Slack Complementarity)} \tag{101}$$

$$\lambda^*(\boldsymbol{s}) = \frac{u_i(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)}{b_i - \beta_i^*(\boldsymbol{s})} \tag{102}$$

Combining the above conditions, with Equation (112), and adding to it Equation (98), and ensuring that the KKT primal feasibility conditions hold as well, we obtain the following necessary and sufficient conditions that need to hold for all states $\boldsymbol{s} \in \mathcal{S}$:

$$\boldsymbol{x}_i^*(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) \cdot \boldsymbol{p}(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) + \beta_i^*(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) \leq b_i \tag{103}$$

$$x_{ij}^*(\boldsymbol{s}) > 0 \implies \frac{\frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)}{p_j} = \frac{u_i(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)}{b_i - \beta_i^*(\boldsymbol{s})} \qquad \forall j \in [m] \tag{104}$$

$$\beta_i^*(\boldsymbol{s}) > 0 \implies \frac{\partial \nu_i}{\partial b_i}(\boldsymbol{s}) = \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} [\nu_i(\boldsymbol{t}', \boldsymbol{b}' + \beta_i^*(\boldsymbol{s}), \boldsymbol{q}')] \tag{105}$$

□

Theorem 4.1. *A stochastic Fisher market with savings $(\mathcal{S}, \mathcal{U}, \boldsymbol{b}^{(0)}, p, \gamma)$ in which $\mathcal{U}$ is a vector of continuous and homogeneous utility functions has at least one recCE. Additionally, the recSE $(\boldsymbol{p}^*, \boldsymbol{X}^*, \boldsymbol{\beta}^*)$ that solves the following Bellman equation corresponds to the recCE of $(\mathcal{S}, \mathcal{U}, \boldsymbol{b}^{(0)}, p, \gamma)$:*

$$v(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}) = \min_{\boldsymbol{p} \in \mathbb{R}_+^m} \max_{(\boldsymbol{X}, \boldsymbol{\beta}) \in \mathbb{R}_+^{n \times (m+1)} : \boldsymbol{X}\boldsymbol{p} + \boldsymbol{\beta} \leq \boldsymbol{b}} \sum_{j \in [m]} q_j p_j + \sum_{i \in [n]} (b_i - \beta_i) \log(u_i(\boldsymbol{x}_i, t_i)) + \gamma \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}') \sim p(\cdot | \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}, \boldsymbol{X}, \boldsymbol{\beta})} [v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}, \boldsymbol{q}')] \tag{6}$$

Proof of Theorem 4.1. Define $\mathcal{L}(\boldsymbol{p}, \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \sum_{j \in [m]} q_j p_j + \sum_{i \in [n]} (b_i - \beta_i) \log(u_i(\boldsymbol{x}_i, t_i)) + \gamma \mathbb{E}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}') \sim p(\cdot | \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}, \boldsymbol{X}, \boldsymbol{\beta})} [v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}, \boldsymbol{q}')] + \sum_{j \in [m]} \lambda_i (b_i - \boldsymbol{x}_i \cdot \boldsymbol{p} + \beta_i)$. By Theorem 3.1, the necessary optimality conditions for the stochastic Stackelberg game

$$\max_{(\boldsymbol{X}, \boldsymbol{\beta}) \in \mathbb{R}_+^{n \times m} \times \mathbb{R}_+^n : \boldsymbol{X}\boldsymbol{p} + \boldsymbol{\beta} \leq \boldsymbol{b}} \sum_{j \in [m]} q_j p_j + \sum_{i \in [n]} (b_i - \beta_i) \log(u_i(\boldsymbol{x}_i, t_i)) + \gamma \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}') \sim p(\cdot | \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}, \boldsymbol{X}, \boldsymbol{\beta})} [v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}, \boldsymbol{q}')]$$

are that for all states $\boldsymbol{s} \in \mathcal{S}$ there exists $\boldsymbol{\mu}^*(\boldsymbol{s}) \in \mathbb{R}^{n \times (m+1)}$, and $\boldsymbol{\nu}^*(\boldsymbol{s}) \in \mathbb{R}_+^{n \times (m+1)}$ associated with the non-negativity constraints for $(\boldsymbol{X}, \boldsymbol{\beta})$, and $\boldsymbol{p}$ respectively, and $\boldsymbol{\lambda}^*(\boldsymbol{s}) \in \mathbb{R}_+^m$ such that:

$$q_j - \sum_{i \in [n]} \lambda_i^*(\boldsymbol{s}) x_{ij}^*(\boldsymbol{s}) - \nu_j^*(\boldsymbol{s}) \doteq 0 \tag{106}$$

$$\frac{b_i - \beta_i^*(\boldsymbol{s})}{u_i(\boldsymbol{x}_i^*(\boldsymbol{s}))} \frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i) - \lambda_i^*(\boldsymbol{s}) p_j(\boldsymbol{s}) + \mu_{ij}^*(\boldsymbol{s}) \doteq 0 \quad \forall i \in [n], j \in [m] \tag{107}$$

$$-\log(u_i(\boldsymbol{x}_i^*(\boldsymbol{s}))) + \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} [v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}^*(\boldsymbol{s}), \boldsymbol{q}')] - \lambda_i^*(\boldsymbol{s}) + \mu_{i(m+1)}^*(\boldsymbol{s}) \doteq 0 \quad \forall i \in [n] \tag{108}$$

Note that by Theorem 3.1, we also have $\mu_{i(m+1)}^*(\boldsymbol{s}) \beta_i^*(\boldsymbol{s})) = \mu_{i+1)}^*(\boldsymbol{s}) x_{ij}^*(\boldsymbol{s}) = 0$ which gives us:

$$p_j(\boldsymbol{s}) > 0 \implies q_j - \sum_{i \in [n]} \lambda_i^*(\boldsymbol{s}) x_{ij}^*(\boldsymbol{s}) = 0 \quad \forall [m] \in [m] \tag{109}$$

$$x_{ij}^*(\boldsymbol{s}) > 0 \implies \frac{b_i - \beta_i^*(\boldsymbol{s})}{u_i(\boldsymbol{x}_i^*(\boldsymbol{s}))} \frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}_i^*(\boldsymbol{s})) - \lambda_i^*(\boldsymbol{s}) p_j(\boldsymbol{s}) = 0 \quad \forall i \in [n], j \in [m] \tag{110}$$

$$\beta_i^*(\boldsymbol{s}) > 0 \implies -\log(u_i(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)) + \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} [v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}^*(\boldsymbol{s}), \boldsymbol{q}')] - \lambda_i^*(\boldsymbol{s}) + \mu_{i(m+1)}^*(\boldsymbol{s}) = 0 \quad \forall i \in [n] \tag{111}$$

Re-organizing expressions, we obtain:

$$p_j(\boldsymbol{s}) > 0 \implies q_j = \sum_{i \in [n]} \lambda_i^*(\boldsymbol{s}) x_{ij}^*(\boldsymbol{s}) \quad \forall j \in [m] \tag{112}$$

$$x_{ij}^*(\boldsymbol{s}) > 0 \implies \frac{u_i(\boldsymbol{x}_i^*(\boldsymbol{s}))}{b_i - \beta_i^*(\boldsymbol{s})} \lambda_i^*(\boldsymbol{s}) = \frac{\frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)}{p_j(\boldsymbol{s})} \quad \forall i \in [n], j \in [m] \tag{113}$$

$$\beta_i^*(\boldsymbol{s}) > 0 \implies -\log(u_i(\boldsymbol{x}_i^*(\boldsymbol{s}))) + \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} [v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}^*(\boldsymbol{s}), \boldsymbol{q}')] - \lambda_i^*(\boldsymbol{s}) = 0 \quad \forall i \in [n] \tag{114}$$

Using the envelope theorem, we can compute $\frac{\partial v}{\partial b_i}$ as follows:

$$\frac{\partial v^*}{\partial b_i}(\boldsymbol{s}) = \log\left(u_i(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)\right) + \lambda_i^*(\boldsymbol{s}) \tag{115}$$

Re-organizing expressions, we get:

$$\lambda_i^*(\boldsymbol{s}) = \frac{\partial v}{\partial b_i}(\boldsymbol{s}) - \log\left(u_i(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)\right) \tag{116}$$

Combining Equation (116) and Equation (114), we obtain:

$$\beta_i^*(\boldsymbol{s}) > 0 \implies -\log\left(u_i(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)\right) + \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} \left[v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}^*(\boldsymbol{s}), \boldsymbol{q}')\right] - \frac{\partial v}{\partial b_i}(\boldsymbol{b}) + \log\left(u_i(\boldsymbol{x}_i^*(\boldsymbol{b}); t_i)\right) = 0 \quad \forall i \in [n] \tag{117}$$

$$\beta_i^*(\boldsymbol{s}) > 0 \implies \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} \left[v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}^*(\boldsymbol{s}), \boldsymbol{q}')\right] - \frac{\partial v}{\partial b_i}(\boldsymbol{b}) = 0 \quad \forall i \in [n] \tag{118}$$

$$\beta_i^*(\boldsymbol{s}) > 0 \implies \frac{\partial v}{\partial b_i}(\boldsymbol{b}) = \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} \left[v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}^*(\boldsymbol{s}), \boldsymbol{q}')\right] \quad \forall i \in [n] \tag{119}$$

Going back to Equation (107), multiplying both sides by $x_{ij}^*(\boldsymbol{s})$ and summing up across all $j \in [m]$, we get:

$$\sum_{j \in j} \frac{b_i - \beta_i^*(\boldsymbol{s})}{u_i(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)} \sum_{j \in [m]} x_{ij}(\boldsymbol{s})^* \frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}_i^*(\boldsymbol{s})) - \lambda_i^*(\boldsymbol{s}) \sum_{j \in [m]} p_j(\boldsymbol{s}) x_{ij}^*(\boldsymbol{s}) = 0 \tag{120}$$

$$\frac{b_i - \beta_i^*(\boldsymbol{s})}{u_i(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)} u_i(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i) - \lambda_i^*(\boldsymbol{s}) \sum_{j \in [m]} p_j(\boldsymbol{s}) x_{ij}^*(\boldsymbol{s}) = 0 \quad \text{(Euler's Theorem)} \tag{121}$$

$$b_i - \beta_i^*(\boldsymbol{s}) - \lambda_i^*(\boldsymbol{s}) \sum_{j \in [m]} p_j(\boldsymbol{s}) x_{ij}^*(\boldsymbol{s}) = 0 \tag{122}$$

By Theorem 3.1, we have that $\lambda_i^*(\boldsymbol{s}) \left(b_i - \sum_{j \in [m]} p_j(\boldsymbol{s}) x_{ij}^*(\boldsymbol{s}) - \beta_i^*(\boldsymbol{s})\right) = 0$, which gives us:

$$b_i - \beta_i^*(\boldsymbol{s}) - \lambda_i^*(\boldsymbol{s})(b_i - \beta_i^*(\boldsymbol{s})) = 0 \tag{123}$$

$$\lambda_i^*(\boldsymbol{s}) = 1 \tag{124}$$

Combining the above with Equations (112) to (114) we obtain:

$$p_j(\boldsymbol{s}) > 0 \implies q_j = \sum_{i \in [n]} x_{ij}^*(\boldsymbol{s}) \quad \forall j \in [m] \tag{125}$$

$$x_{ij}^*(\boldsymbol{s}) > 0 \implies \frac{u_i(\boldsymbol{x}_i^*(\boldsymbol{s}))}{b_i - \beta_i^*(\boldsymbol{s})} = \frac{\frac{\partial u_i}{\partial x_{ij}}(\boldsymbol{x}_i^*(\boldsymbol{s}); t_i)}{p_j(\boldsymbol{s})} \quad \forall i \in [n], j \in [m] \tag{126}$$

$$\beta_i^*(\boldsymbol{s}) > 0 \implies \frac{\partial v}{\partial b_i}(\boldsymbol{b}) = \gamma \frac{\partial}{\partial \beta_i} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{b}', \boldsymbol{q}')} \left[v(\boldsymbol{t}', \boldsymbol{b}' + \boldsymbol{\beta}^*(\boldsymbol{s}), \boldsymbol{q}')\right] \quad \forall i \in [n] \tag{127}$$

Since the utility functions are non-satiated, and by the second equation, the buyers are utility maximizing at state $\boldsymbol{s}$ over all allocations, we must also have that Walras' law holds, i.e., $\boldsymbol{p} \cdot \left(\boldsymbol{q} - \sum_{i \in [n]} \boldsymbol{x}_i\right) - \sum_{i \in [n]} \beta_i$. Walras' law combined with the first equation above then imply the second condition of a recursive competitive equilibrium. Finally, by Lemma D.1, the last two equations imply the first condition of recursive competitive equilibrium.

□

E Experiment Details

We initialized a stochastic Fisher market with $n = 2$ buyers and $m = 5$ goods. To simplify the analysis, we assumed deterministic transitions such that the buyers do not get new budgets at each time period, and their types/valuations as well as the supply of goods does not change at each state, i.e., the type/valuation space and supply space has cardinality 1. This reduced the market to a deterministic repeated market setting in which the amount of budget saved by the buyers differentiates different states of the market. To initialize the state space of the market, we first fixed a range of $[10, 50]^m$ for the buyers' valuations and drew for all buyers $i \in [n]$ valuations $\boldsymbol{\theta}_i$ from that range uniformly at random at the beginning of the experiment. We have assumed the supply of goods is $\mathbf{1}_m$ and that the budget space was $[0, 1]^n$. This means that our state space for our experiments was $\mathcal{S} = \{(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\} \times \{\mathbf{1}_m\} \times [0, 1]^n$. We note that although the assumption that buyers valuations/type space has cardinality one does simplify the problem, the supply of the goods being $\mathbf{1}$ at each state is wlog because goods are divisible and the allocation of goods to buyers at each state can then be interpreted as the percentage of a particular good allocated to a buyer. We assumed initial budgets of $\boldsymbol{b}^{(0)} = \mathbf{1}_n$ for both buyers.

Since the state space is continuous, the value function is also continuous in the stochastic Fisher market setting. As a result, we had to use fitted variant of value iteration. In particular, we assumed that the value function had a linear form at each state such that $v(\boldsymbol{t}, \boldsymbol{b}, \boldsymbol{q}; \boldsymbol{a}, c) = \boldsymbol{a}^T \boldsymbol{b} + c$ for some parameters $\boldsymbol{a} \in \mathbb{R}^n, c \in \mathbb{R}$, and we tried to approximate the value function at the next step of value iteration by using linear regression. That is, at each value iteration step, we uniformly sampled 25 budget vectors from the range $[0, 1]^n$. Next, for each sampled budget $\boldsymbol{b}$, we solved the min-max step given that budget as a state. This process gave us (budget, value) pairs on which we ran linear regression to approximate the value function at the next iterate.

To solve the generalized min-max operator at each step of value iteration, we used two different methods for comparison: **nested gradient descent ascent (GDA)** Goktas & Greenwald (2021) (Algorithm 2), which is not guaranteed to converge to a global optimum since the min-max Stackelberg game for stochastic Fisher markets is convex-non-concave and **max-oracle gradient descent** Goktas & Greenwald (2021) where the max-oracle is the simulated annealing algorithm Bertsimas & Tsitsiklis (1993), a metaheuristic which probabiliticly aims to find a global maximum. Although, simulated annealing annealing is not guaranteed to converge to a global maximum, we observed that it often performed better than nested GDA at finding a global optimum for the inner player's strategy. For both methods, we have run value iteration for 30 iterations. We ran nested GDA with learning rates $\eta_{\boldsymbol{X}} = 0.5, \eta_{\boldsymbol{p}} = 0.02$ for linear, $\eta_{\boldsymbol{X}} = 0.5, \eta_{\boldsymbol{p}} = 0.01$ for leontief, and $\eta_{\boldsymbol{X}} = 0.5$, $\eta_{\boldsymbol{p}} = 0.0005$ for Cobb-Douglas. We also ran max-oracle gradient descent with a learning rate of $\eta_{\boldsymbol{p}} = 0.5$. The outer loop of nested GDA was run for $T_{\boldsymbol{p}} = 30$ iterations, while its inner loop was run for $T_{\boldsymbol{X}} = 100$ iterations, and the max-oracle gradient descent algorithm was run for $T_{\boldsymbol{p}} = 30$ iterations. We depict the trajectory of the average value of the value function at each iteration of value iteration, under nested GDA in Section 5, and under max-oracle gradient descent in Section 5.

Utility Type	Cumulative Utility based on VI	Maximized cumulative Utility	Distance to Utility Maximization	Normalized Distance to Utility Maximization	Distance to Market Clearance (Average Excess Demands)
Linear (Nested GDA)	[96.11864217, 872.44457586]	[84.79659719559233, 90.3457392997638]	782.1807839960541	6.312670581380343	1.51360722680293
Linear (Max-oracle GD)	[58.83670238, 65.37844446]	[96.96171820101995, 106.36975094410113]	55.98038976844064	0.3889393929909976	2.1650851972578646
Leontief (Nested GDA)	[0.01241015, 0.01197241]	[0.009253314180356959, 0.008070244600082384]	0.005019216377258324	0.408792752480498	2.1885091150520655
Leontief (Max-oracle GD)	[0.0044322, 0.00609951]	[0.012505356366874554, 0.009612660130480 25]	0.008804435107563718	0.5581970083770711	2.178489762142351
Cobb-Douglas (Nested GDA)	[-759048.31311689, -844533.23523943]	[-5410.446761817995, -5757.897160340484]	1127614.3407095883	142.71734067443717	2.1892506104891303
Cobb-Douglas (Max-oracle GD)	[-763010.21531314 -847841.51430998]	[-5482.763563167677, -5790.554467234857]	1132650.7224766547	142.0356741620785	2.1697254289647336

Figure 3: Exploitability of the computed recursive competitive equilbirium by both methods.