

## A Pseudo-Code for Algorithms

---

### Algorithm 2 Value Iteration (with Min-Max Oracle)

---

**Inputs:**  $\mathcal{S}, \mathcal{X}, \mathcal{Y}, r, \mathbf{g}, p, \gamma, T$

**Outputs:**  $v^{(T)}$

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

---

### Algorithm 3 Nested GDA for stochastic Fisher Markets with saving

---

**Inputs:**  $v, T, \mathbf{b}, \mathbf{q}, \eta_p, \eta_{\mathbf{X}}, T_p, T_{\mathbf{X}}, \mathbf{p}^{(0)}, \mathbf{X}^{(0)}, \boldsymbol{\beta}^{(0)}$

**Output:**  $(\mathbf{p}^{(t)}, \mathbf{X}^{(t)})_{t=1}^T$

- 1: **for**  $t = 1, \dots, T_p$  **do**
  - 2:     **for**  $s = 1, \dots, T_{\mathbf{X}}$  **do**
  - 3:         For all  $i \in [n]$ ,  $\mathbf{x}_i^{(t)} = \mathbf{x}_i^{(t-1)} + \eta_{\mathbf{X}} \left( \frac{b_i - \beta_i^{(t)}}{u_i(\mathbf{x}_i^{(t-1)}; \mathbf{t}_i)} \nabla_{\mathbf{x}_i} u_i(\mathbf{x}_i^{(t-1)}; \mathbf{t}_i) \right)$
  - 4:         For all  $i \in [n]$ ,  $\beta_i^{(t)} = \beta_i^{(t-1)} + \eta_{\mathbf{X}} \left( -\log(u_i(\mathbf{x}_i^{(t-1)}; \mathbf{t}_i)) + \gamma \frac{\partial v(\mathbf{T}, \mathbf{b}, \mathbf{q})}{\partial b_i} \right)$
  - 5:          $(\mathbf{X}^{(t)}, \boldsymbol{\beta}^{(t)}) = \Pi_{\{(\mathbf{X}, \boldsymbol{\beta}) \in \mathbb{R}_+^{n \times m} \times \mathbb{R}_+^n : \mathbf{X} \cdot \mathbf{p}^{(t-1)} + \boldsymbol{\beta} \leq \mathbf{b}\}}((\mathbf{X}^{(t-1)}, \boldsymbol{\beta}^{(t-1)}))$
  - 6:          $\mathbf{p}^{(t)} = \Pi_{\mathbb{R}_+^m}(\mathbf{p}^{(t-1)} - \eta_p(\mathbf{1} - \sum_{i \in [n]} \mathbf{x}_i^{(t-1)}))$
  - 7: **return**  $(\mathbf{p}^{(t)}, \mathbf{X}^{(t)})_{t=1}^{T_p}$
- 

## 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 [66] and which follows from their definitions:

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

$$v(\mathbf{s}) = q(\mathbf{s}, \pi_{\mathbf{x}}(\mathbf{s}), \pi_{\mathbf{y}}(\mathbf{s})) \quad (14)$$

$$q(\mathbf{s}, \mathbf{x}, \mathbf{y}) = \mathbb{E}_{S' \sim p(\cdot | \mathbf{s}, \mathbf{x}, \mathbf{y})} [r(\mathbf{s}, \mathbf{x}, \mathbf{y}) + \gamma v(S')] \quad (15)$$

*Proof of Theorem B.1.* By the definition of the state value function we have  $v_i^{\pi_{\mathbf{x}} \pi_{\mathbf{y}}} = q_i(\mathbf{s}, \pi_{\mathbf{x}}(\mathbf{s}), \pi_{\mathbf{y}}(\mathbf{s}))$ , hence by Equation (14) we must have that  $v_i = v_i^{\pi_{\mathbf{x}} \pi_{\mathbf{y}}}$ . Additionally, by Equation (15) and the definition of the action-value functions this also implies that  $q_i(\mathbf{s}, \mathbf{x}, \mathbf{y}) = q_i^{\pi_{\mathbf{x}} \pi_{\mathbf{y}}}(\mathbf{s}, \mathbf{x}, \mathbf{y})$   $\square$

**Theorem 2.1.**  $(\pi_{\mathbf{x}}^*, \pi_{\mathbf{y}}^*)$  is a recSE of  $\mathcal{G}^{(0)}$  of  $v^{\pi_{\mathbf{x}} \pi_{\mathbf{y}}}$  iff it induces a value function which is a fixed point of  $C$ : i.e.,  $(\pi_{\mathbf{x}}^*, \pi_{\mathbf{y}}^*)$  is a Stackelberg equilibrium iff, for all  $\mathbf{s} \in \mathcal{S}$ ,  $(Cv^{\pi_{\mathbf{x}}^* \pi_{\mathbf{y}}^*})(\mathbf{s}) = v^{\pi_{\mathbf{x}}^* \pi_{\mathbf{y}}^*}(\mathbf{s})$ .

*Proof of Theorem 2.1.* We prove one direction, the other direction follows symmetrically. (Fixed Point  $\implies$  recursive Stackelberg equilibrium) Suppose that a value function  $v^{\pi_{\mathbf{x}}^* \pi_{\mathbf{y}}^*}$  which is induced

by a policy profile  $(\pi_x^*, \pi_y^*)$  is a fixed point of  $C$ , we then have for all states  $s \in \mathcal{S}$ :

$$v^{\pi_x^* \pi_y^*} = (Cv^{\pi_x^* \pi_y^*})(s) \quad (16)$$

$$= \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}: g(s, x, y) \geq 0} \mathbb{E}_{S' \sim p(\cdot | s, x, y)} \left[ r(s, x, y) + \gamma v^{\pi_x^* \pi_y^*}(S') \right] \quad (17)$$

$$= \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}: g(s, x, y) \geq 0} q^{\pi_x^* \pi_y^*}(s, x, y) \quad (18)$$

Hence, by Definition 1.2,  $(\pi_x^*, \pi_y^*)$  is recursive Stackelberg equilibrium.  $\square$

**Lemma 2.2.** Suppose that  $f, h : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ ,  $g : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^d$  are continuous functions, and  $\mathcal{X}, \mathcal{Y}$  are compact sets. Then  $|\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}: g(x, y) \geq 0} f(x, y) - \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}: g(x, y) \geq 0} h(x, y)| \leq \max_{(x, y) \in \mathcal{X} \times \mathcal{Y}} |f(x, y) - h(x, y)|$ .

*Proof of Lemma 2.2* Let  $(x^*, y^*)$  be a Stackelberg equilibrium of  $\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}: g(x, y) \geq 0} f(x, y)$ , and  $(x', y')$  be a Stackelberg equilibrium of  $\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}: g(x, y) \geq 0} h(x, y)$ . Additionally, let  $\bar{y} \in \arg \max_{y \in \mathcal{Y}: g(x', y) \geq 0} f(x', y)$ , then by the definition of a Stackelberg equilibrium, we have  $f(x^*, y^*) = \min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}: g(x, y) \geq 0} f(x, y) \leq \max_{y \in \mathcal{Y}: g(x', y) \geq 0} f(x', y) = f(x', \bar{y})$ , and  $h(x', y') = \max_{y \in \mathcal{Y}: g(x', y) \geq 0} h(x', y) \geq h(x', \bar{y})$ .

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

$$\left| \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}: g(x, y) \geq 0} f(x, y) - \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}: g(x, y) \geq 0} h(x, y) \right| \quad (19)$$

$$= |f(x^*, y^*) - h(x', y')| \quad (20)$$

$$\leq |f(x', \bar{y}) - h(x', y')| \quad (21)$$

$$\leq |f(x', \bar{y}) - h(x', \bar{y})| \quad (22)$$

$$\leq \max_{(x, y) \in \mathcal{X} \times \mathcal{Y}} |f(x, y) - h(x, y)| \quad (23)$$

The opposite case follows similarly by symmetry.  $\square$

**Theorem 2.3.** Consider the operator  $C$  associated with a stochastic Stackelberg game  $\mathcal{G}^{(0)}$ . Under Assumption 1.1,  $C$  is a contraction mapping w.r.t. to the sup norm  $\|\cdot\|_\infty$  with constant  $\gamma$ .

*Proof of Theorem 2.3* 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:

$$\|Cv - Cv'\|_\infty \quad (24)$$

$$\leq \max_{s \in \mathcal{S}} \left| \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}: g(s, x, y) \geq 0} q(s, x, y) - \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}: g(s, x, y) \geq 0} q'(s, x, y) \right| \quad (25)$$

$$\leq \max_{s \in \mathcal{S}} \max_{(x, y) \in \mathcal{X} \times \mathcal{Y}} |q(s, x, y) - q'(s, x, y)| \quad (\text{Lemma 2.2}) \quad (26)$$

$$\leq \max_{s \in \mathcal{S}} \max_{(x, y) \in \mathcal{X} \times \mathcal{Y}} \left| \mathbb{E}_{S' \sim p(\cdot | s, x, y)} [r(s, x, y) + \gamma v(S')] - \mathbb{E}_{S' \sim p(\cdot | s, x, y)} [r(s, x, y) + \gamma v'(S')] \right| \quad (27)$$

$$\leq \max_{s \in \mathcal{S}} \max_{(x, y) \in \mathcal{X} \times \mathcal{Y}} \left| \mathbb{E}_{S' \sim p(\cdot | s, x, y)} [\gamma v(S') - \gamma v'(S')] \right| \quad (28)$$

$$\leq \gamma \max_{s \in \mathcal{S}} \max_{(x, y) \in \mathcal{X} \times \mathcal{Y}} \left| \mathbb{E}_{S' \sim p(\cdot | s, x, y)} [v(S') - v'(S')] \right| \quad (29)$$

$$\leq \gamma \max_{s \in \mathcal{S}} \max_{(x, y) \in \mathcal{X} \times \mathcal{Y}} |v(s) - v'(s)| \quad (30)$$

$$= \gamma \|v - v'\|_\infty \quad (31)$$

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

**Theorem 2.4.** Consider a zero-sum stochastic Stackelberg game  $\mathcal{G}^{(0)}$ . Under Assumption 1.1  $\mathcal{G}^{(0)}$  has a unique value function  $v^{\pi_x^* \pi_y^*}$  associated with all recSE  $(\pi_x^*, \pi_y^*)$ , which can be computed by iteratively applying  $C$  to any initial state-value function  $v^{(0)} \in \mathcal{V}$ : i.e.,  $\lim_{t \rightarrow \infty} v^{(t)} = v^{\pi_x^* \pi_y^*}$ .

*Proof of Theorem 2.4* By combining Theorem 2.3 and the Banach fixed point theorem [57], we obtain that a fixed point of  $C$  exists. Hence, by Theorem 2.1 a recursive Stackelberg equilibrium of  $(\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mu^{(0)}, r, \mathbf{g}, p, \gamma)$  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  $\lim_{t \rightarrow \infty} v^{(t)} = v^{\pi_x^* \pi_y^*}$ . □

For any  $q \in \mathcal{Q}$ , we define a **greedy policy profile** with respect to  $q$  as a policy profile  $(\pi_x^q, \pi_y^q)$  such that  $\pi_x^q \in \arg \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbf{y} \in \mathcal{Y}: \mathbf{g}(\mathbf{s}, \mathbf{x}, \mathbf{y}) \geq 0} q(\mathbf{s}, \mathbf{x}, \mathbf{y})$  and  $\pi_y^q \in \arg \max_{\mathbf{y} \in \mathcal{Y}: \mathbf{g}(\mathbf{s}, \pi_x^q(\mathbf{x}), \mathbf{y}) \geq 0} q(\mathbf{s}, \pi_x^q(\mathbf{x}), \mathbf{y})$ . 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.

**Theorem 2.6** (Convergence of Value Iteration). Suppose value iteration is run on input  $\mathcal{G}$ . Let  $(\pi_x^*, \pi_y^*)$  be recSE of  $\mathcal{G}$  with value function  $v^{\pi_x^* \pi_y^*}$ . Under Assumption 1.1 if we initialize  $v^{(0)}(\mathbf{s}) = 0$ , for all  $\mathbf{s} \in \mathcal{S}$ , then for  $k \geq \frac{1}{1-\gamma} \log \frac{r_{\max}}{\epsilon(1-\gamma)}$ , it holds that  $v^{(k)}(\mathbf{s}) - v^{\pi_x^* \pi_y^*}(\mathbf{s}) \leq \epsilon$ .

*Proof of Theorem 2.6* First note that by Assumption 1.1 we have that  $\|v^{\pi_x^* \pi_y^*}\|_{\infty} \leq \frac{r_{\max}}{1-\gamma}$ . Applying the operator  $C$  repeatedly and using the fact that  $v^{\pi_x^* \pi_y^*} = C v^{\pi_x^* \pi_y^*}$  from Theorem 2.1 we obtain

$$\|v^{(k)} - v^{\pi_x^* \pi_y^*}\|_{\infty} \tag{32}$$

$$= \|(C)^k v^{(0)} - (C)^k v^{\pi_x^* \pi_y^*}\|_{\infty} \tag{33}$$

$$\leq \gamma^k \|v^{(0)} - v^{\pi_x^* \pi_y^*}\|_{\infty} \tag{34}$$

$$= \gamma^k \|v^{\pi_x^* \pi_y^*}\|_{\infty} \tag{35}$$

$$\leq \gamma^k \frac{r_{\max}}{1-\gamma} \tag{36}$$

where Equation (35) was obtained as  $v^{(0)} = 0$  Since  $1 - x \leq e^{-x}$  for any  $x \in \mathbb{R}$ , we have

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

Thus, for any  $\mathbf{s} \in \mathcal{S}$

$$\begin{aligned} v^{(k)}(\mathbf{s}) - v^{\pi_x^* \pi_y^*}(\mathbf{s}) &\leq \|v^{(k)} - v^{\pi_x^* \pi_y^*}\|_{\infty} \\ &= \gamma^k \frac{r_{\max}}{1-\gamma} \\ &\leq e^{-(1-\gamma)k} \frac{r_{\max}}{1-\gamma} \end{aligned}$$

Thus it suffices to solve for  $k$  such that

$$e^{-(1-\gamma)k} \frac{r_{\max}}{1-\gamma} \leq \epsilon .$$

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 and Greenwald [\[6\]](#), which characterizes the set of subdifferentials of parametrized constrained optimization problems, i.e., the set of subgradients w.r.t.  $\mathbf{x}$  of  $f^*(\mathbf{x}) = \max_{\mathbf{y} \in \mathcal{Y}: h(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y})$ . In particular, we note that Goktas and Greenwald's proof goes through even without assuming the concavity of  $f(\mathbf{x}, \mathbf{y}), h_1(\mathbf{x}, \mathbf{y}), \dots, h_d(\mathbf{x}, \mathbf{y})$  in  $\mathbf{y}$ , for all  $\mathbf{x} \in \mathcal{X}$ .

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

$$\text{conv} \left( \bigcup_{\mathbf{y}^*(\hat{\mathbf{x}}) \in \mathcal{Y}^*(\hat{\mathbf{x}})} \bigcup_{\lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \in \Lambda^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}}))} \left\{ \nabla_{\mathbf{x}} f(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) + \sum_{k=1}^d \lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \nabla_{\mathbf{x}} g_k(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \right\} \right), \quad (37)$$

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

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

$$v(\mathbf{s}, \mathbf{x}) = \max_{\mathbf{y} \in \mathcal{Y}: g(\mathbf{s}, \mathbf{x}, \mathbf{y}) \geq 0} \left\{ r(\mathbf{s}, \mathbf{x}, \mathbf{y}) + \gamma \mathbb{E}_{\mathbf{S}' \sim p(\cdot | \mathbf{s}, \mathbf{x}, \mathbf{y})} [v(\mathbf{S}', \mathbf{x})] \right\} \quad (38)$$

Suppose that Assumption [1.1](#) holds, and that 1. for all  $\mathbf{s} \in \mathcal{S}, \mathbf{y} \in \mathcal{Y}$ ,  $r(\mathbf{s}, \mathbf{x}, \mathbf{y}), g_1(\mathbf{s}, \mathbf{x}, \mathbf{y}), \dots, g_d(\mathbf{s}, \mathbf{x}, \mathbf{y})$  are concave in  $\mathbf{x}$ , 2.  $\nabla_{\mathbf{x}} r(\mathbf{s}, \mathbf{x}, \mathbf{y}), \nabla_{\mathbf{x}} g_1(\mathbf{s}, \mathbf{x}, \mathbf{y}), \dots, \nabla_{\mathbf{x}} g_d(\mathbf{s}, \mathbf{x}, \mathbf{y}), \nabla_{\mathbf{x}} p(\mathbf{s}' | \mathbf{s}, \mathbf{x}, \mathbf{y})$  are continuous in  $(\mathbf{s}, \mathbf{s}', \mathbf{x}, \mathbf{y})$ , 3.  $v(\mathbf{s}, \mathbf{x})$  is convex in  $\mathbf{x}$  4. Slater's condition holds for the optimization problem, i.e.,  $\forall \mathbf{x} \in \mathcal{X}, \mathbf{s} \in \mathcal{S}, \exists \hat{\mathbf{y}} \in \mathcal{Y}$  s.t.  $g_k(\mathbf{s}, \mathbf{x}, \hat{\mathbf{y}}) > 0$ , for all  $k = 1, \dots, d$ .

Let  $\mathcal{Y}^*(\mathbf{s}, \mathbf{x}) = \max_{\mathbf{y} \in \mathcal{Y}: g(\mathbf{s}, \mathbf{x}, \mathbf{y}) \geq 0} \{ r(\mathbf{s}, \mathbf{x}, \mathbf{y}) + \gamma \mathbb{E}_{\mathbf{S}' \sim p(\cdot | \mathbf{s}, \mathbf{x}, \mathbf{y})} [v(\mathbf{S}', \mathbf{x})] \}$ , then  $v$  is subdifferentiable and  $\partial_{\mathbf{x}} v(\mathbf{s}, \hat{\mathbf{x}}) =$

$$\text{conv} \left( \bigcup_{\mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}) \in \mathcal{Y}^*(\mathbf{s}, \hat{\mathbf{x}})} \bigcup_{\lambda_k^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \in \Lambda^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}))} \left\{ \nabla_{\mathbf{x}} r(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) + \gamma \nabla_{\mathbf{x}} \mathbb{E}_{\mathbf{S}' \sim p(\cdot | \mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}))} [v(\mathbf{S}', \hat{\mathbf{x}})] \right. \right. \\ \left. \left. + \sum_{k=1}^d \lambda_k^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \nabla_{\mathbf{x}} g_k(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \right\} \right). \quad (39)$$

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

$$\text{conv} \left( \bigcup_{\mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}) \in \mathcal{Y}^*(\mathbf{s}, \hat{\mathbf{x}})} \bigcup_{\lambda_k^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \in \Lambda^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}))} \left\{ \nabla_{\mathbf{x}} r(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) + \gamma \mathbb{E}_{\mathbf{S}' \sim p(\cdot | \mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}))} [\nabla_{\mathbf{x}} v(\mathbf{S}', \hat{\mathbf{x}})] \right. \right. \\ \left. \left. + \gamma \mathbb{E}_{\mathbf{S}' \sim p(\cdot | \mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}))} [v(\mathbf{S}', \hat{\mathbf{x}}) \nabla_{\mathbf{x}} \log(p(\mathbf{S}' | \mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})))] + \sum_{k=1}^d \lambda_k^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \nabla_{\mathbf{x}} g_k(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \right\} \right). \quad (40)$$

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

*Proof of Theorem C.2:* From Theorem C.1, we obtain the first part of the theorem:

$$\begin{aligned} & \partial_{\mathbf{x}} v(\mathbf{s}, \hat{\mathbf{x}}) \tag{41} \\ &= \text{conv} \left( \bigcup_{\mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}) \in \mathcal{Y}^*(\mathbf{s}, \hat{\mathbf{x}})} \bigcup_{\lambda_k^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \in \Lambda^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}))} \left\{ \nabla_{\mathbf{x}} r(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) + \gamma \nabla_{\mathbf{x}} \mathbb{E}_{\mathbf{s}' \sim p(\cdot | \mathbf{s}, \mathbf{x}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}))} [v(\mathbf{S}', \mathbf{x})] \right. \right. \\ & \left. \left. + \sum_{k=1}^d \lambda_k^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \nabla_{\mathbf{x}} g_k(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \right\} \right). \tag{42} \end{aligned}$$

By the Leibniz integral rule [67], 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  $\mathbf{s}, \mathbf{s}' \in \mathcal{S}, \mathbf{x} \in \mathcal{X}, \mathbf{y}^*(\mathbf{s}, \mathbf{x}) \in \mathcal{Y}^*(\mathbf{s}, \mathbf{x}) \nabla_{\mathbf{x}} p(\mathbf{s}' | \mathbf{s}, \mathbf{x}, \mathbf{y}^*(\mathbf{s}, \mathbf{x})) > 0$  we have:

$$\nabla_{\mathbf{x}} \mathbb{E}_{\mathbf{s}' \sim p(\cdot | \mathbf{s}, \mathbf{x}, \mathbf{y})} [v(\mathbf{S}', \mathbf{x})] \tag{43}$$

$$= \nabla_{\mathbf{x}} \int_{\mathbf{z} \in \mathcal{S}} p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y}) v(\mathbf{z}, \mathbf{x}) d\mathbf{z} \tag{44}$$

$$= \int_{\mathbf{z} \in \mathcal{S}} \nabla_{\mathbf{x}} [p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y}) v(\mathbf{z}, \mathbf{x})] d\mathbf{z} \tag{Leibniz Integral Rule} \tag{45}$$

$$= \int_{\mathbf{z} \in \mathcal{S}} [p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y}) \nabla_{\mathbf{x}} v(\mathbf{z}, \mathbf{x}) + v(\mathbf{z}, \mathbf{x}) \nabla_{\mathbf{x}} p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y})] d\mathbf{z} \tag{Product Rule} \tag{46}$$

$$= \int_{\mathbf{z} \in \mathcal{S}} \left[ p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y}) \nabla_{\mathbf{x}} v(\mathbf{z}, \mathbf{x}) + v(\mathbf{z}, \mathbf{x}) p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y}) \frac{\nabla_{\mathbf{x}} p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y})}{p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y})} \right] d\mathbf{z} \tag{p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y}) > 0} \tag{47}$$

$$= \int_{\mathbf{z} \in \mathcal{S}} [p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y}) \nabla_{\mathbf{x}} v(\mathbf{z}, \mathbf{x}) + v(\mathbf{z}, \mathbf{x}) p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y}) \nabla_{\mathbf{x}} \log p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y})] d\mathbf{z} \tag{48}$$

$$= \int_{\mathbf{z} \in \mathcal{S}} [p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y}) \nabla_{\mathbf{x}} v(\mathbf{z}, \mathbf{x})] d\mathbf{z} + \int_{\mathbf{z} \in \mathcal{S}} [v(\mathbf{z}, \mathbf{x}) p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y}) \nabla_{\mathbf{x}} \log p(\mathbf{z} | \mathbf{s}, \mathbf{x}, \mathbf{y})] d\mathbf{z} \tag{49}$$

$$= \mathbb{E}_{\mathbf{s}' \sim p(\cdot | \mathbf{s}, \mathbf{x}, \mathbf{y})} [\nabla_{\mathbf{x}} v(\mathbf{S}', \mathbf{x})] + \mathbb{E}_{\mathbf{s}' \sim p(\cdot | \mathbf{s}, \mathbf{x}, \mathbf{y})} [v(\mathbf{S}', \mathbf{x}) \nabla_{\mathbf{x}} \log p(\mathbf{S}' | \mathbf{s}, \mathbf{x}, \mathbf{y})] \tag{50}$$

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

$$\begin{aligned} & \text{conv} \left( \bigcup_{\mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}) \in \mathcal{Y}^*(\mathbf{s}, \hat{\mathbf{x}})} \bigcup_{\lambda_k^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \in \Lambda^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}))} \left\{ \nabla_{\mathbf{x}} r(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) + \gamma \mathbb{E}_{\mathbf{s}' \sim p(\cdot | \mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}))} [\nabla_{\mathbf{x}} v(\mathbf{S}', \hat{\mathbf{x}})] \right. \right. \\ & \left. \left. + \gamma \mathbb{E}_{\mathbf{s}' \sim p(\cdot | \mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}))} [v(\mathbf{S}', \hat{\mathbf{x}}) \nabla_{\mathbf{x}} \log (p(\mathbf{S}' | \mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})))] + \sum_{k=1}^d \lambda_k^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \nabla_{\mathbf{x}} g_k(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \right\} \right). \tag{51} \end{aligned}$$

□

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

transition function  $\tau$  (Proposition 1 of [68]). In this case, the subdifferential of the Bellman equation reduces to

$$\partial_{\mathbf{x}} v(\mathbf{s}, \hat{\mathbf{x}}) \quad (52)$$

$$= \text{conv} \left( \bigcup_{\mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}) \in \mathcal{Y}^*(\mathbf{s}, \hat{\mathbf{x}})} \bigcup_{\lambda_k^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \in \Lambda^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}}))} \{ \nabla_{\mathbf{x}} r(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) + \gamma \nabla_{\mathbf{x}} \tau(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}) \nabla_{\mathbf{s}} v(\tau(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}), \hat{\mathbf{x}}) \} \right) \quad (53)$$

$$+ \nabla_{\mathbf{x}} v(\tau(\mathbf{s}, \hat{\mathbf{x}}), \hat{\mathbf{x}}) + \sum_{k=1}^d \lambda_k^*(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \nabla_{\mathbf{x}} g_k(\mathbf{s}, \hat{\mathbf{x}}, \mathbf{y}^*(\mathbf{s}, \hat{\mathbf{x}})) \quad (54)$$

**Theorem 3.1.** Consider a zero-sum stochastic Stackelberg game  $\mathcal{G}$ , where  $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid q_1(\mathbf{x}) \leq 0, \dots, q_p(\mathbf{x}) \leq 0\}$  and  $\mathcal{Y} = \{\mathbf{y} \in \mathbb{R}^m \mid r_1(\mathbf{y}) \geq 0, \dots, r_l(\mathbf{y}) \geq 0\}$  are convex. Let  $\mathcal{L}_{\mathbf{s}, \mathbf{x}}(\mathbf{y}, \boldsymbol{\lambda}) = r(\mathbf{s}, \mathbf{x}, \mathbf{y}) + \gamma \mathbb{E}_{\mathbf{S}' \sim p(\cdot | \mathbf{s}, \mathbf{x}, \mathbf{y})} [v(\mathbf{S}', \mathbf{x})] + \sum_{k=1}^d \lambda_k g_k(\mathbf{s}, \mathbf{x}, \mathbf{y})$  where  $Cv = v$ .

Suppose that Assumption [I.1] holds, and that 1. for all  $\mathbf{s} \in \mathcal{S}$ ,  $\max_{\mathbf{y} \in \mathcal{Y}: g(\mathbf{s}, \mathbf{x}, \mathbf{y}) \geq 0} \{r(\mathbf{s}, \mathbf{x}, \mathbf{y}) + \gamma \mathbb{E}_{\mathbf{S}' \sim p(\cdot | \mathbf{s}, \mathbf{x}, \mathbf{y})} [v(\mathbf{S}', \mathbf{x})]\}$  is concave in  $\mathbf{x}$ , 2.  $\nabla_{\mathbf{x}} r(\mathbf{s}, \mathbf{x}, \mathbf{y}), \nabla_{\mathbf{x}} g_1(\mathbf{s}, \mathbf{x}, \mathbf{y}), \dots, \nabla_{\mathbf{x}} g_d(\mathbf{s}, \mathbf{x}, \mathbf{y}), \nabla_{\mathbf{y}} r(\mathbf{s}, \mathbf{x}, \mathbf{y}), \nabla_{\mathbf{y}} g_1(\mathbf{s}, \mathbf{x}, \mathbf{y}), \dots, \nabla_{\mathbf{y}} g_d(\mathbf{s}, \mathbf{x}, \mathbf{y})$  exist, for all  $\mathbf{s} \in \mathcal{S}, \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}$ , 4.  $p(\mathbf{s}' \mid \mathbf{s}, \mathbf{x}, \mathbf{y})$  is continuous and differentiable in  $(\mathbf{x}, \mathbf{y})$ , and 5. Slater's condition holds, i.e.,  $\forall \mathbf{s} \in \mathcal{S}, \mathbf{x} \in \mathcal{X}, \exists \hat{\mathbf{y}} \in \mathcal{Y}$  s.t.  $g_k(\mathbf{s}, \mathbf{x}, \hat{\mathbf{y}}) > 0$ , for all  $k = 1, \dots, d$  and  $r_j(\hat{\mathbf{y}}) > 0$ , for all  $j = 1, \dots, l$ , and  $\exists \mathbf{x} \in \mathbb{R}^n$  s.t.  $q_k(\mathbf{x}) < 0$  for all  $k = 1 \dots, p$ . Then, there exists  $\boldsymbol{\mu}^* : \mathcal{S} \rightarrow \mathbb{R}_+^p$ ,  $\boldsymbol{\lambda}^* : \mathcal{S} \times \mathcal{X} \rightarrow \mathbb{R}_+^d$ , and  $\boldsymbol{\nu}^* : \mathcal{S} \times \mathcal{X} \rightarrow \mathbb{R}_+^l$  s.t. a policy profile  $(\boldsymbol{\pi}_{\mathbf{x}}^*, \boldsymbol{\pi}_{\mathbf{y}}^*) \in \mathcal{X}^{\mathcal{S}} \times \mathcal{Y}^{\mathcal{S}}$  is a recSE of  $\mathcal{G}$  only if it satisfies the following conditions, for all  $\mathbf{s} \in \mathcal{S}$ :

$$\nabla_{\mathbf{x}} \mathcal{L}_{\mathbf{s}, \boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s})}(\boldsymbol{\pi}_{\mathbf{y}}^*(\mathbf{s}), \boldsymbol{\lambda}^*(\mathbf{s}, \boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s}))) + \sum_{k=1}^p \mu_k^*(\mathbf{s}) \nabla_{\mathbf{x}} q_k(\boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s})) = 0 \quad (8)$$

$$\nabla_{\mathbf{y}} \mathcal{L}_{\mathbf{s}, \boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s})}(\boldsymbol{\pi}_{\mathbf{y}}^*(\mathbf{s}), \boldsymbol{\lambda}^*(\mathbf{s}, \boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s}))) + \sum_{k=1}^l \nu_k^*(\mathbf{s}, \boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s})) \nabla_{\mathbf{x}} r_k(\boldsymbol{\pi}_{\mathbf{y}}^*(\mathbf{s})) = 0 \quad (9)$$

$$\mu_k^*(\mathbf{s}) q_k(\boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s})) = 0 \quad q_k(\boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s})) \leq 0 \quad \forall k \in [p] \quad (10)$$

$$g_k(\mathbf{s}, \boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s}), \boldsymbol{\pi}_{\mathbf{y}}^*(\mathbf{s})) \geq 0 \quad \lambda_k^*(\mathbf{s}, \boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s})) g_k(\mathbf{s}, \boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s}), \boldsymbol{\pi}_{\mathbf{y}}^*(\mathbf{s})) = 0 \quad \forall k \in [d] \quad (11)$$

$$\nu_k^*(\mathbf{s}, \boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s})) \nabla_{\mathbf{x}} r_k(\boldsymbol{\pi}_{\mathbf{y}}^*(\mathbf{s})) = 0 \quad r_k(\boldsymbol{\pi}_{\mathbf{y}}^*(\mathbf{s})) \geq 0 \quad \forall k \in [l] \quad (12)$$

*Proof of Theorem 3.1* By Theorem [2.1] and Theorem [2.3] we know that  $(\boldsymbol{\pi}_{\mathbf{x}}^*, \boldsymbol{\pi}_{\mathbf{y}}^*)$  is a recursive Stackelberg equilibrium iff

$$v^{\boldsymbol{\pi}_{\mathbf{y}}^* \boldsymbol{\pi}_{\mathbf{x}}^*}(\mathbf{s}) = \left( Cv^{\boldsymbol{\pi}_{\mathbf{y}}^* \boldsymbol{\pi}_{\mathbf{x}}^*} \right) (\mathbf{s}) . \quad (55)$$

Note that for any policy profile  $(\boldsymbol{\pi}_{\mathbf{x}}^*, \boldsymbol{\pi}_{\mathbf{y}}^*)$  that satisfies  $v^{\boldsymbol{\pi}_{\mathbf{y}}^* \boldsymbol{\pi}_{\mathbf{x}}^*}(\mathbf{s}) = \left( Cv^{\boldsymbol{\pi}_{\mathbf{y}}^* \boldsymbol{\pi}_{\mathbf{x}}^*} \right) (\mathbf{s})$  by Definition [1.2] we have that  $(\boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s}), \boldsymbol{\pi}_{\mathbf{y}}^*(\mathbf{s}))$  is a Stackelberg equilibrium of

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

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

Fix a state  $\mathbf{s} \in \mathcal{S}$ , under the assumptions of the theorem, the conditions of Theorem [C.2] are satisfied and  $u^*(\mathbf{s}, \mathbf{x}) = \max_{\mathbf{y} \in \mathcal{Y}: g(\mathbf{s}, \mathbf{x}, \mathbf{y}) \geq 0} \{r(\mathbf{s}, \mathbf{x}, \mathbf{y}) + \gamma \mathbb{E}_{\mathbf{S}' \sim p(\cdot | \mathbf{s}, \mathbf{x}, \mathbf{y})} [v(\mathbf{S}')] \}$  is subdifferentiable in  $\mathbf{x}$ . Since  $u^*(\mathbf{s}, \mathbf{x})$  is convex in  $\mathbf{x}$ , and Slater's condition are satisfied by the assumptions of the theorem, the necessary and sufficient conditions for  $\boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s})$  to be an optimal solution to  $\min_{\mathbf{x} \in \mathcal{X}} u^*(\mathbf{s}, \mathbf{x})$  are given by the KKT conditions [69] for  $\min_{\mathbf{x} \in \mathcal{X}} u^*(\mathbf{s}, \mathbf{x})$ . Note that we can state the first order KKT conditions explicitly thanks to the subdifferential Benveniste-Scheinkman theorem (Theorem [C.2]). That is,  $\boldsymbol{\pi}_{\mathbf{x}}^*(\mathbf{s})$  is an optimal solution to  $\min_{\mathbf{x} \in \mathcal{X}} u^*(\mathbf{s}, \mathbf{x})$  if there exists  $\boldsymbol{\mu}^*(\mathbf{s}) \in \mathbb{R}_+^p$  such that:

$$\nabla_{\mathbf{x}} \mathcal{L}_{s, \pi_{\mathbf{x}}^*}(y^*(s, \pi_{\mathbf{x}}^*(s)), \boldsymbol{\lambda}^*(s, \pi_{\mathbf{x}}^*(s), y^*(s, \pi_{\mathbf{x}}^*(s))) + \sum_{k=1}^p \mu_k^*(s) \nabla_{\mathbf{x}} q_k(\pi_{\mathbf{x}}^*(s)) = 0 \quad (56)$$

$$\mu_k^*(s) q_k(\pi_{\mathbf{x}}^*(s)) = 0 \quad \forall k \in [p] \quad (57)$$

$$q_k(\pi_{\mathbf{x}}^*(s)) \leq 0 \quad \forall k \in [p] \quad (58)$$

where  $y^*(s, \pi_{\mathbf{x}}^*(s)) \in \arg \max_{\mathbf{y} \in \mathcal{Y}: g(s, \pi_{\mathbf{x}}^*(s), \mathbf{y}) \geq 0} \{r(s, \pi_{\mathbf{x}}^*(s), \mathbf{y}) + \gamma \mathbb{E}_{S' \sim p(\cdot | s, \pi_{\mathbf{x}}^*(s), \mathbf{y})} [v(S')]\}$  and  $\boldsymbol{\lambda}^*(s, \pi_{\mathbf{x}}^*(s), y^*(s, \pi_{\mathbf{x}}^*(s))) = (\lambda_1^*(s, \pi_{\mathbf{x}}^*(s), y^*(s, \pi_{\mathbf{x}}^*(s))), \dots, \lambda_d^*(s, \pi_{\mathbf{x}}^*(s), y^*(s, \pi_{\mathbf{x}}^*(s))))^T \in \Lambda^*(s, \pi_{\mathbf{x}}^*(s), y^*(s, \pi_{\mathbf{x}}^*(s)))$  are the optimal KKT multipliers associated with  $y^*(s, \pi_{\mathbf{x}}^*(s)) \in \mathcal{Y}^*(s, \pi_{\mathbf{x}}^*(s))$  which are guaranteed to exist since Slater's condition is satisfied for  $\max_{\mathbf{y} \in \mathcal{Y}: g(s, \mathbf{x}, \mathbf{y}) \geq 0} \{r(s, \mathbf{x}, \mathbf{y}) + \gamma \mathbb{E}_{S' \sim p(\cdot | s, \mathbf{x}, \mathbf{y})} [v(S')]\}$ .

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

$$\nabla_{\mathbf{y}} \mathcal{L}_{s, \mathbf{x}}(\pi_{\mathbf{y}}^*(s), \boldsymbol{\lambda}^*(s, \mathbf{x})) + \sum_{k=1}^l \nu_k^*(s) \nabla_{\mathbf{x}} r_k(\pi_{\mathbf{y}}^*(s)) = 0 \quad (59)$$

$$g_k(s, \mathbf{x}, \pi_{\mathbf{y}}^*(s)) \geq 0 \quad \forall k \in [d] \quad (60)$$

$$\lambda_k^*(s, \mathbf{x}) g_k(s, \mathbf{x}, \pi_{\mathbf{y}}^*(s)) = 0 \quad \forall k \in [d] \quad (61)$$

$$\nu_k^*(s, \mathbf{x}) \nabla_{\mathbf{x}} r_k(\pi_{\mathbf{y}}^*(s)) = 0 \quad \forall k \in [l] \quad (62)$$

$$r_k(\mathbf{x}) \geq 0 \quad \forall k \in [l] \quad (63)$$

Combining the necessary and *sufficient* conditions in Equations (56) to (58) with the necessary conditions in Equations (59) to (63), we obtain the necessary conditions for  $(\pi_{\mathbf{x}}^*, \pi_{\mathbf{y}}^*)$  to be a recursive Stackelberg equilibrium. □

If additionally, the objective of the inner player at each state  $s \in \mathcal{S}$ ,  $r(s, \mathbf{x}, \mathbf{y}) + \gamma \mathbb{E}_{S' \sim p(\cdot | s, \mathbf{x}, \mathbf{y})} [v(S', \mathbf{x})]$  is concave in  $\mathbf{y}$ , then the above conditions become necessary and sufficient. The proof follows exactly the same, albeit the optimality conditions on the inner player's policy become necessary and sufficient.

**Theorem C.3** (Recursive Stackelberg Equilibrium Necessary and Sufficient Optimality Conditions). *Consider a zero-sum stochastic Stackelberg game  $\mathcal{G}$ , where  $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid q_1(\mathbf{x}) \leq 0, \dots, q_p(\mathbf{x}) \leq 0\}$  and  $\mathcal{Y} = \{\mathbf{y} \in \mathbb{R}^m \mid r_1(\mathbf{y}) \geq 0, \dots, r_l(\mathbf{y}) \geq 0\}$  are convex. Let  $\mathcal{L}_{s, \mathbf{x}}(\mathbf{y}, \boldsymbol{\lambda}) = r(s, \mathbf{x}, \mathbf{y}) + \gamma \mathbb{E}_{S' \sim p(\cdot | s, \mathbf{x}, \mathbf{y})} [v(S', \mathbf{x})] + \sum_{k=1}^d \lambda_k g_k(s, \mathbf{x}, \mathbf{y})$  where  $Cv = v$ . Suppose that Assumption 1.1 holds, and that 1. for all  $s \in \mathcal{S}$ ,  $\mathbf{y} \in \mathcal{Y}$ ,  $\max_{\mathbf{x} \in \mathcal{X}: r_1(s, \mathbf{x}, \mathbf{y}) \geq 0} \{r(s, \mathbf{x}, \mathbf{y}) + \gamma \mathbb{E}_{S' \sim p(\cdot | s, \mathbf{x}, \mathbf{y})} [v(S', \mathbf{x})]\}$  is convex in  $\mathbf{x}$  and  $r_1(s, \mathbf{x}, \mathbf{y}), g_1(s, \mathbf{x}, \mathbf{y}), \dots, g_d(s, \mathbf{x}, \mathbf{y})$  are concave in  $\mathbf{y}$  for all  $\mathbf{x} \in \mathcal{X}$  and  $s \in \mathcal{S}$ , 2.  $\nabla_{\mathbf{x}} r(s, \mathbf{x}, \mathbf{y}), \nabla_{\mathbf{x}} g_1(s, \mathbf{x}, \mathbf{y}), \dots, \nabla_{\mathbf{x}} g_d(s, \mathbf{x}, \mathbf{y}), \nabla_{\mathbf{y}} r(s, \mathbf{x}, \mathbf{y}), \nabla_{\mathbf{y}} g_1(s, \mathbf{x}, \mathbf{y}), \dots, \nabla_{\mathbf{y}} g_d(s, \mathbf{x}, \mathbf{y})$  exist, for all  $s \in \mathcal{S}$ ,  $\mathbf{x} \in \mathcal{X}$ ,  $\mathbf{y} \in \mathcal{Y}$ , 4.  $p(s' \mid s, \mathbf{x}, \mathbf{y})$  is continuous and differentiable in  $(\mathbf{x}, \mathbf{y})$  and CSD concave in  $\mathbf{y}$ , and 5. Slater's condition holds, i.e.,  $\forall s \in \mathcal{S}, \mathbf{x} \in \mathcal{X}, \exists \hat{\mathbf{y}} \in \mathcal{Y}$  s.t.  $g_k(s, \mathbf{x}, \hat{\mathbf{y}}) > 0$ , for all  $k = 1, \dots, d$  and  $r_j(\hat{\mathbf{y}}) > 0$ , for all  $j = 1, \dots, l$ , and  $\exists \hat{\mathbf{x}} \in \mathbb{R}^n$  s.t.  $q_k(\hat{\mathbf{x}}) < 0$  for all  $k = 1, \dots, p$ . Then, there exists  $\boldsymbol{\mu}^* : \mathcal{S} \rightarrow \mathbb{R}_+^p$ ,  $\boldsymbol{\lambda}^* : \mathcal{S} \times \mathcal{X} \rightarrow \mathbb{R}_+^d$ , and  $\boldsymbol{\nu}^* : \mathcal{S} \times \mathcal{X} \rightarrow \mathbb{R}_+^l$  s.t. a policy profile*

$(\boldsymbol{\pi}_x^*, \boldsymbol{\pi}_y^*) \in \mathcal{X}^S \times \mathcal{Y}^S$  is a recSE of  $\mathcal{G}$  only if it satisfies the following conditions, for all  $\mathbf{s} \in \mathcal{S}$ :

$$\nabla_{\mathbf{x}} \mathcal{L}_{\mathbf{s}, \boldsymbol{\pi}_x^*}(\boldsymbol{\pi}_y^*(\mathbf{s}), \boldsymbol{\lambda}^*(\mathbf{s}, \boldsymbol{\pi}_x^*(\mathbf{s}))) + \sum_{k=1}^p \mu_k^*(\mathbf{s}) \nabla_{\mathbf{x}} q_k(\boldsymbol{\pi}_x^*(\mathbf{s})) = 0 \quad (64)$$

$$\nabla_{\mathbf{y}} \mathcal{L}_{\mathbf{s}, \boldsymbol{\pi}_x^*}(\boldsymbol{\pi}_y^*(\mathbf{s}), \boldsymbol{\lambda}^*(\mathbf{s}, \boldsymbol{\pi}_x^*(\mathbf{s}))) + \sum_{k=1}^l \nu_k^*(\mathbf{s}, \boldsymbol{\pi}_x^*(\mathbf{s})) \nabla_{\mathbf{y}} r_k(\boldsymbol{\pi}_y^*(\mathbf{s})) = 0 \quad (65)$$

$$\mu_k^*(\mathbf{s}) q_k(\boldsymbol{\pi}_x^*(\mathbf{s})) = 0 \quad q_k(\boldsymbol{\pi}_x^*(\mathbf{s})) \leq 0 \quad \forall k \in [p] \quad (66)$$

$$g_k(\mathbf{s}, \boldsymbol{\pi}_x^*(\mathbf{s}), \boldsymbol{\pi}_y^*(\mathbf{s})) \geq 0 \quad \lambda_k^*(\mathbf{s}, \boldsymbol{\pi}_x^*(\mathbf{s})) g_k(\mathbf{s}, \boldsymbol{\pi}_x^*(\mathbf{s}), \boldsymbol{\pi}_y^*(\mathbf{s})) = 0 \quad \forall k \in [d] \quad (67)$$

$$\nu_k^*(\mathbf{s}, \boldsymbol{\pi}_x^*(\mathbf{s})) \nabla_{\mathbf{y}} r_k(\boldsymbol{\pi}_y^*(\mathbf{s})) = 0 \quad r_k(\boldsymbol{\pi}_y^*(\mathbf{s})) \geq 0 \quad \forall k \in [l] \quad (68)$$

## 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 associated 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.** Consider a stochastic Fisher market  $\mathcal{F}$  such that the transition probability function  $p$  is continuous in  $\beta_i$  and independent of  $\mathbf{x}_i$ . For any price system  $\mathbf{p} \in \mathbb{R}_+^{S \times m}$ , a tuple  $(\mathbf{x}_i^*, \beta_i^*) \in \mathbb{R}_+^{S \times m} \times \mathbb{R}_+^m$  consisting of an allocation system and saving system for a buyer  $i \in [n]$ , given by a continuous, and homogeneous utility function  $u_i : \mathbb{R}_+^m \times \mathcal{T} \rightarrow \mathbb{R}$  representing a locally non-satiated preference, is expected utility maximizing constrained by the saving and spending constraints, i.e.,  $(\mathbf{x}_i^*, \beta_i^*)$  is the optimal policy solving the following recursive Bellman equation  $\nu_i(\mathbf{s}) = \max_{(\mathbf{x}_i, \beta_i) \in \mathbb{R}_+^{m+1} : \mathbf{x}_i \cdot \mathbf{p}^*(\mathbf{s}) + \beta_i \leq b_i} \left\{ u_i(\mathbf{x}_i, \mathbf{t}_i) + \gamma \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}') \sim p(\cdot | \mathbf{s}, (\beta_i, \beta_{-i}^*(\mathbf{s})))} [\nu_i(\mathbf{T}', \mathbf{b}' + \beta_i, \mathbf{q}')] \right\}$ , only if we have for all states  $\mathbf{s} \in \mathcal{S}$ ,  $\mathbf{x}_i^*(\mathbf{T}, \mathbf{b}, \mathbf{q}) \cdot \mathbf{p}(\mathbf{T}, \mathbf{b}, \mathbf{q}) + \beta_i^*(\mathbf{T}, \mathbf{b}, \mathbf{q}) \leq b_i$ , and,

$$x_{ij}^*(\mathbf{s}) > 0 \implies \frac{\frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)}{p_j(\mathbf{s})} = \frac{u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)}{b_i - \beta_i^*(\mathbf{s})} \quad \forall j \in [m] \quad (69)$$

$$\beta_i^*(\mathbf{s}) > 0 \implies \frac{\partial \nu_i^*}{\partial b_i}(\mathbf{s}) = \gamma \frac{\partial}{\partial \beta_i} \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}') \sim p(\cdot | \mathbf{s}, (\beta_i, \beta_{-i}^*(\mathbf{s})))} [\nu_i^*(\mathbf{T}', \mathbf{b}' + \beta_i^*(\mathbf{s}), \mathbf{q}')] \quad (70)$$

If additionally,  $u_i$  is concave and  $p$  is CSD concave in  $(\mathbf{x}_i, \beta_i)$ , then the above condition becomes also sufficient.

*Proof of Lemma D.1* Fix a buyer  $i \in [n]$ . Throughout we use  $\mathbf{b} + \beta_i$  as shortcut for  $\mathbf{b} + (\beta_i, \mathbf{0}_{-i})$ . Suppose that  $\nu_i^*$  solves the following recursive Bellman equation:

$$\nu_i(\mathbf{s}) = \max_{(\mathbf{x}_i, \beta_i) \in \mathbb{R}_+^{m+1} : \mathbf{x}_i \cdot \mathbf{p}^*(\mathbf{s}) + \beta_i \leq b_i} \left\{ u_i(\mathbf{x}_i, \mathbf{t}_i) + \gamma \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}') \sim p(\cdot | \mathbf{s}, (\beta_i, \beta_{-i}^*(\mathbf{s})))} [\nu_i(\mathbf{T}', \mathbf{b}' + \beta_i, \mathbf{q}')] \right\} \quad (71)$$

Define the Lagrangian associated with the consumption-saving problem:

$$\max_{(\mathbf{x}_i, \beta_i) \in \mathbb{R}_+^{m+1} : \mathbf{x}_i \cdot \mathbf{p}^*(\mathbf{s}) + \beta_i \leq b_i} \left\{ u_i(\mathbf{x}_i, \mathbf{t}_i) + \gamma \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}') \sim p(\cdot | \mathbf{s}, (\beta_i, \beta_{-i}^*(\mathbf{s})))} [\nu_i^*(\mathbf{T}', \mathbf{b}' + \beta_i, \mathbf{q}')] \right\} \quad (72)$$

as follows:

$$\begin{aligned} \mathcal{L}(\mathbf{x}_i, \beta_i, \lambda, \boldsymbol{\mu}; \mathbf{p}) &= u_i(\mathbf{x}_i; \mathbf{t}_i) + \gamma \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}') \sim p(\cdot | \mathbf{T}, \mathbf{b}, \mathbf{q}, (\beta_i, \beta_{-i}^*(\mathbf{s})))} [\nu_i^*(\mathbf{T}', \mathbf{b}' + \beta_i, \mathbf{q}')] \\ &\quad + \lambda(b_i - \mathbf{x}_i \mathbf{p}) + \sum_{j \in [m]} \mu_j x_{ij} + \mu_{m+1} \beta_i \end{aligned} \quad (73)$$



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

Assume that for any state  $\mathbf{s} \in \mathcal{S}$ , we have  $b_i > 0$ . We can ignore states such that  $b_i > 0$  since at those states the buyer cannot be allocated goods, and can also not put aside savings. Then, Slater's condition holds and the necessary first order optimality conditions for an allocation  $\mathbf{x}_i^*(\mathbf{s}) \in \mathbb{R}_+^m$ , saving  $\beta_i^*(\mathbf{s}) \in \mathbb{R}_+$  and associated Lagrangian multipliers  $\lambda^*(\mathbf{s}) \in \mathbb{R}_+$ ,  $\mu^*(\mathbf{s}) \in \mathbb{R}^{m+1}$  to be optimal for any prices  $\mathbf{p}(\mathbf{s}) \in \mathbb{R}_+^m$  and state  $\mathbf{s} \in \mathcal{S}$  are given by the following pair of KKT conditions [69] for all  $j \in [m]$ :

$$\frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i) - \lambda^*(\mathbf{s})p_j(\mathbf{s}) + \mu_j^*(\mathbf{s}) \doteq 0 \quad (74)$$

$$\gamma \frac{\partial}{\partial \beta_i} \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}')} [\nu_i^*(\mathbf{T}', \mathbf{b}' + \beta_i^*, \mathbf{q}')] - \lambda_i^*(\mathbf{s}) + \mu_{m+1}^*(\mathbf{s}) \doteq 0 \quad (75)$$

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

$$\frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i) - \lambda^*(\mathbf{s})p_j(\mathbf{s}) = 0 \quad \forall j \in [m] \quad (76)$$

$$\beta_i^*(\mathbf{s}) > 0 \implies \gamma \frac{\partial}{\partial \beta_i} \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}')} [\nu_i^*(\mathbf{T}', \mathbf{b}' + \beta_i^*, \mathbf{q}')] - \lambda^*(\mathbf{s}) = 0 \quad \forall j \in [m] \quad (77)$$

Re-organizing expressions, yields:

$$x_{ij}^*(\mathbf{s}) > 0 \implies \lambda^* = \frac{\frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)}{p_j(\mathbf{s})} \quad \forall j \in [m] \quad (78)$$

$$\beta_i^*(\mathbf{s}) > 0 \implies \lambda^*(\mathbf{s}) = \gamma \frac{\partial}{\partial \beta_i} \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}')} [\nu_i^*(\mathbf{T}', \mathbf{b}' + \beta_i^*, \mathbf{q}')] \quad (79)$$

Using the envelope theorem [70, 71], we can also compute  $\frac{\partial \nu_i^*}{\partial \beta_i}(\mathbf{s})$  as follows:

$$\frac{\partial \nu_i^*}{\partial \beta_i}(\mathbf{s}) = \lambda^*(\mathbf{s}) \quad (80)$$

We note that for all states  $\mathbf{s} \in \mathcal{S}$ ,  $\frac{\partial \nu_i^*}{\partial \beta_i}(\mathbf{s})$  is well-defined since  $\lambda^*(\mathbf{s})$  is uniquely defined for all states by Equation (78). Hence, combining the above with Equation (78) and Equation (79), we get:

$$x_{ij}^*(\mathbf{s}) > 0 \implies \lambda^* = \frac{\frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)}{p_j} \quad \forall j \in [m] \quad (81)$$

$$\beta_i^*(\mathbf{s}) > 0 \implies \frac{\partial \nu_i^*}{\partial \beta_i}(\mathbf{s}) = \gamma \frac{\partial}{\partial \beta_i} \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}')} [\nu_i^*(\mathbf{T}', \mathbf{b}' + \beta_i^*, \mathbf{q}')] \quad (82)$$

Finally, going back to Equation (74), multiplying by  $x_{ij}^*(\mathbf{s})$  and summing up across all  $j \in [m]$ , we obtain:

$$\sum_{j \in [m]} x_{ij}^*(\mathbf{s}) \frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i) - \lambda^*(\mathbf{s}) \sum_{j \in [m]} p_j(\mathbf{s}) x_{ij}^*(\mathbf{s}) + \mu_j^* x_{ij}^*(\mathbf{s}) = 0 \quad (83)$$

Using Euler's theorem for homogeneous functions on the partial derivatives of the utility functions, we then have:

$$u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i) - \lambda^*(\mathbf{s}) \sum_{j \in [m]} p_j(\mathbf{s}) x_{ij}^*(\mathbf{s}) + \mu_j^* x_{ij}^*(\mathbf{s}) = 0 \quad (84)$$

Additionally, the KKT Slack complementarity conditions, we have  $\lambda^*(\mathbf{s})(b_i - \beta_i^*(\mathbf{s})) = \lambda^*(\mathbf{s}) \sum_{j \in [m]} p_j(\mathbf{s}) x_{ij}^*(\mathbf{s})$ :

$$u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i) - \lambda^*(\mathbf{s})(b_i - \beta_i^*(\mathbf{s})) = 0 \quad (85)$$

$$\lambda^*(\mathbf{s}) = \frac{u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)}{b_i - \beta_i^*(\mathbf{s})} \quad (86)$$

Combining the above conditions, with Equation (98), and adding to it Equation (82), and ensuring that the KKT primal feasibility conditions hold as well, we obtain the following necessary conditions that need to hold for all states  $\mathbf{s} \in \mathcal{S}$ :

$$x_{ij}^*(\mathbf{s}) > 0 \implies \frac{\frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)}{p_j(\mathbf{s})} = \frac{u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)}{b_i - \beta_i^*(\mathbf{s})} \quad \forall j \in [m] \quad (87)$$

$$\beta_i^*(\mathbf{s}) > 0 \implies \frac{\partial \nu_i^*}{\partial b_i}(\mathbf{s}) = \gamma \frac{\partial}{\partial \beta_i} \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}')} [\nu_i^*(\mathbf{T}', \mathbf{b}' + \beta_i^*(\mathbf{s}), \mathbf{q}')] \quad (88)$$

If additionally the transition probability function  $p$  is CSD concave in  $\beta_i$ , then  $\nu_i$  is concave and the utility maximization problem is concave, which in turn implies that the above conditions are also sufficient. □

**Theorem 4.1.** *A stochastic Fisher market with savings  $\mathcal{F}$  in which  $\mathcal{U}$  is a set of continuous and homogeneous utility functions and the transition function is continuous in  $\beta_i$  has at least one recCE. Additionally, the recSE that solves the following Bellman equation corresponds to a recCE of  $\mathcal{F}$ :*

$$v(\mathbf{s}) = \min_{\mathbf{p} \in \mathbb{R}_+^m} \max_{(\mathbf{X}, \beta) \in \mathbb{R}_+^{n \times (m+1)}: \mathbf{X}\mathbf{p} + \beta \leq \mathbf{b}} \sum_{j \in [m]} q_j p_j + \sum_{i \in [n]} (b_i - \beta_i) \log(u_i(\mathbf{x}_i, \mathbf{t}_i)) + \gamma \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}') \sim p(\cdot | \mathbf{s}, \beta)} [v(\mathbf{T}', \mathbf{b}' + \beta, \mathbf{q}')] \quad (13)$$

*Proof of Theorem 4.1* Fix a buyer  $i \in [n]$ . Suppose that  $v^*$  solves the following Stochastic Stackelberg game:

$$v(\mathbf{s}) = \min_{\mathbf{p} \in \mathbb{R}_+^m} \max_{(\mathbf{X}, \beta) \in \mathbb{R}_+^{n \times (m+1)}: \mathbf{X}\mathbf{p} + \beta \leq \mathbf{b}} \sum_{j \in [m]} q_j p_j + \sum_{i \in [n]} (b_i - \beta_i) \log(u_i(\mathbf{x}_i, \mathbf{t}_i)) + \gamma \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}') \sim p(\cdot | \mathbf{s}, \beta)} [v(\mathbf{T}', \mathbf{b}' + \beta, \mathbf{q}')] \quad (89)$$

Define the Lagrangian associated with the following optimization problem:

$$\min_{\mathbf{p} \in \mathbb{R}_+^m} \max_{(\mathbf{X}, \beta) \in \mathbb{R}_+^{n \times (m+1)}: \mathbf{X}\mathbf{p} + \beta \leq \mathbf{b}} \sum_{j \in [m]} q_j p_j + \sum_{i \in [n]} (b_i - \beta_i) \log(u_i(\mathbf{x}_i, \mathbf{t}_i)) + \gamma \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}') \sim p(\cdot | \mathbf{s}, \beta)} [v^*(\mathbf{T}', \mathbf{b}' + \beta, \mathbf{q}')] \quad (90)$$

as follows:

$$\begin{aligned} \mathcal{L}(\mathbf{p}, \mathbf{X}, \beta, \lambda) = & \sum_{j \in [m]} q_j p_j + \sum_{i \in [n]} (b_i - \beta_i) \log(u_i(\mathbf{x}_i, \mathbf{t}_i)) + \gamma \mathbb{E}_{(\mathbf{s}') \sim p(\cdot | \mathbf{s}, \beta)} [v(\mathbf{T}', \mathbf{b}' + \beta, \mathbf{q}')] \\ & + \sum_{j \in [m]} \lambda_j (b_j - \mathbf{x}_j \cdot \mathbf{p} + \beta_j) \quad . \end{aligned} \quad (91)$$

By Theorem 3.1, the necessary optimality conditions for the stochastic Stackelberg game are that for all states  $\mathbf{s} \in \mathcal{S}$  there exists  $\boldsymbol{\mu}^*(\mathbf{s}) \in \mathbb{R}^{n \times (m+1)}$ , and  $\boldsymbol{\nu}^*(\mathbf{s}) \in \mathbb{R}_+^{n \times (m+1)}$  associated with the non-negativity constraints for  $(\mathbf{X}(\mathbf{s}), \beta(\mathbf{s}))$ , and  $\mathbf{p}$  respectively, and  $\boldsymbol{\lambda}^*(\mathbf{s}) \in \mathbb{R}_+^m$  associated with the spending constraint for  $(\mathbf{X}(\mathbf{s}), \beta(\mathbf{s}))$  such that:

$$q_j - \sum_{i \in [n]} \lambda_i^*(\mathbf{s}) x_{ij}^*(\mathbf{s}) - \nu_j^*(\mathbf{s}) \doteq 0 \quad \forall j \in [m] \quad (92)$$

$$\frac{b_i - \beta_i^*(\mathbf{s})}{u_i(\mathbf{x}_i^*(\mathbf{s}))} \frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i) - \lambda_i^*(\mathbf{s}) p_j(\mathbf{s}) + \mu_{ij}^*(\mathbf{s}) \doteq 0 \quad \forall i \in [n], j \in [m] \quad (93)$$

$$-\log(u_i(\mathbf{x}_i^*(\mathbf{s}))) + \gamma \frac{\partial}{\partial \beta_i} \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}')} [v(\mathbf{T}', \mathbf{b}' + \beta^*(\mathbf{s}), \mathbf{q}')] - \lambda_i^*(\mathbf{s}) + \mu_{i(m+1)}^*(\mathbf{s}) \doteq 0 \quad \forall i \in [n] \quad (94)$$

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

$$p_j(\mathbf{s}) > 0 \implies q_j - \sum_{i \in [n]} \lambda_i^*(\mathbf{s})x_{ij}^*(\mathbf{s}) = 0 \quad \forall j \in [m] \quad (95)$$

$$x_{ij}^*(\mathbf{s}) > 0 \implies \frac{b_i - \beta_i^*(\mathbf{s})}{u_i(\mathbf{x}_i^*(\mathbf{s}))} \frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s})) - \lambda_i^*(\mathbf{s})p_j(\mathbf{s}) = 0 \quad \forall i \in [n], j \in [m] \quad (96)$$

$$\begin{aligned} \beta_i^*(\mathbf{s}) > 0 \implies & -\log(u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)) + \gamma \frac{\partial}{\partial \beta_i}(\mathbf{T}', \mathbf{b}', \mathbf{q}') \mathbb{E} [v(\mathbf{T}', \mathbf{b}' + \boldsymbol{\beta}^*(\mathbf{s}), \mathbf{q}')] \\ & - \lambda_i^*(\mathbf{s}) + \mu_{i(m+1)}^*(\mathbf{s}) = 0 \quad \forall i \in [n] \quad (97) \end{aligned}$$

Re-organizing expressions, we obtain:

$$p_j(\mathbf{s}) > 0 \implies q_j = \sum_{i \in [n]} \lambda_i^*(\mathbf{s})x_{ij}^*(\mathbf{s}) \quad \forall j \in [m] \quad (98)$$

$$x_{ij}^*(\mathbf{s}) > 0 \implies \frac{u_i(\mathbf{x}_i^*(\mathbf{s}))}{b_i - \beta_i^*(\mathbf{s})} \lambda_i^*(\mathbf{s}) = \frac{\frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s}))}{p_j(\mathbf{s})} \quad \forall i \in [n], j \in [m] \quad (99)$$

$$\begin{aligned} \beta_i^*(\mathbf{s}) > 0 \implies & -\log(u_i(\mathbf{x}_i^*(\mathbf{s}))) + \gamma \frac{\partial}{\partial \beta_i}(\mathbf{T}', \mathbf{b}', \mathbf{q}') \mathbb{E} [v(\mathbf{T}', \mathbf{b}' + \boldsymbol{\beta}^*(\mathbf{s}), \mathbf{q}')] \\ & - \lambda_i^*(\mathbf{s}) = 0 \quad \forall i \in [n] \quad (100) \end{aligned}$$

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

$$\frac{\partial v}{\partial b_i}(\mathbf{s}) = \log(u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)) + \lambda_i^*(\mathbf{s}) \quad (101)$$

Once again note that  $\frac{\partial v}{\partial b_i}$  is well-defined by Equation 99.

Re-organizing expressions, we get:

$$\lambda_i^*(\mathbf{s}) = \frac{\partial v}{\partial b_i}(\mathbf{s}) - \log(u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)) \quad (102)$$

Combining Equation 102 and Equation 100, we obtain:

$$\begin{aligned} \beta_i^*(\mathbf{s}) > 0 \implies & -\log(u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)) + \gamma \frac{\partial}{\partial \beta_i}(\mathbf{T}', \mathbf{b}', \mathbf{q}') \mathbb{E} [v(\mathbf{T}', \mathbf{b}' + \boldsymbol{\beta}^*(\mathbf{s}), \mathbf{q}')] \\ & - \frac{\partial v}{\partial b_i}(\mathbf{s}) + \log(u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)) = 0 \quad \forall i \in [n] \quad (103) \end{aligned}$$

$$\beta_i^*(\mathbf{s}) > 0 \implies \gamma \frac{\partial}{\partial \beta_i}(\mathbf{T}', \mathbf{b}', \mathbf{q}') \mathbb{E} [v(\mathbf{T}', \mathbf{b}' + \boldsymbol{\beta}^*(\mathbf{s}), \mathbf{q}')] - \frac{\partial v}{\partial b_i}(\mathbf{s}) = 0 \quad \forall i \in [n] \quad (104)$$

$$\beta_i^*(\mathbf{s}) > 0 \implies \frac{\partial v}{\partial b_i}(\mathbf{s}) = \gamma \frac{\partial}{\partial \beta_i}(\mathbf{T}', \mathbf{b}', \mathbf{q}') \mathbb{E} [v(\mathbf{T}', \mathbf{b}' + \boldsymbol{\beta}^*(\mathbf{s}), \mathbf{q}')] \quad \forall i \in [n] \quad (105)$$

Going back to Equation 93, multiplying both sides by  $x_{ij}^*(\mathbf{s})$  and summing up across all  $j \in [m]$ , we get:

$$\sum_{j \in [m]} \frac{b_i - \beta_i^*(\mathbf{s})}{u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)} \sum_{j \in [m]} x_{ij}^*(\mathbf{s}) \frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s})) - \lambda_i^*(\mathbf{s}) \sum_{j \in [m]} p_j(\mathbf{s})x_{ij}^*(\mathbf{s}) = 0 \quad (106)$$

$$\frac{b_i - \beta_i^*(\mathbf{s})}{u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)} u_i(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i) - \lambda_i^*(\mathbf{s}) \sum_{j \in [m]} p_j(\mathbf{s})x_{ij}^*(\mathbf{s}) = 0 \quad (\text{Euler's Theorem}) \quad (107)$$

$$b_i - \beta_i^*(\mathbf{s}) - \lambda_i^*(\mathbf{s}) \sum_{j \in [m]} p_j(\mathbf{s})x_{ij}^*(\mathbf{s}) = 0 \quad (108)$$

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

$$b_i - \beta_i^*(\mathbf{s}) - \lambda_i^*(\mathbf{s})(b_i - \beta_i^*(\mathbf{s})) = 0 \quad (109)$$

$$\lambda_i^*(\mathbf{s}) = 1 \quad (110)$$

Combining the above with Equations (98) to (100) we obtain:

$$p_j(\mathbf{s}) > 0 \implies q_j = \sum_{i \in [n]} x_{ij}^*(\mathbf{s}) \quad \forall j \in [m] \quad (111)$$

$$x_{ij}^*(\mathbf{s}) > 0 \implies \frac{u_i(\mathbf{x}_i^*(\mathbf{s}))}{b_i - \beta_i^*(\mathbf{s})} = \frac{\frac{\partial u_i}{\partial x_{ij}}(\mathbf{x}_i^*(\mathbf{s}); \mathbf{t}_i)}{p_j(\mathbf{s})} \quad \forall i \in [n], j \in [m] \quad (112)$$

$$\beta_i^*(\mathbf{s}) > 0 \implies \frac{\partial v}{\partial b_i}(\mathbf{b}) = \gamma \frac{\partial}{\partial \beta_i} \mathbb{E}_{(\mathbf{T}', \mathbf{b}', \mathbf{q}')} [v(\mathbf{T}', \mathbf{b}' + \beta^*(\mathbf{s}), \mathbf{q}')] \quad \forall i \in [n] \quad (113)$$

Since the utility functions are non-satiated, and by the second equation, the buyers are utility maximizing at state  $\mathbf{s}$  over all allocations, we must also have that Walras' law holds, i.e.,  $\mathbf{p} \cdot \left( \mathbf{q} - \sum_{i \in [n]} \mathbf{x}_i \right) - \sum_{i \in [n]} \beta_i = 0$ . 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.  $\square$

## E Experiment Details

### E.1 Stochastic Fisher market without interest rates

For the without interest rates setup, we initialized a stochastic Fisher market with  $n = 2$  buyers and  $m = 2$  goods. To simplify the analysis, we assumed deterministic transitions such that the buyers get a constant new budgets of 9.5 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  $\theta_i$  from that range uniformly at random at the beginning of the experiment. (We scaled the valuations differently for different markets to ensure positive utilities though.) We have assumed the supply of goods is  $\mathbf{1}_m$  and that the budget space was  $[9, 10]^n$ . This means that our state space for our experiments was  $\mathcal{S} = \{(\theta_1, \theta_2)\} \times \{\mathbf{1}_m\} \times [9, 10]^n$ . 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  $\mathbf{b}^{(0)} = 10_n$  for buyers.

Since the state space is continuous, the value function has continuous domain 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  $v(\mathbf{T}, \mathbf{b}, \mathbf{q}; \mathbf{a}, c) = \mathbf{a}^T \mathbf{b} + c$  for some parameters  $\mathbf{a} \in \mathbb{R}^n, c \in \mathbb{R}$ , 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  $[9, 10]^n$ . Next, for each sampled budget  $\mathbf{b}$ , 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 the **nested gradient descent ascent (GDA)** [6] (Algorithm 3) along with JAX gradients, which is not guaranteed to converge to a global optimum since the min-max Stackelberg game for stochastic Fisher markets is convex-non-concave. Then, we have run value iteration for 30 iterations. We ran nested GDA with learning rates  $\eta_X = 1.4, \eta_p = 1.5 \times 10^{-2}$  for linear,  $\eta_X = 1.5, \eta_p = 6.5 \times 10^{-4}$  for leontief, and  $\eta_X = 1.4, \eta_p = 5 \times 10^{-3}$  for Cobb-Douglas. The outer loop of nested GDA was run for  $T_p = 60$

iterations, while its inner loop was run for  $T_X = 100$  iterations, and break from the nested GDA if we obtain an excess demand with norm lower than 0.01. We depict the trajectory of the average value of the value function at each iteration of value iteration under nested GDA in Figure 1.

## E.2 Stochastic Fisher market with interest rates

For the with interest rates setup, we initialized a stochastic Fisher market with  $n = 5$  buyers and  $m = 5$  goods. This time, we implemented a stochastic transitions. Though buyers still get a constant new budgets of 9.5 at each time step, and their types/valuations as well as the supply of goods does not change at each state, their savings from last time step may decrease or increase according to some probabilistic interest rates. In specific, at each time step, we consider five interest rates  $\{0.9, 1.0, 1.1, 1.2, 1.5\}$ , each with probability 0.2. Thus, we have a stochastic market setting in which the amount of budget possessed by the buyers at the beginning of each time step differentiates different states of the market. To initialize the state space of the market, we first fixed a range of  $[0, 1]^m$  for the buyers' valuations and drew for all buyers  $i \in [n]$  valuations  $\theta_i$  from that range uniformly at random at the beginning of the experiment. (We scaled the valuations differently for different markets to ensure positive utilities though.) We have assumed the supply of goods is  $\mathbf{1}_m$  and that the budget space was  $[9, 10]^n$ . This means that our state space for our experiments was  $\mathcal{S} = \{(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)\} \times \{\mathbf{1}_m\} \times [9, 10]^n$ . 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  $\mathbf{b}^{(0)} = 10_n$  for buyers.

Since the state space is continuous, the value function has continuous domain 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  $v(\mathbf{T}, \mathbf{b}, \mathbf{q}; \mathbf{a}, c) = \mathbf{a}^T \mathbf{b} + c$  for some parameters  $\mathbf{a} \in \mathbb{R}^n, c \in \mathbb{R}$ , 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  $[9, 10]^n$ . Next, for each sampled budget  $\mathbf{b}$ , 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 the **nested gradient descent ascent (GDA)** [6] (Algorithm 3) along with JAX gradients, which is not guaranteed to converge to a global optimum since the min-max Stackelberg game for stochastic Fisher markets is convex-non-concave. Then, we have run value iteration for 30 iterations. We ran nested GDA with learning rates  $\eta_X = 1.7, \eta_p = 2 \times 10^{-2}$  for linear,  $\eta_X = 2, \eta_p = 5 \times 10^{-5}$  for leontief, and  $\eta_X = 1.8, \eta_p = 2.5 \times 10^{-2}$  for Cobb-Douglas. The outer loop of nested GDA was run for  $T_p = 60$  iterations, while its inner loop was run for  $T_X = 100$  iterations, and break from the nested GDA if we obtain an excess demand with norm lower than 0.01. We depict the trajectory of the average value of the value function at each iteration of value iteration under nested GDA in Figure 2.

## E.3 Other Details

**Programming Languages, Packages, and Licensing** We ran our experiments in Python 3.7 [72], using NumPy [73], CVXPY [74], and JAX [63]. Figure 1 and Figure 2 were graphed using Matplotlib [75].

Python software and documentation are licensed under the PSF License Agreement. Numpy is distributed under a liberal BSD license. Matplotlib only uses BSD compatible code, and its license is based on the PSF license. CVXPY is licensed under an APACHE license.

**Implementation Details** In our execution of Algorithm 3 in order to project each allocation computed onto the budget set of the consumers, i.e.,  $\{\mathbf{X} \in \mathbb{R}_+^{n \times m} \mid \mathbf{X}\mathbf{p} \leq \mathbf{b}\}$ , we used the CVXPY with MOSEK solver with warm start option, a feature that enables the solver to exploit work from previous solves.

**Computational Resources** Our experiments were run on Google Colab with 12.68GB RAM, and took about 8 hours to run the Stochastic Fisher market without interest rates experiment (for each

utility function class) and about 8.5 hours to run the Stochastic Fisher market with interest rates experiment (for each utility function class). Only CPU resources were used.

**Code Repository** The data our experiments generated, and the code used to produce our visualizations, can be found in our code repository (<https://github.com/Sadie-Zhao/Zero-Sum-Stochastic-Stackelberg-Games-NeurIPS>).