Discounted Robust Stochastic Games and an Application to Queueing Control

Erim Kardeş
University of Southern California, kardes@usc.edu,

Fernando Ordóñez
University of Southern California, fordon@usc.edu,

Randolph W. Hall
University of Southern California, rwhall@usc.edu,

This paper presents a robust optimization model for n-person finite state/action discounted stochastic games with incomplete information. We consider n-player, non-zero sum discounted stochastic games in which none of the players knows the true data of the game and each player considers a distribution-free incomplete information stochastic game to be played using robust optimization. We call such games "*discounted robust stochastic games*". Discounted robust stochastic games allow us to use simple uncertainty sets for the unknown data of the game, and eliminate the need to have an a-priori probability distribution over a set of games. We prove the existence of equilibrium points and propose an explicit mathematical programming formulation for an equilibrium calculation. We illustrate the use of discounted robust stochastic games in a single server queueing control problem.

Subject classifications: Games: stochastic; dynamic programming/optimal control: Markov: finite state; queues: optimization.
Area of review: Optimization
History:

1. Introduction

There has been an extensive body of research on stochastic games in various fields including operations research, mathematics, and economics since the 1950s. The first paper on finite state/action, discrete time, two person zero-sum stochastic games was introduced by Shapley (1953). Many extensions to this basic model have been proposed after this seminal paper such as games with infinite states and actions, n-person games, games with incomplete information, continuous time games, and semi-Markov games among others.

Our research is motivated by the fact that in many practical applications, the data of the game is not known exactly, but it is known that it belongs to a set over which we may not have probabilistic information. Futhermore, solutions to stochastic games may be very sensitive to payoff and transition probability data for which the estimates could be inaccurate. In this paper, we focus on finite state/action, discrete time, non-zero sum, discounted stochastic games in which the payoffs and/or the transition probabilitites among states are ambigious and players adopt a robust optimization approach to cope with the ambiguity. Here, we use the conventional terminology used in decision analysis where ambiguity refers to quantities with unknown probabilities, unlike the term "uncertainty" that refers to random quantities with known probability distributions.

Since it considers two or more competing decision makers acting as adversaries, a stochastic game could be viewed as a generalization of a Markov Decision Process (MDP) (Filar and Vrieze (1997)). They could also be viewed as a collection of auxillary one-shot matrix games (Shapley (1953)). Hence, two lines of related research in the literature are MDPs and one-shot games. Some authors have addressed the issue of ambiguity in the transition probabilitites of MDPs. A Bayesian approach is presented by Shapiro and Kleywegt (2002), where a prior distribution on the transition matrix should be known. Satia and Lave (1973), White and Eldeib (1994), and Givan et al. (1997) have modeled an MDP where the transition matrix lies in a given set, which is most typically a polytope. Nilim and Ghaoui (2005) considers robust control in MDPs where a proof of the robust value iteration is presented. Bagnell et al. (2001) considers a similar problem and present the robust value iteration without proof. Iyengar (2005) considers robust dynamic programming problems and provides an independent proof of the robust value iteration. It is important to note here that these recent efforts consider the robustness in the context of MDPs, where an opponent player is not modeled explicitly.

Using a worst-case approach has been prevalent in game theory since the "max-min" formulation of von Neumann's and Morgenstern's. For instance, Gilboa and Schmeidler (1989), Lo (1996), and Marinacci (2000) have presented a max-min approach to cope with ambiguous uncertainty in normal form games. Although these authors adopt a worst-case approach, their models are fundamentally probabilistic and are based on prior probability distributions. Futhermore these authors address complete information games and adopt a worst-case approach with respect to players' behaviors towards each other, rather than adressing ambiguity in the data of a game. Harsanyi (1967,1968) modeled incomplete information games by considering that each player could use a prior probability distribution to obtain a conditional distribution on the data of the game unknown to himself. Unlike these approaches, a robust optimization approach to ambiguos payoff uncertainty in one shot games is considered by Aghassi and Bertsimas (2006). In that work, the authors prove the existence of a robust equilibrium and formulate the robust game by considering that the payoffs belong to a polytope, yielding a method to compute an equilibrium point.

The incomplete information case within the repeated games is first introduced by Aumann and Maschler (1968). Sorin (1984) and Sorin (1985) consider stochastic games with incomplete information on one side that have a single nonabsorbing state. It is proven in that paper that these games have a min-max and a max-min value. However, there is no explicit computational scheme for these values. In a more recent effort, Rosenberg et al. (2004) consider two-player zero-sum stochastic games with incomplete information where the incomplete information is described by a finite collection of stochastic games, and a game is to be played out of the finite set of games over which a probability distribution is specified. That paper focuses on stochastic games in which one player controls the transitions. In other words, they consider that the evolution of the game is independent of one of the opponents' actions and only depends on one player's actions. We note that the approach adopted in Rosenberg et al. (2004) is based on the approach proposed in Harsanyi (1967,1968), and requires a probability distribution over set of games.

In this paper, we consider n-player, non-zero sum discounted stochastic games in which none of the players knows the true transition probabilities and/or payoffs of the game and each player uses a robust optimization approach to address this ambiguity. We call such games "*robust stochastic games*". We propose a distribution-free model for discounted stochastic games with incomplete information relaxing the assumptions on former efforts and present an explicit mathematical programming formulation for equilibrium calculation. Specifically, in section 2, we review basic ideas on discounted stochastic games and robust optimization, followed by the formulation of discounted robust stochastic games. In section 3, we prove the existence of a robust equilibrium point. In section 4, we show that when the ambiguity in the transition data comes from a polytope instersected with the probability simplex, the robust equilibrium can be formulated as a feasibility problem, the solution of which gives an equilibrium point of the discounted robust stochastic game. We also show in this section that if there is data ambiguity in the game, the approach players adopt for this ambiguity may differ, resulting in different ambiguity perspectives for the players. In section 5, we present an application of discounted robust stochastic games to queueing control. Finally, section 6 concludes the paper with remarks and future research directions.

2. Problem Setup

2.1. Stochastic Games

This section reviews basics of stochastic game theory, as presented in Shapley (1953) and Fink (1964). In stochastic games, the play proceeds from one state to the other according to transition probabilities controlled jointly by two or more players. It consists of states and actions associated with each player. Once the game starts in a state, each player chooses their respective actions. The play then moves into the next state with some probability and continues from thereon. The probability that the game moves into the next state is determined by the current state and the actions chosen in the current state.

Let the set of states $S = \{1, ..., M\}$ and the set of players $I = \{1, ..., N\}$ be finite. If the play is in state s, player i can choose the action $a_s^i \in A_s^i$, where A_s^i of cardinality m_s^i is the set of actions of player i in state s. Suppose that each player makes a choice in state s, i.e., we have an action tuple $a_s = (a_s^1, ..., a_s^i, ..., a_s^N) \in A_s$, the set of all possible action tuples in state s. Then the game moves into state k with probability $P_{sa_sk} \geq 0$, $\sum_{k=1}^{M} P_{sa_sk} = 1$.

At each stage, players may consider to use mixed strategies. Let x_s^i be the probability distribution over the set A_s^i. In other words, the probability vector for player i in state s is $x_s^i = (x_{s,1}^i, ..., x_{s,m_s^i}^i)$, where $x_{s,k}^i \geq 0, \sum_{k=1}^{m_s^i} x_{s,k}^i = 1$. If we denote the set of mixed strategies of player i in state s by X_s^i, then X_s^i is the m_s^i-dimensional probability simplex:

$$X_s^i = \{x_s^i \in \Re_+^{m_s^i} | \sum_{k=1}^{m_s^i} x_{s,k}^i = 1\}.$$

We consider a certain class of strategies as introduced by Shapley (1953), namely, stationary strategies. Stationary strategies prescribe a player the same probabilities for her choices each time the player visits a certain state, no matter what route she follows to reach that state. Let us represent the stationary strategies of a player i by $x^i = (x_1^i, ..., x_M^i)$ and denote the set of mixed strategies of all players in the state space of the game by $x = (x^1, ..., x^N)$. We denote the mixed strategies of all players for all states except for player i by $x^{-i} = (x^1, ..., x^{i-1}, x^{i+1}, ..., x^N)$. The following notation is used to distinguish a mixed strategy of player i from those of others: $(x^{-i}, u^i) = (x^1, ..., x^{i-1}, u^i, x^{i+1}, ..., x^N)$. Finally, we use the following notation. $X^i = \prod_{s \in S} X_s^i$, $X_s = \prod_{i \in I} X_s^i$, and $X = \prod_{i \in I} X^i$.

Let $\{R_t^i\}_{t=0}^{\infty}$ denote the sequence of costs to player i throughout the process. The expected cost at stage t to player i resulting from the strategy $x \in X$ and the initial state s is denoted by $E_{s,x}(R_t^i)$. The discounted value of a strategy $x \in X$ to player i is defined by

$$v_\beta^i(s,x) := \sum_{t=0}^{\infty} \beta^t E_{s,x}(R_t^i),$$

where $0 \leq \beta < 1$ is the discount factor. To ease the notation, given $x^{-i} \in X^{-i}$, we denote the value of the $\beta-$discounted stochatic game to player i starting in state s by v_s^i. If the need arises, we denote the $\beta-$discounted value to player i in state s corresponding to $x \in X$ by $v_s^i(x)$.

Let $C_{sa_s}^i$ be the immediate cost to player i induced by a_s in state s. Suppose that players minimize their expected overall costs throughout the process and that they choose their actions independently of each other at a given state. It is shown in Fink (1964) that $\forall i \in I, s \in S$, given x^{-i}, the value v_s^i to player i in state s, satisfies

$$v_s^i = \min_{u_s^i \in X_s^i} \sum_{a_s \in A_s} \prod_{\substack{m=1 \\ m \neq i}}^{N} x_{s,a_s^m}^m u_{s,a_s^i}^i \{C_{sa_s}^i + \beta_i \sum_{k=1}^{M} P_{sa_sk} v_k^i\}. \tag{1}$$

It is in fact another result that, for any $x = (x^1, ..., x^N) \in X$, there exists a unique corresponding value v_s^i, $\forall i \in I, \forall s \in S$. We are now ready for the following definition.

DEFINITION 1. A tuple of stationary strategies $x = (x^1, ..., x^N) \in X$ is a *Nash equilibrium point* in a stochastic game if and only if, $\forall i \in I$ and $\forall s \in S$, $v_s^i(x^1, ..., x^N) \leq v_s^i(x^{-i}, u^i), \forall u^i \in X^i$.

The interpretation of equation (1) is that if a player knew how to play optimally from the next stage on, then, at the current stage, he would select the strategy that minimizes the expected immediate cost at the current stage plus the total expected future costs. Hence, player i is not only concerned with the immediate outcome of his actions but also with the future consequences of his strategies in the current stage.

We next state an equivalent equilibrium definition for the purposes of developments in what follows. Let $g_s^i(x_s^{-i}, u_s^i; v^i) = \sum_{a_s \in A_s} \prod_{\substack{m=1 \\ m \neq i}}^{N} x_{s,a_s^m}^m u_{s,a_s^i}^i \{C_{sa_s}^i + \beta_i \sum_{k=1}^{M} P_{sa_sk} v_k^i\}$.

DEFINITION 2. A point $x \in X$ is a Nash equilibrium in a stochastic game if and only if, $\exists\ v = (v^1, ..., v^N)$, such that, $\forall i \in I, \forall s \in S$,

$$v_s^i = \min_{u_s^i \in X_s^i} g_s^i(x_s^{-i}, u_s^i; v^i) \quad \text{and} \quad x_s^i \in \text{argmin}_{u_s^i \in X_s^i} g_s^i(x_s^{-i}, u_s^i; v^i). \tag{2}$$

This definition states that x_s^i is an optimal (stationary) strategy for player i in state s if, when equation (1) is satisfied, the corresponding minimizer of the objective function of player i is the strategy that other players expect player i to use. If this statement holds for all players and all states, then no player would wish to deviate from their strategies, resulting in an equilibrium.

2.2. Robust Optimization

This section briefly reviews the basics of robust optimization, as introduced in Ben-Tal and Nemirovski (1998). Consider the following optimization problem $P_\gamma : \min_{x \in \Re^n} f(x, \gamma)$ s.t. $F(x, \gamma) \in K \subset \Re^m$, where $\gamma \in \Re^M$ is the data vector, $x \in \Re^n$ is the decision vector, and K is a convex cone. Suppose that the data of P_γ is uncertain and all that is known about the data is that it belongs to an uncertainty set $U \in \Re^M$. Now, consider the problem $P = \{P_\gamma\}_{\gamma \in U}$, where the constraints $F(x, \gamma) \in$ K must be satisfied no matter what the actual realization of $\gamma \in U$ is. An optimal solution to the uncertain problem P is defined as a solution that must give the best possible guaranteed value under all possible realizations of constraints. Formally, it should be an optimal solution of the following program: $P_R: \ \min_{x \in \Re^n} \{\sup_{\gamma \in U} f(x, \gamma) \ \ s.t. \ \ F(x, \gamma) \in K, \ \ \forall \gamma \in U\}$. Problem P_R is called the robust counterpart of P, and its feasible and optimal solutions are called robust feasible and robust optimal solutions, respectively Ben-Tal and Nemirovski (1998). Prior work, e.g. Bertsimas and Sim (2004), Ben-Tal and Nemirovski (1998) has also shown that for many function types and uncertainty sets, the robust counterpart problem P_R can be solved as a single optimization problem of size comparable to a deterministic version of the problem.

2.3. Formulation of Robust Stochastic Games

In this section, we formalize our robust model for incomplete information stochastic games and the robust equilibrium concept by considering that both payoffs and transition probabilities of the game belong to respective uncertainty sets. In discounted robust stochastic games, it is assumed that the players commonly know the uncertainty set of payoffs C_s at each state and the set of transition probabilities P_s out of each state. Unlike the approach in Rosenberg et al. (2004), players need not have distributional information on the uncertainty sets with respect to which they adopt a worst-case approach. Now, in light of the results summarized in the previous section, we notice the following: If a player knew how to play in the robust stochastic game optimally from the next stage on, then, at the current stage, she would play with such strategies so that she *minimizes the maximum expected immediate cost* at the current stage and also *minimizes the maximum expected costs possibly incurred in future stages*. Hence, *if* optimal robust values for player i exist, given x^{-i}, they must satisfy the following,

$$\omega_s^i = \min_{u_s^i \in X_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \sum_{a_s \in A_s} \prod_{\substack{m=1 \\ m \neq i}}^{N} x_{s,a_s^m}^m u_{s,a_s^i}^i \{\tilde{C}_{sa_s} + \beta_i \sum_{k=1}^{M} \tilde{P}_{sa_s k} \omega_k^i\}, \tag{3}$$

where the inner maximization problem is with respect to the ambiguous transition probabilities and immediate costs. Note that we could have modeled each player as wishing to minimize her expected maximum total cost, rather than her maximum expected total cost. We use the latter model for the following reasons. In the former case, players would have the advantage of observing each others' randomized actions before they adopt their own perspectives on the ambiguous data. In the latter case, the adversaries do not have this advantage and the worst-case perspective on the ambiguity is considered with respect to the mixed strategies. Note that while the latter results in a pessimistic approach, the former model would be overly pessimistic.

To ease the notation, let us define

$$\psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i) = \sum_{a_s \in A_s} \prod_{\substack{m=1 \\ m \neq i}}^{N} x_{s,a_s^m}^m u_{s,a_s^i}^i \{\tilde{C}_{sa_s} + \beta_i \sum_{k=1}^{M} \tilde{P}_{sa_s k} \omega_k^i\}.$$

Equation (3) now reads as follows.

$$\omega_s^i = \min_{u_s^i \in X_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i). \tag{4}$$

We will in fact show that such robust values exist. Similar to condition (2), we are now ready to state our definition of equilibrium in discounted robust stochastic games.

DEFINITION 3. A point x is a *robust equilibrium* point in a robust stochastic game if and only if, $\exists\ \omega = (\omega^1, ..., \omega^N)$, such that, $\forall i \in I, \forall s \in S$,

$$\omega_s^i = \min_{u_s^i \in X_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i) \tag{5}$$

$$x_s^i \in \operatorname{argmin}_{u_s^i \in X_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i). \tag{6}$$

Equivalently, a tuple of strategies $x = (x^1, ..., x^N)$ is a robust equilibrium point in a robust stochastic game if and only if, $\forall i \in I$ and $\forall s \in S$, $w_s^i(x^1, ..., x^N) \leq w_s^i(x^{-i}, u^i), \forall u^i \in X^i$.

3. Existence of Equilibrium

Our proof of existence of equilibrium points in discounted robust stochastic games parallels Fink's (1964). However, a different point-to-set mapping (correspondence) is defined that takes into account the robustness. This mapping uses a maximum expected total cost function with respect to mixed strategies. We show that the fixed point of this suitably constructed correspondence is an equilibrium point.

Let $W^i \equiv \{\omega_s^i \in \Re\}_{s \in S}$, $W \equiv \{W^i\}_{i \in I}$. The infinity norm on W is defined as follows: $||\omega - \theta||_\infty = \max_{i \in I, s \in S} |\omega_s^i - \theta_s^i|$.

Given the strategies of all other players, x_s^{-i}, we define a mapping below that takes any robust value vector for player i, ω^i, and minimizes the maximum expected total cost with respect to the mixed strategies for player i. The next theorem shows that this mapping is a contraction mapping and Theorem 2 below shows that such a robust value vector exists for any given x_s^{-i}.

THEOREM 1. *Let* $\gamma^i_{s,x_s^{-i}} : W^i \rightarrow \Re$ *be defined by*

$$\gamma^i_{s,x_s^{-i}}(\omega^i) = \min_{u_s^i \in X_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi^i_s(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i).$$

For $x \in X$, *define* $\gamma_x(\omega) : W \rightarrow W$ *by* $(\gamma_x(\omega))_{is} = \gamma^i_{s,x_s^{-i}}(\omega^i)$. *The function* $\gamma_x(\omega)$ *is a contraction mapping.*

Proof.
Let $\omega, \theta \in W$. For x_s^{-i} fixed, $\forall i \in I, s \in S$,

$$\begin{aligned}\gamma^i_{s,x_s^{-i}}(\omega^i) &= \min_{u_s^i \in X_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi^i_s(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i) \\ &= \psi^i_s(C^i_s(x_s^{-i}, u_s^{*i}), P^i_s(x_s^{-i}, u_s^{*i}, \omega^i); x_s^{-i}, u_s^{*i}; \omega^i),\end{aligned}$$

where u_s^{*i} is the minimizer, and $C^i_s(x_s^{-i}, u_s^{*i}) \in C_s$ and $P^i_s(x_s^{-i}, u_s^{*i}, \omega^i) \in P_s$ are the optimizers that now depend on (x_s^{-i}, u_s^{*i}). Similarly, with z_s^{*i} and $C^i_s(x_s^{-i}, z_s^{*i}) \in C_s, P^i_s(x_s^{-i}, z_s^{*i}, \theta^i) \in P_s$, we have

$$\begin{aligned}\gamma^i_{s,x_s^{-i}}(\theta^i) &= \min_{z_s^i \in X_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi^i_s(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, z_s^i; \theta^i) \\ &= \psi^i_s(C^i_s(x_s^{-i}, z_s^{*i}), P^i_s(x_s^{-i}, z_s^{*i}, \theta^i); x_s^{-i}, z_s^{*i}; \theta^i).\end{aligned}$$

Now,

$$\gamma^i_{s,x_s^{-i}}(\omega^i) - \gamma^i_{s,x_s^{-i}}(\theta^i)$$

$$\begin{aligned}
&= \psi^i_s(C^i_s(x_s^{-i}, u_s^{*i}), P^i_s(x_s^{-i}, u_s^{*i}, \omega^i); x_s^{-i}, u_s^{*i}; \omega^i) - \psi^i_s(C^i_s(x_s^{-i}, z_s^{*i}), P^i_s(x_s^{-i}, z_s^{*i}, \theta^i); x_s^{-i}, z_s^{*i}; \theta^i) \\
&\leq \psi^i_s(C^i_s(x_s^{-i}, z_s^{*i}), P^i_s(x_s^{-i}, z_s^{*i}, \omega^i); x_s^{-i}, z_s^{*i}; \omega^i) - \psi^i_s(C^i_s(x_s^{-i}, z_s^{*i}), P^i_s(x_s^{-i}, z_s^{*i}, \theta^i); x_s^{-i}, z_s^{*i}; \theta^i) \\
&= \sum_{a_s \in A_s} \prod_{\substack{m=1 \\ m \neq i}}^{N} x^m_{s,a_s^m} z_s^{*i} \{C^i_{sa_s}(x_s^{-i}, z_s^{*i}) + \beta \sum_{k=1}^{M} P^i_{sa_sk}(x_s^{-i}, z_s^{*i}, \omega^i_k)\omega^i_k\} \\
&\qquad - \sum_{a_s \in A_s} \prod_{\substack{m=1 \\ m \neq i}}^{N} x^m_{s,a_s^m} z_s^{*i} \{C^i_{sa_s}(x_s^{-i}, z_s^{*i}) + \beta \sum_{k=1}^{M} P^i_{sa_sk}(x_s^{-i}, z_s^{*i}, \theta^i_k)\theta^i_k\} \\
&\leq \sum_{a_s \in A_s} \prod_{\substack{m=1 \\ m \neq i}}^{N} x^m_{s,a_s^m} z_s^{*i} \beta \{\sum_{k=1}^{M} P^i_{sa_sk}(x_s^{-i}, z_s^{*i}, \omega^i_k)(\omega^i_k - \theta^i_k)\} \\
&\leq \sum_{a_s \in A_s} \prod_{\substack{m=1 \\ m \neq i}}^{N} x^m_{s,a_s^m} z_s^{*i} \beta (\sum_{k=1}^{M} P^i_{sa_sk}(x_s^{-i}, z_s^{*i}, \omega^i_k)) ||\omega - \theta||_\infty = \beta ||\omega - \theta||_\infty.
\end{aligned}$$

The second to the last inequality above follows from the fact that

$$\sum_{a_s \in A_s} \prod_{\substack{m=1 \\ m \neq i}}^{N} x^m_{s,a_s^m} z_s^{*i} \sum_{k=1}^{M} P^i_{sa_sk}(x_s^{-i}, z_s^{*i}, \omega^i_k)\theta^i_k \leq \sum_{a_s \in A_s} \prod_{\substack{m=1 \\ m \neq i}}^{N} x^m_{s,a_s^m} z_s^{*i} \sum_{k=1}^{M} P^i_{sa_sk}(x_s^{-i}, z_s^{*i}, \theta^i_k)\theta^i_k,$$

because for a given $(x_s^{-i}, z_s^{*i}, \theta^i_k)$, $[P^i_{sa_sk}(x_s^{-i}, z_s^{*i}, \theta^i_k)]_{k=1,\ldots,M}$ is the maximizer of $\psi^i_s(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, z_s^i; \theta^i)$ over $\tilde{P}_s \in P_s$.

Similar to the above arguments, we have for x_s^{-i} fixed that , $\forall i \in I, s \in S$,

$$\gamma^i_{s,x_s^{-i}}(\theta^i) - \gamma^i_{s,x_s^{-i}}(\omega^i) \leq \beta ||\omega - \theta||_\infty.$$

Thus, $||\gamma_x(\omega) - \gamma_x(\theta)||_\infty \leq \beta ||\omega - \theta||_\infty$. □

Theorem 2. ***Application of Banach's Theorem.*** *For any $x \in X$, and $\forall i \in I, s \in S$, there exists a unique w_s^i such that $\omega_s^i = \min_{u_s^i \in S_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i)$.*

Proof.

Note that $(W, ||.||_\infty)$ is a complete metric space and by Theorem 1, $\gamma_x : W \to W$ is a contraction mapping. Therefore, by Banach's Theorem, $\gamma_x(\omega)$ has a unique fixed point, ω. That is, there exists a unique vector, ω, such that $\gamma_x(\omega) = \omega$, which means

$$\omega_s^i = \min_{u_s^i \in S_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i), \quad \forall i \in I, s \in S. \tag{7}$$

□

The above theorem states that for all players, states, and any given $x^{-i} \in X^{-i}$, a unique robust value vector w^i exists satisfying (7). This also implies that if $\beta < 1$, and we consider any fixed x_s^{-i}, applying the above transformation over and over again starting with an arbitrary robust value vector, will converge to the unique fixed point of the transformation.

We next state the definition of upper-semi continuity for correspondences and Kakutani's fixed point theorem (Kakutani (1941)).

Definition 4. A correspondence $\phi : S \to 2^S$ is upper semi-continuous if $y^n \in \phi(x^n)$, $\lim_{n\to\infty} x^n = x$, $\lim_{n\to\infty} y^n = y$ imply that $y \in \phi(x)$.

Theorem 3. *(Kakutani's Fixed Point Theorem). If S is a closed, bounded, and convex set in a Euclidean space, and ϕ is an upper semi-continuous correspondence mapping S into the family of closed, convex subsets of S, then $\exists \ x \in S, \ s.t. \ x \in \phi(x)$.*

Let $f_s^i(x_s^{-i}, u_s^i; \omega^i) = \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i)$.

Define the metrics $d_{X_s}(x_s, u_s) = \max_{i\in I} ||x_s^i - u_s^i||_\infty$, $d_{W^i}(w^i, \theta^i) = \max_{s\in S} ||w_s^i - \theta_s^i||_\infty$, and $d_1(p, q) = d_{X_s}(x_s, u_s) + d_{W^i}(w^i, \theta^i)$. We need the following lemma to show that f_s^i satisfies the properties needed to use Kakutani's theorem.

Lemma 1. *Let $p = (x_s, w^i)$, $q = (u_s, \theta^i)$.*

Given $\epsilon > 0$, $\exists \delta(\epsilon) > 0$ such that if for any $p, q \in X_s \times W^i$, $d_1(p, q) < \delta(\epsilon)$, then, $\forall \tilde{C}_s \in C_s$, $\forall \tilde{P}_s \in P_s$, $\left|\psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s, \omega^i) - \psi_s^i(\tilde{C}_s, \tilde{P}_s; u_s, \theta^i)\right| < \epsilon$.

Proof.

Since, $\tilde{C}_s \in C_s$ and C_s is bounded $\forall s \in S$, we have $\left|\tilde{C}_{sa_s}\right| \le K$, where $K < \infty$. It is clear that robust values are bounded. Hence, we have $\forall i \in I, s \in S$, that $|\omega_s^i| \le W$, where $W < \infty$. Note that

$$\left|\psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s, \omega^i) - \psi_s^i(\tilde{C}_s, \tilde{P}_s; u_s, \theta^i)\right| =$$

$$| \sum_{a_s \in A_s} \prod_{m=1}^{N} x_{s,a_s^m}^m \tilde{C}_{sa_s} + \beta_i \sum_{a_s \in A_s} (\prod_{m=1}^{N} x_{s,a_s^m}^m)(\sum_{k=1}^{M} \tilde{P}_{sa_sk} \omega_k^i)$$

$$- \sum_{a_s \in A_s} \prod_{m=1}^{N} u_{s,a_s^m}^m \tilde{C}_{sa_s} - \beta_i \sum_{a_s \in A_s} (\prod_{m=1}^{N} u_{s,a_s^m}^m)(\sum_{k=1}^{M} \tilde{P}_{sa_sk} \theta_k^i)|$$

$$= \left| \sum_{a_s \in A_s} \tilde{C}_{sa_s} (\prod_{m=1}^{N} x_{s,a_s^m}^m - \prod_{m=1}^{N} u_{s,a_s^m}^m) + \beta_i \sum_{a_s \in A_s} \sum_{k=1}^{M} \tilde{P}_{sa_sk} (\prod_{m=1}^{N} x_{s,a_s^m}^m \omega_k^i - \prod_{m=1}^{N} u_{s,a_s^m}^m \theta_k^i) \right|$$

$$\leq \left| \sum_{a_s \in A_s} \tilde{C}_{sa_s} (\prod_{m=1}^{N} x^m_{s,a^m_s} - \prod_{m=1}^{N} u^m_{s,a^m_s}) \right| + \left| \beta_i \sum_{a_s \in A_s} \sum_{k=1}^{M} \tilde{P}_{sa_sk} (\prod_{m=1}^{N} x^m_{s,a^m_s} \omega^i_k - \prod_{m=1}^{N} u^m_{s,a^m_s} \theta^i_k) \right|$$
$$\leq K \sum_{a_s \in A_s} \left| \prod_{m=1}^{N} x^m_{s,a^m_s} - \prod_{m=1}^{N} u^m_{s,a^m_s} \right| + \beta_i \sum_{a_s \in A_s} \sum_{k=1}^{M} \left| \prod_{m=1}^{N} x^m_{s,a^m_s} \omega^i_k - \prod_{m=1}^{N} u^m_{s,a^m_s} \theta^i_k \right|.$$

Let

$$\delta_1(\epsilon) = \frac{\min\{\epsilon, 1\}}{3K(2^N - 1)\prod_{i=1}^N m^i_s}, \delta_2(\epsilon) = \frac{\min\{\epsilon, 1\}}{3M\beta_i \prod_{i=1}^N m^i_s}, \quad \delta_3(\epsilon) = \frac{\min\{\epsilon, 1\}}{3WM\beta_i(2^N - 1)\prod_{i=1}^N m^i_s},$$

and let $\delta(\epsilon) = \min\{\delta_1(\epsilon), \delta_2(\epsilon), \delta_3(\epsilon)\}$. Now, $d_1(p, q) < \delta(\epsilon)$ implies that, $\forall i \in I, s \in S$, and $\forall a^i_s \in A^i_s$, $x^m_{s,a^m_s} = u^m_{s,a^m_s} + \alpha^m_{s,a^m_s}$ and $\omega^i_s = \theta^i_s + \gamma^i_s$, where $\left|\alpha^m_{s,a^m_s}\right| < \delta(\epsilon)$, and $|\gamma^i_s| < \delta(\epsilon)$. We will make use of the following algebraic identity.

$$\left| \prod_{m=1}^{N} (u^m_{s,a^m_s} + \alpha^m_{s,a^m_s}) - \prod_{m=1}^{N} u^m_{s,a^m_s} \right| = \left| \sum_{\substack{I \subseteq 1,\dots,N \\ |I| \geq 1}} (\prod_{m \in I} \alpha^m_{s,a^m_s})(\prod_{m \in I^C} u^m_{s,a^m_s}) \right|,$$

where $I^C = \{1, ..., N\} \backslash I$. Note that $\prod_{m \in I} \left|\alpha^m_{s,a^m_s}\right| < (\delta_1(\epsilon))^{|I|} \leq \delta_1(\epsilon)$, and that

$$\left| \prod_{m=1}^{N} (u^m_{s,a^m_s} + \alpha^m_{s,a^m_s}) - \prod_{m=1}^{N} u^m_{s,a^m_s} \right| \leq \sum_{\substack{I \subseteq 1,\dots,N \\ |I| \geq 1}} \left| \prod_{m \in I} \alpha^m_{s,a^m_s} \right| \left| \prod_{m \in I^C} u^m_{s,a^m_s} \right|.$$

Hence, we have

$$K \sum_{a_s \in A_s} \left| \prod_{m=1}^{N} (u^m_{s,a^m_s} + \alpha^m_{s,a^m_s}) - \prod_{m=1}^{N} u^m_{s,a^m_s} \right| \leq K \sum_{a_s \in A_s} \sum_{\substack{I \subseteq 1,\dots,N \\ |I| \geq 1}} \left| \prod_{m \in I} \alpha^m_{s,a^m_s} \right| \left| \prod_{m \in I^C} u^m_{s,a^m_s} \right|$$
$$\leq K \sum_{a_s \in A_s} \sum_{\substack{I \subseteq 1,\dots,N \\ |I| \geq 1}} \left| \prod_{m \in I} \alpha^m_{s,a^m_s} \right| < K \sum_{a_s \in A_s} \sum_{\substack{I \subseteq 1,\dots,N \\ |I| \geq 1}} \gamma_1(\epsilon) = \frac{\epsilon}{3}.$$

We also have

$$\beta_i \sum_{a_s \in A_s} \sum_{k=1}^{M} \left| \prod_{m=1}^{N} x^m_{s,a^m_s} \omega^i_k - \prod_{m=1}^{N} u^m_{s,a^m_s} \theta^i_k \right| = \beta_i \sum_{a_s \in A_s} \sum_{k=1}^{M} \left| \prod_{m=1}^{N} (u^m_{s,a^m_s} + \alpha^m_{s,a^m_s}) \omega^i_k - \prod_{m=1}^{N} u^m_{s,a^m_s} \theta^i_k \right|$$

$$= \beta_i \sum_{a_s \in A_s} \sum_{k=1}^{M} \left| \prod_{m=1}^{N} u^m_{s,a^m_s} (\omega^i_k - \theta^i_k) + \omega^i_k \sum_{\substack{I \subseteq 1,\dots,N \\ |I| \geq 1}} \prod_{m \in I} \alpha^m_{s,a^m_s} \prod_{m \in I^C} u^m_{s,a^m_s} \right|$$
$$\leq \beta_i \sum_{a_s \in A_s} \sum_{k=1}^{M} \left| \prod_{m=1}^{N} u^m_{s,a^m_s} \right| \left| (\omega^i_k - \theta^i_k) \right| + \beta_i W \sum_{a_s \in A_s} \sum_{k=1}^{M} \sum_{\substack{I \subseteq 1,\dots,N \\ |I| \geq 1}} \left| \prod_{m \in I} \alpha^m_{s,a^m_s} \right| \left| \prod_{m \in I^C} u^m_{s,a^m_s} \right|$$
$$\leq \beta_i \sum_{a_s \in A_s} \sum_{k=1}^{M} |\gamma^i_s| + \beta_i W \sum_{a_s \in A_s} \sum_{k=1}^{M} \sum_{\substack{I \subseteq 1,\dots,N \\ |I| \geq 1}} \left| \prod_{m \in I} \alpha^m_{s,a^m_s} \right|$$

$$< \beta_i \sum_{a_s \in A_s} \sum_{k=1}^{M} \delta_2(\epsilon) + \beta_i W \sum_{a_s \in A_s} \sum_{k=1}^{M} \sum_{\substack{I \subseteq 1,\ldots,N \\ |I| \geq 1}} \delta_3(\epsilon)$$
$$= \frac{\epsilon}{3} + \frac{\epsilon}{3} = 2\frac{\epsilon}{3}.$$

Thus,

$$\left| \psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s, \omega^i) - \psi_s^i(\tilde{C}_s, \tilde{P}_s; u_s, \theta^i) \right| \leq$$
$$K \sum_{a_s \in A_s} \left| \prod_{m=1}^{N} x_{s,a_s^m}^m - \prod_{m=1}^{N} u_{s,a_s^m}^m \right| + \beta_i \sum_{a_s \in A_s} \sum_{k=1}^{M} \left| \prod_{m=1}^{N} x_{s,a_s^m}^m \omega_k^i - \prod_{m=1}^{N} u_{s,a_s^m}^m \theta_k^i \right| < \frac{\epsilon}{3} + 2\frac{\epsilon}{3} = \epsilon.$$

□

The following two lemmas are direct consequences of Lemma 1 and the definition of $f_s^i(x_s^{-i}, u_s^i; \omega^i)$.

Lemma 2. *The function $f_s^i(x_s^{-i}, u_s^i; \omega^i)$ is continuous $\forall i \in I$, and $s \in S$.*

Lemma 3. *$f_s^i(x_s^{-i}, u_s^i; \omega^i)$ is convex in u_s^i for fixed x_s^{-i} and ω^i.*

We need the following transformation and definition in order to prove Lemma 5, which is required to show the upper semi-continuity result in the main existence theorem below (Theorem 4):

Let $\gamma_{s,x_s^{-i}}^i(\omega^i) = \alpha_{s,\omega^i}^i(x_s^{-i})$.

Define

$$\tau^i(x^{-i}) = \{\omega^i = (\omega_1^i, \ldots, \omega_M^i) : \omega_s^i = \min_{u_s^i \in S_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i), s \in S\},$$

and denote the s^{th} element of $\tau^i(x^{-i})$ by $\tau_s^i(x^{-i})$.

Proof of Lemma 4 follows directly from Fink (1964) and Lemma 1 above. Proof of Lemma 5 follows directly from Lemma 4 as shown in Fink (1964). These proofs are presented using our notation in the Appendix. Lemma 4 is used to prove Lemma 5, and Lemma 5 is used to show the upper semi-continuity result required by Kakutani's fixed point theorem.

Lemma 4. *$\alpha_{s,\omega^i}^i(x_s^{-i})$ is continuous on X_s^{-i}.*
Furthermore, the set $\{\alpha_{s,\omega^i}^i | \omega^i$ is bounded $\}$ is equicontinuous.

Lemma 5. *If $x^{-i,n} \to x^{-i}$ and $\tau_s^i(x^{-i,n}) \to \omega_s^i$ as $n \to \infty$, then $\tau_s^i(x^{-i}) = \omega_s^i$.*

Theorem 4. (Existence of Equilibrium in Robust Stochastic Games)

Suppose that uncertain transition probabilities and payoffs in a discounted robust stochastic game belong to compact sets and that the set of actions and players, who use stationary strategies, are finite. Then, an equilibrium point of this game exists.

Proof.

We show the existence of an equilibrium point that satisfies conditions (5) and (6) by using Kakutani's fixed point theorem. We will show that the fixed point of a suitably constructed correspondence is this equilibrium point. To this end, let

$$\phi(x) = \{(y) \in X \quad | \quad y_s^i \in \mathrm{argmin}_{u_s^i \in X_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_{sa_s}, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i),$$
$$\text{and } \omega_s^i = \min_{u_s^i \in X_s^i} \max_{\substack{\tilde{C}_s \in Cs \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i), \forall s \in S, i \in I\}.$$

Note that by definition, $\phi(x) \subseteq X$, $\forall x \in X$.

Next, we show that $\phi(x)$ is a convex set. Suppose that $(z^1,...,z^N),(v^1,...,v^N) \in \phi(x^1,...,x^N)$. Then, $\forall u_s^i$, and $s \in S, i \in I$, $w_s^i = f_s^i(x_s^{-i}, z_s^i; \omega^i) = f_s^i(x_s^{-i}, v_s^i; \omega^i) \leq f_s^i(x_s^{-i}, u_s^i; \omega^i)$. Hence, for any $\lambda \in [0,1]$ and $\forall i \in I, s \in S$,

$$w_s^i = \lambda f_s^i(x_s^{-i}, z_s^i; \omega^i) + (1-\lambda) f_s^i(x_s^{-i}, v_s^i; \omega^i) \leq f(x_s^{-i}, u_s^i; \omega^i)$$

By the convexity of $f_s^i(x_s^{-i}, u_s^i; \omega^i)$, we obtain

$$w_s^i = f_s^i(x_s^{-i}, ((\lambda) z_s^i + (1-\lambda) v_s^i); \omega^i) \leq \lambda f_s^i(x_s^{-i}, z_s^i; \omega^i) + (1-\lambda) f_s^i(x_s^{-i}, v_s^i; \omega^i) \leq f_s^i(x_s^{-i}, u_s^i; \omega^i),$$

and hence, $(\lambda)(z^1,...,z^N) + (1-\lambda)(v^1,...,v^N) \in \phi(x^1,...,x^N)$.

Finally, we must show that $\phi(x)$ is an upper semi-continuous correspondence. Suppose $x_n \to x$, $y_n \to y$, and $y_n \in \phi(x_n)$. Taking a subsequence, we have $\tau_s^i(x^{-i,n}) \to \omega_s^i$ and by Lemma 5, $\tau_s^i(x^{-i}) = w_s^i$. Using the triangle inequality, we have $\forall i \in I, s \in S$ that

$$\begin{aligned} &|f_s^i(x_s^{-i}, y_s^i; \omega^i) \; - \; w_s^i| \\ &\quad \leq |f_s^i(x_s^{-i}, y_s^i; \omega^i) - f_s^i(x_s^{-i,n}, y_s^{i,n}; \tau^i(x^{-i,n}))| + |f_s^i(x_s^{-i,n}, y_s^{i,n}; \tau^i(x^{-i,n})) - w_s^i| \\ &\quad = |f_s^i(x_s^{-i}, y_s^i; \omega^i) - f_s^i(x_s^{-i,n}, y_s^{i,n}; \tau^i(x^{-i,n}))| + |\tau^i(x^{-i,n}) - w_s^i| \to 0 \text{ as } n \to \infty. \end{aligned}$$

Therefore, $w_s^i = f_s^i(x_s^{-i}, y_s^i; \omega^i)$, and since $\tau_s^i(x^{-i}) = w_s^i$, we obtain that

$$\omega_s^i = \min_{u_s^i \in X_s^i} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; x_s^{-i}, u_s^i; \omega^i).$$

Therefore, $y \in \phi(x)$, completing the proof that ϕ is an upper semi-continuous correspondence. The fact that $\phi(x)$ is a closed set for any x follows from the fact that it is an upper-semi continuous correspondence. Therefore, ϕ satisfies the assumptions of Kakutani's fixed point theorem. $\square$

4. Calculation of an Equilibrium Point

Now that we have proved the existence of an equilibrium point in a discounted robust stochastic game, our next step is to calculate such a point. We will show that when the ambiguity in the probability transition data of the game belongs to a polytope intersected with the probability simplex, the problem of finding an equilibrium point could be cast as a feasibility problem that has multi-linear constraints. For simplicity, we only consider ambiguity in the transition data in constructing a feasibility problem that characterizes an equilibrium point. An analogous approach can be used to consider ambiguity both in payoffs and the transition data of a game.

Recall the definition of a robust equilibrium given in Conditions (5) and (6). These conditions are equivalent to the requirement that $\forall i \in I, s \in S, \exists q_s^i \in \Re$ such that $(\mathbf{x}_s^i, q_s^i)$ is an optimizer of the following robust mathematical program P_R with the objective value at optimality being equal to w_s^i:

$$P_R := \{w_s^i = \min_{u_s^i, q_s^i} q_s^i \quad : \quad q_s^i \geq \max_{\tilde{P}_s \in P_s} \psi_s^i(C_s, \tilde{P}_s; \mathbf{x}_s^{-i}, \mathbf{u}_s^i; \mathbf{w}^i), \mathbf{1}\mathbf{u}_s^i = 1, \mathbf{u}_s^i \geq \mathbf{0}\}.$$

Here, $(\mathbf{x}^{-i}, \mathbf{w}^i)$ is treated as data. Define the uncertain probability transition matrix induced by a strategy $(\mathbf{x}^{-i}, \mathbf{u}^i)$:

$$\tilde{\mathbf{P}}(\mathbf{x}^{-i}, \mathbf{u}^i) = \left[\sum_{a_s \in A_s} \prod_{\substack{m=1 \\ m \neq i}}^{N} x_{s,a_s^m}^m u_{s,a_s^i}^i \tilde{P}_{s a_s k} \right]_{s=1,k=1}^{M, \quad M}.$$

Denote the s^{th} row of $\tilde{\mathbf{P}}(\mathbf{x}^{-i}, \mathbf{u}^i)$ by $\tilde{\mathbf{p}}_s(\mathbf{x}^{-i}, \mathbf{u}^i)$. Let $\tilde{\mathbf{p}}_s$ denote the uncertain transition probability vector associated with the starting state s, that is, $\tilde{\mathbf{p}}_s = [\tilde{P}_{s a_s k}]_{a_s \in A_s; k \in S}$. Let $\mathbf{1}$ be a vector of ones of appropriate dimension. Let $\mathbf{E}_s^i(\mathbf{x}_s^{-i}, C^i) \in \Re^{\prod_{n=1, n\neq i}^{N} m_s^n m_s^i}$ denote the matrix a row of which is given by the vector

$$\left[\prod_{\substack{m=1\\m\neq i}}^{N} x^m_{s,a_s^m} C^i_{s,a_s^{-i},a_s^i}\right]_{a_s^i \in 1,\dots,m_s^i}.$$

Note that we have the following requirement in P_R:

$$q_s^i \geq \max_{\tilde{P}_s \in P_s} \psi_s^i(C_s, \tilde{P}_s; \mathbf{x}_s^{-i}, \mathbf{u}_s^i; \mathbf{w}^i) = \max_{\tilde{\mathbf{p}}_s} \beta \tilde{\mathbf{p}}_s(\mathbf{x}^{-i}, \mathbf{u}^i)\mathbf{w}^i + \mathbf{1}^T \mathbf{E}_s^i(\mathbf{x}_s^{-i}, C^i)\mathbf{u}_s^i \tag{8}$$

We assume that for any alternative combination of the players, the uncertain transition probabilitites belong to a polytope intersected with the probability simplex.

Let $\mathbf{Q}_s \in \Re^{(\prod_{i=1}^N m_s^i)(M \prod_{i=1}^N m_s^i)}$ be a matrix of 0s and 1s (such that each of its rows that cooresponds to a pure strategy combination $\mathbf{a}_s \in A_s$ satisfies $\sum_{k\in S} \tilde{P}_{s,a_s,k} = 1$). In other words, we assume that the transition probabilities belong to the following uncertainty set: $P = \{\tilde{\mathbf{p}}_s, s \in S \ : \mathbf{A}_s \tilde{\mathbf{p}}_s \geq \mathbf{b}_s, \mathbf{Q}_s \tilde{\mathbf{p}}_s = \mathbf{1}, \tilde{\mathbf{p}}_s \geq 0\}$, where $\mathbf{A}_s \in \Re^{l_s M \prod_{i=1}^N m_s^i}$.

Consider the maximization problem in P_R, where $(\mathbf{x}_s^{-i}, \mathbf{u}_s^i, \mathbf{w}^i)$ is regarded as data. Given that the uncertainty set is as stated, for fixed $(\mathbf{x}_s^{-i}, \mathbf{u}_s^i, \mathbf{w}^i)$, this maximization problem is equivalent to the following LP:

$$\{\max_{\tilde{\mathbf{p}}_s} \beta \tilde{\mathbf{p}}_s(\mathbf{x}^{-i}, \mathbf{u}^i)\mathbf{w}^i \ : \mathbf{A}_s \tilde{\mathbf{p}}_s \geq \mathbf{b}_s, \mathbf{Q}_s \tilde{\mathbf{p}}_s = \mathbf{1}, \tilde{\mathbf{p}}_s \geq 0\}. \tag{9}$$

Define the column vector $\mathbf{z}_s^i = [\prod_{\substack{m=1\\m\neq i}}^N x^m_{s,a_s^m} u^i_{s,a_s^i} w_k^i]_{a_s^m \in A_s^m; a_s^i \in A_s^i; k \in S}$ such that the indices of $\mathbf{z}_s^i$ match the ones of $\tilde{\mathbf{p}}_s$. Let $\mathbf{Y}_s^i(\mathbf{x}_s^{-i}, \mathbf{w}^i) \in \Re^{\prod_{n=1}^N m_s^n M m_s^i}$ be the matrix such that $\mathbf{Y}_s^i(\mathbf{x}_s^{-i}, \mathbf{w}^i)\mathbf{u}_s^i = \mathbf{z}_s^i$. Let $\mathbf{m}_s^i$ and $\mathbf{n}_s^i$ be the dual variable vectors of problem (9). The dual of problem (9) is

$$\{\min_{\mathbf{m}_s, \mathbf{n}_s} \ \left[[\mathbf{b}_s]'; [\mathbf{1}]'\right] \begin{bmatrix} \mathbf{m}_s^i \\ \mathbf{n}_s^i \end{bmatrix} \ : \ \begin{bmatrix} \mathbf{A}_s \\ \mathbf{Q}_s \end{bmatrix}' \begin{bmatrix} \mathbf{m}_s^i \\ \mathbf{n}_s^i \end{bmatrix} \geq \beta \mathbf{Y}_s^i(\mathbf{x}_s^{-i}, \mathbf{w}^i)\mathbf{u}_s^i, \ \ \mathbf{m}_s^i \leq 0\}. \tag{10}$$

By the definition of our uncertainty set, problem (9) is feasible and it is clear that it is bounded. By strong duality, problem (10) is bounded and feasible and its optimal objective value is equal to that of problem (9). Therefore, if $(\mathbf{x}^{-i}, \mathbf{u}^i, \mathbf{w}^i)$ satisfies condition (8), then (8) is equivalent to the condition that $\exists \mathbf{m}_s^i \in \Re^{l_s}$ and $\exists \mathbf{n}_s^i \in \Re^{\prod_{i=1}^N m_s^i}$ such that:

$$\begin{aligned} q_s^i - \mathbf{1}^T \mathbf{E}_s^i(\mathbf{x}_s^{-i}, C^i)\mathbf{u}_s^i &\geq \left[[\mathbf{b}_s]'; [\mathbf{1}]'\right] \begin{bmatrix} \mathbf{m}_s^i \\ \mathbf{n}_s^i \end{bmatrix} \\ \begin{bmatrix} \mathbf{A}_s \\ \mathbf{1}_s \end{bmatrix}' \begin{bmatrix} \mathbf{m}_s^i \\ \mathbf{n}_s^i \end{bmatrix} &\geq \beta \mathbf{Y}_s^i(x_s^{-i}, w^i)\mathbf{u}_s^i \\ \mathbf{m}_s^i &\leq 0 \end{aligned} \tag{11}$$

Conversely, if condition (11) is satisfied, then problem (10) is feasible. Then by weak duality, any feasible solution $\left[[\mathbf{b}_s]'; [\mathbf{1}]'\right] \begin{bmatrix} \mathbf{m}_s^i \\ \mathbf{n}_s^i \end{bmatrix}$ of problem (10) is greater than or equal to any solution $\beta \tilde{\mathbf{p}}_s(\mathbf{x}^{-i}, \mathbf{u}^i)\mathbf{w}^i$ of problem (9), so,

$$q_s^i - \mathbf{1}^T \mathbf{E}_s^i(\mathbf{x}_s^{-i}, C^i)\mathbf{u}_s^i \geq \left[[\mathbf{b}_s]'; [\mathbf{1}]'\right] \begin{bmatrix} \mathbf{m}_s^i \\ \mathbf{n}_s^i \end{bmatrix} \geq \max_{\tilde{\mathbf{p}}_s} \beta \tilde{\mathbf{p}}_s(\mathbf{x}^{-i}, \mathbf{u}^i)\mathbf{w}^i.$$

Therefore conditions (8) and (11) are equivalent. This proves:

LEMMA 6. *Condition (8) is equivalent to condition (11).*

Let

$$\mathbf{T}^i(\mathbf{x}) = \left[\sum_{a_s \in A_s} \prod_{m=1}^{N} x^m_{s,a^m_s} t^i_{sa_sk} \right]^{M,\ \ M}_{s=1,k=1},$$

and denote the s^{th} row of $\mathbf{T}^i(\mathbf{x})$ by $\mathbf{t}^i_s(\mathbf{x})$. Let $\mathbf{t}^i_s$ denote the variables representing the transition probabilities adopted by player i according to i's worst-case perspective, associated with the starting state s. That is, $\mathbf{t}^i_s = [t^i_{sa_sk}]_{a_s \in A_s; k \in S}$.

THEOREM 5. *A stationary strategy $\boldsymbol{x}$ is a robust equilibrium point with the robust value vector $\boldsymbol{w}^i$, iff $\forall i \in I, s \in S, \exists \boldsymbol{m}^i_s \in \Re^{l_s}, \boldsymbol{n}^i_s \in \Re^{\prod_{i=1}^N m^i_s}, \boldsymbol{t}^i_s \in \Re^{M \prod_{i=1}^N m^i_s}$ such that for $j = 1, ..., m^i_s$, $(\boldsymbol{w}^i, \boldsymbol{x}_s, \boldsymbol{m}^i_s, \boldsymbol{n}^i_s, \boldsymbol{t}^i_s)$ satisfies*

$$w^i_s = \boldsymbol{1}^T \boldsymbol{E}^i_s(\boldsymbol{x}^{-i}_s, C^i)\boldsymbol{x}^i_s + \beta \boldsymbol{t}^i_s(\boldsymbol{x})\boldsymbol{w}^i$$

$$[\boldsymbol{e}^i_{s,j}]' \boldsymbol{E}'^i_s(\boldsymbol{x}^{-i}_s, C^i)\boldsymbol{1} + \beta[\boldsymbol{e}^i_{s,j}]' \boldsymbol{Y}'^i_s(\boldsymbol{x}^{-i}_s, \boldsymbol{w}^i)\boldsymbol{t}^i_s \geq \boldsymbol{w}^i_s \geq \left[[\boldsymbol{b}_s]'; [\boldsymbol{1}]'\right] \begin{bmatrix} \boldsymbol{m}^i_s \\ \boldsymbol{n}^i_s \end{bmatrix} - \boldsymbol{1}^T \boldsymbol{E}^i_s(\boldsymbol{x}^{-i}_s, C^i)\boldsymbol{x}^i_s$$

$$\begin{bmatrix} \boldsymbol{A}_s \\ \boldsymbol{Q}_s \end{bmatrix}' \begin{bmatrix} \boldsymbol{m}^i_s \\ \boldsymbol{n}^i_s \end{bmatrix} - \beta \boldsymbol{Y}^i_s(\boldsymbol{x}^{-i}_s, \boldsymbol{w}^i)\boldsymbol{x}^i_s \geq 0$$

$$\boldsymbol{1}\boldsymbol{x}^i_s = 1, \quad \boldsymbol{m}^i_s \leq 0, \quad \boldsymbol{x}^i_s \geq 0, \quad \boldsymbol{A}_s \boldsymbol{t}^i_s \geq \boldsymbol{b}_s, \quad \boldsymbol{Q}_s \boldsymbol{t}^i_s = \boldsymbol{1}.$$

Proof.

Recall problem P_R. By lemma 6, if $\mathbf{x}$ is a robust equilibrium point, $\forall i \in I, s \in S, \exists q^i_s \in \Re, \mathbf{m}^i_s \in \Re^{l_s}, \mathbf{n}^i_s \in \Re^{\prod_{i=1}^N m^i_s}$ such that $(\mathbf{x}^i_s, q^i_s, \mathbf{m}^i_s, \mathbf{n}^i_s)$ is an optimizer of

$$\begin{aligned} w^i_s = &\min_{\mathbf{u}^i_s, q^i_s, \mathbf{m}^i_s, \mathbf{n}^i_s} \quad q^i_s \\ &q^i_s - \mathbf{1}^T \mathbf{E}^i_s(\mathbf{x}^{-i}_s, C^i)\mathbf{u}^i_s \geq \left[[\mathbf{b}_s]'; [\mathbf{1}]'\right] \begin{bmatrix} \mathbf{m}^i_s \\ \mathbf{n}^i_s \end{bmatrix} \\ &\begin{bmatrix} \mathbf{A}_s \\ \mathbf{Q}_s \end{bmatrix}' \begin{bmatrix} \mathbf{m}^i_s \\ \mathbf{n}^i_s \end{bmatrix} \geq \beta \mathbf{Y}^i_s(\mathbf{x}^{-i}_s, \mathbf{w}^i)\mathbf{u}^i_s, \quad \mathbf{m}^i_s \leq 0, \quad \mathbf{1}\mathbf{u}^i_s = 1, \quad \mathbf{u}^i_s \geq 0 \end{aligned} \tag{12}$$

Let $\mathbf{e}^i_{s,j}$ be the j^{th} unit vector. Dual of the above is:

$$\max_{\nu^i_s, \mathbf{t}^i_s} \ \nu^i_s : \mathbf{A}_s \mathbf{t}_s \geq \mathbf{b}_s, \ \ \mathbf{Q}_s \mathbf{t}^i_s = \mathbf{1}, \ \ \nu^i_s \leq [\mathbf{e}^i_{s,j}]' \mathbf{E}'^i_s(\mathbf{x}^{-i}_s, C^i)\mathbf{1} + \beta[\mathbf{e}^i_{s,j}]' \mathbf{Y}'^i_s(\mathbf{x}^{-i}_s, \mathbf{w}^i)\mathbf{t}^i_s, j = 1, ..., m^i_s. \tag{13}$$

The statement in the theorem follows from strong duality and Theorem 2. For the other direction, now suppose a given $\mathbf{x}_s, \mathbf{w}^i$, and that $\forall i \in I, s \in S, \exists \mathbf{m}^i_s \in \Re^{l_s}, \mathbf{n}^i_s \in \Re^{\prod_{i=1}^N m^i_s}, \mathbf{t}^i_s \in \Re^{M \prod_{i=1}^N m^i_s}$, such that $(\mathbf{w}^i, \mathbf{x}_s, \mathbf{m}^i_s, \mathbf{n}^i_s, \mathbf{t}^i_s)$ satisfies the above system. Let

$$\nu^i_s = \min_{j \in 1, \ldots, m^i_s} [\mathbf{e}^i_{s,j}]' \mathbf{E}'^i_s(\mathbf{x}^{-i}_s, C^i)\mathbf{1} + \beta[\mathbf{e}^i_{s,j}]' \mathbf{Y}'^i_s(\mathbf{x}^{-i}_s, \mathbf{w}^i)\mathbf{t}^i_s, \qquad j = 1, ..., m^i_s,$$

$$q^i_s = \left[[\mathbf{b}_s]'; [\mathbf{1}]'\right] \begin{bmatrix} \mathbf{m}^i_s \\ \mathbf{n}^i_s \end{bmatrix} + \mathbf{1}^T \mathbf{E}^i_s(\mathbf{x}^{-i}_s, C^i)\mathbf{u}^i_s$$

Then, for $(\mathbf{x}^{-i}_s, \mathbf{w}^i)$, $(\mathbf{x}^i_s, q^i_s, \mathbf{m}^i_s, \mathbf{n}^i_s)$ is feasible for problem 12, and $(\nu^i_s, \mathbf{t}^i_s)$ is feasible for problem (8) with $\nu^i_s \geq q^i_s$. By weak duality $\nu^i_s \leq q^i_s$, so $\nu^i_s = q^i_s$. Hence, $(\mathbf{x}^i_s, q^i_s, \mathbf{m}^i_s, \mathbf{n}^i_s)$ is optimal for problem 12. Hence, $(\mathbf{x}^i_s, q^i_s)$ is optimal in P_R. Therefore, $\mathbf{x}$ is an equilibrium point of the discounted robust stochastic game. □

5. A Queueing Control Application

Game theoretical analysis has been widely applied to queueing control problems, e.g., Altman and Shimkin (1998), Heyman (1968), Sobel (1969), Stidham and Weber (1989), Yechiali (1971), and to flow control problems in particular, e.g., Altman (1994b), Altman (1994a), Altman and Hordijk (1995). In this section, we present an application of our robust model for incomplete information stochastic games to a single server queuing system. Such a queueing control problem arises in telecommunication systems (Altman (1994a)). In this problem, a controller, called player 2, dynamically controls the flow of arriving customers into a finite buffer. The service rate, controlled by the service controller (player 1), depends on the state of the system and varies in time, and is unknown to both of the controllers. The goal of the controllers' is then to design robust strategies that could guarantee the best performance under worst-case data conditions. We next provide the details for our model.

Let X_t represent the number of customers in the system at time t, $t=0,1,2,\ldots$. The state space is denoted by $\mathbf{X}=\{0,1,\ldots,L\}$, where $L<\infty$ is the buffer size. It is assumed that at most one arrival can occur at the beginning of a time slot. At the end of each time slot, if the state is x the service controller (player 1) chooses an interval from a finite set of possibly disjoint intervals of service rates. Hence, the exact service rate is unknown and belongs to an interval. Formally, if the state is x, service controller choses an interval $I^1_j \in A^1_x$, where A^1_x is a finite collection of possibly disjoint closed intervals. That is, $A^1_x=\{I^1_j\}_{j=1}^J$, $I^1_j=[\underline{\mu}_j,\overline{\mu}_j]$. At the beginning of a time slot, if there are x customers in the system, the flow controller chooses an arrival rate λ from a finite set A^2_x of arrival rates. We assume that no arrivals are allowed when the buffer is full. Let us denote the unknown service rate by $\tilde{\mu}$. We assume that the alternative sets of both players are the same for all states, that is, $A^1_x=A^1, A^2_x=A^2, \forall x\in X$. Therefore, we have the following transition rule that is unknown commonly to both players:

$$\tilde{p}(y|x,\tilde{\mu},\lambda)=\left\{\begin{array}{cc} \tilde{\mu}/(\lambda+\tilde{\mu}), & 1\le x\le L,\quad y=x-1\\ \lambda/(\lambda+\tilde{\mu}), & 0\le x<L-1,\quad y=x+1\\ 1-\lambda/(\lambda+\tilde{\mu}), & y=x=0\\ 1-\tilde{\mu}/(\lambda+\tilde{\mu}), & y=x=L \end{array}\right\}.$$

The immediate payoff function that is frequently used in the literature on flow control models is defined as follows (see Altman (1994b)):

$$C(x,\widehat{\mu},\lambda)=c(x)+\theta(\widehat{\mu})+\rho(\lambda),$$

where $\widehat{\mu}$ denotes the midpoint of a service rate interval. $C(x,\widehat{\mu},\lambda)$ represents the cost that the flow controller pays the service controller given the state of the system and the alternative pairs. Hence, in this two-person zero-sum model, player 2 is the minimizing, whereas player 1 is the maximizing player. Here, $c(x)$ is a nondecreasing real function in x, $\theta\ge 0$ is an increasing real function in $\widehat{\mu}$, and ρ is a nonpositive real function of λ. $c(x)$ can be interpreted as the holding cost per unit time, θ as the cost associated with the service rate, and ρ as the reward associated with an entering customer. Note that the exact immediate costs are known to both players.

We set up an instance of this problem with two alternatives for the server controller at each state. First alternative for player 1 is a service rate of one customer per 14 to 16 seconds, and the second is a rate of one customer per 19 to 21 seconds. Note that the mean for the first and the second alternatives are between 14 and 16 seconds, and between 19 and 21 seconds, respectively. We set up nine more instances of the same problem by enlarging the range of the means for both alternatives by 2 seconds for every new instance. Hence we obtained Table 1 for service rate intervals for each instance.

In each state, the flow controller's (player 2) first and second alternatives, $1/\lambda_1$ and $1/\lambda_2$, are to permit one customer into the system every 15 and every 20 seconds on average, respectively.

Table 1 Intervals of rates for different instances

instance number	1	2	3	4	5
μ_1	[1/16,1/14]	[1/17,1/13]	[1/18,1/12]	[1/19,1/11]	[1/20,1/10]
μ_2	[1/21,1/19]	[1/22,1/18]	[1/23,1/17]	[1/24,1/16]	[1/25,1/15]
instance number	6	7	8	9	10
μ_1	[1/21,1/9]	[1/22,1/8]	[1/23, 1/7]	[1/24,1/6]	[1/25,1/5]
μ_2	[1/26,1/14]	[1/27,1/13]	[1/28,1/12]	[1/29,1/11]	[1/30,1/10]

Note that since players may use mixed strategies, a convex combination of the two alternatives could be chosen in any state. Under these sets of alternatives, we set up the following scenario for payoff functions depicted in Figure 1. This scenario indicates that in the first few states, the service controller pays the flow controller so that he admits more customers into the system, which would be beneficial for the service controller. However, after the admission of approximately 2 customers, the flow controller pays some amount to the service controller so that the admitted customers get served. From this point up to approximately 20 customers in the system, the alternative pair of ($\widehat{\mu_1} = 1/15sec^{-1}$, $\lambda_2 = 1/20sec^{-1}$) incurs the highest amount paid by the flow controller to the service controller, because the cost to the flow controller is high for higher service rates, and the reward associated with an entering customer is low. On the other hand, the alternative pair ($\widehat{\mu_2} = 1/20sec^{-1}$, $\lambda_1 = 1/15sec^{-1}$) that indicates a relatively slower service rate with respect to the flow rate, is less beneficial from the service controller's perspective, since the flow controller pays the service controller less for lower service rates, and the service controller pays the flow controller more for higher admission rates. Up until to the state of 20 customers in the system, alternative pairs ($\widehat{\mu_1} = 1/15sec^{-1}$, $\lambda_1 = 1/15sec^{-1}$) and ($\widehat{\mu_2} = 1/20sec^{-1}$, $\lambda_2 = 1/20sec^{-1}$) incurs costs to the flow controller that lie between these two extreme cases described, with higher service rates incurring slightly more cost to the flow controller. This cost scenario also indicates that after approximately 20 customers in the system, the pair ($\widehat{\mu_1} = 1/15sec^{-1}$, $\lambda_2 = 1/20sec^{-1}$) becomes preferable from the flow controller's perspective. This is so, since choosing slower rates when the service controller serves at a higher rate is beneficial for the flow controller.

Based on the service rate intervals of this birth and death process, we first calculate the intervals of transition probabilities to obtain the ambiguous transition data of the game for each instance in Table 1. We solved each instance using Theorem 5 and by setting $\beta = 0.95$. From the equilibrium strategies, we calculated the expected service and arrival rates at equilibrium for each state from the two players' perspectives. From these rates, we calculated steady state probabilites, the average number of customers in the system (L), the average amount of time a customer spends in the system (W), and the average value of the game from the point of view of each player. The average value for a player is calculated by weighting the value to a player starting in a state by the respective steady state probabilities from that player's perspective, and taking the summation over the states. Furthermore, we solved the nominal zero-sum discounted stochastic game by ignoring the uncertainty in service rates and considering the midpoints of the service intervals. We also calculated each of the above quantities for this nominal stochastic game. Note that each instance has the same rate as its midpoint. The results are depicted in Figures 2, 3, and 4. Figure 2 indicates that as far as the service controller is concerned, the average number of customers in the system is less than that of the nominal solution when there is uncertainty in the system. From the flow controller's perspective, L increases as the intervals becomes larger. Note that these are pessimistic points of views for both players since an increase in the number of customers would be an advantage to the service controller, whereas it would be a disadvantage to the flow controller. Consequently, the gain that the service controller achieves decreases and the cost to the flow controller increases as the length of the service intervals increases. On the other hand, average waiting time for a customer decreases from the service controller's perspective, and increases from the flow controller's

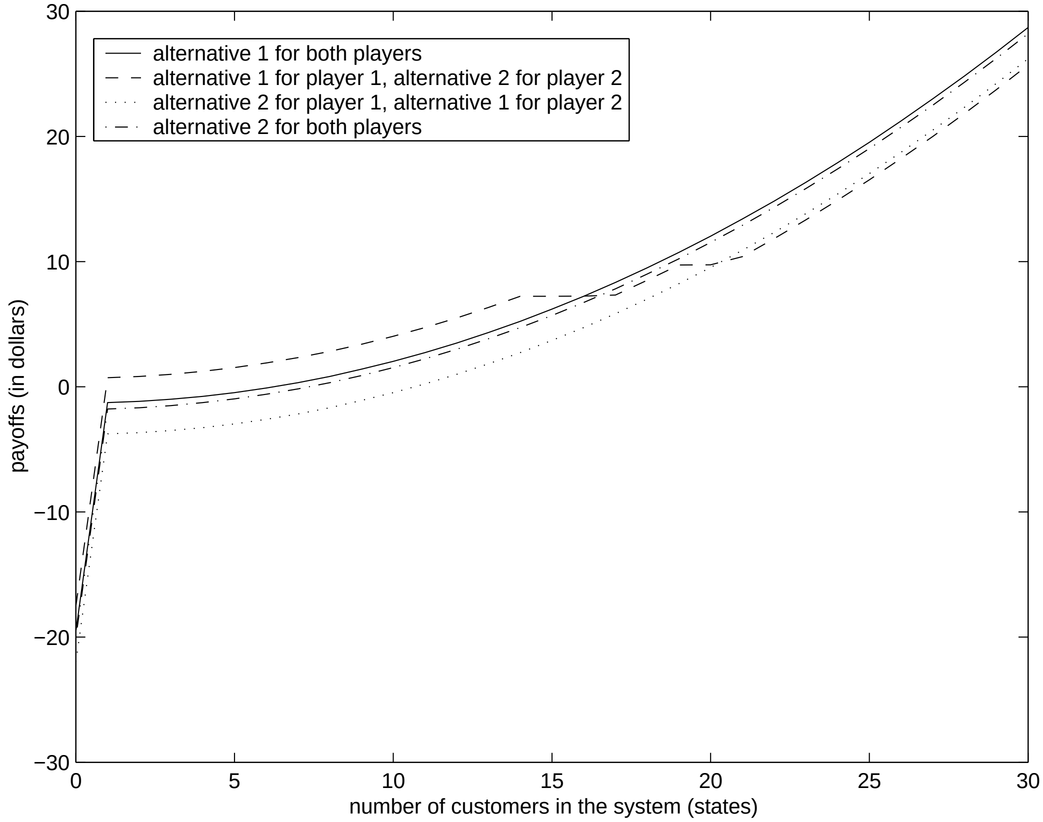

Figure 1 Payoff functions

perspective. The reason for this is that in the robust equilibrium, the service contoller assumes the pessimistic perspective of having less customers in the system, whereas the flow controller assumes the opposite. Note that, although this stochastic game is zero-sum, the resulting values differ for each player when they play robustly. Hence, we justify here that although our example is a zero-sum game, formulations for zero-sum games cannot be used to solve discounted robust stochastic games, despite the fact that one player pays the other player a fixed amount.

We next compare the results obtained for the average values, L, and W associated with the robust equilibrium and those with the nominal equilibrium under its respective worst-case probability transition data. This comparison allows us to contrast the robust equilibrium against the nominal solution under its worst possible data scenario, and to determine the uncertainty scenarios where the robust equilibrium performs better or worse, should the data attain worst-case values for the nominal equilibrium strategies. In order to obtain the latter for each instance, we first fixed the nominal strategies for each player and solved the resulting formulations in which transition data are treated as variables. The objectives in these formulations yield the maximum value (maximum cost) and the minimum value (minimum profit) for players 2 and 1, respectively. Note that results of the nominal solution under its worst possible data scenario do not pertain to an equilibrium. Figure 5 shows that for the fifth instance, the robust equilibrium yields less gain for the service controller than that of the nominal solution's under its worst-case data. This is possible in the context of stochastic games since, unlike the dotted line, the dashed line in Figure 5 pertains to a robust equilibrium point where no player is willing to deviate from their strategies. In robust equilibrium, players reach an equilibrium by not only considering their opponents' stratgeies but also the worst-case transition data with respect to the opponents' strategies. The dotted line is obtained by taking into account the nominal stochastic game's equilibrium strategies with respect

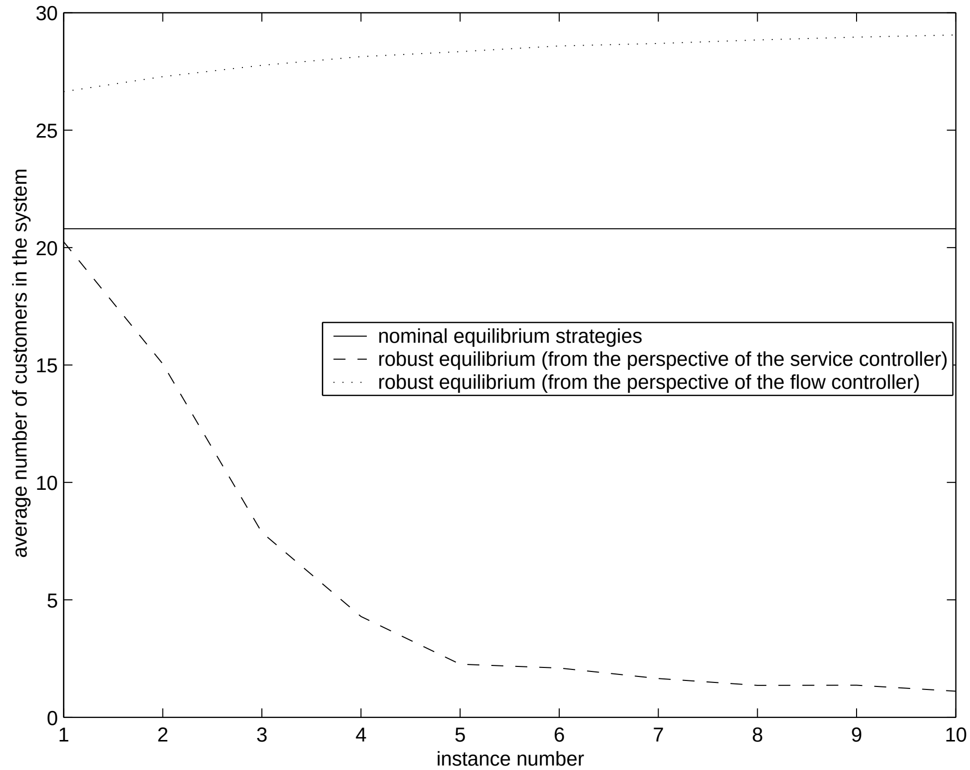

Figure 2 L, the average number of customers in the system

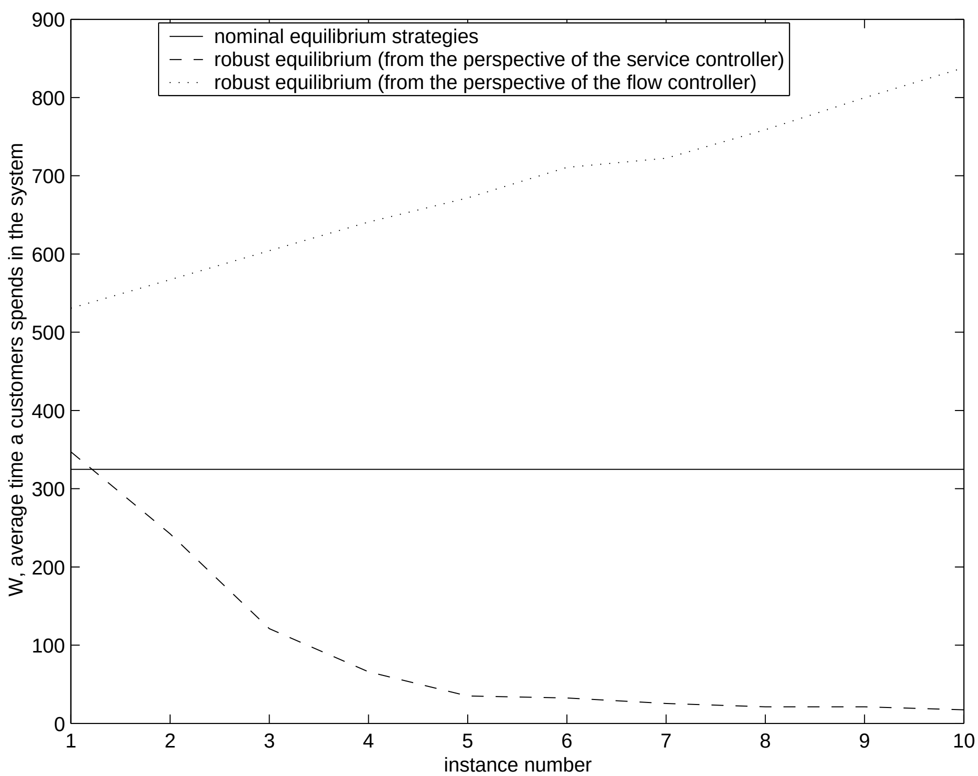

Figure 3 W, average time a customers spends in the system

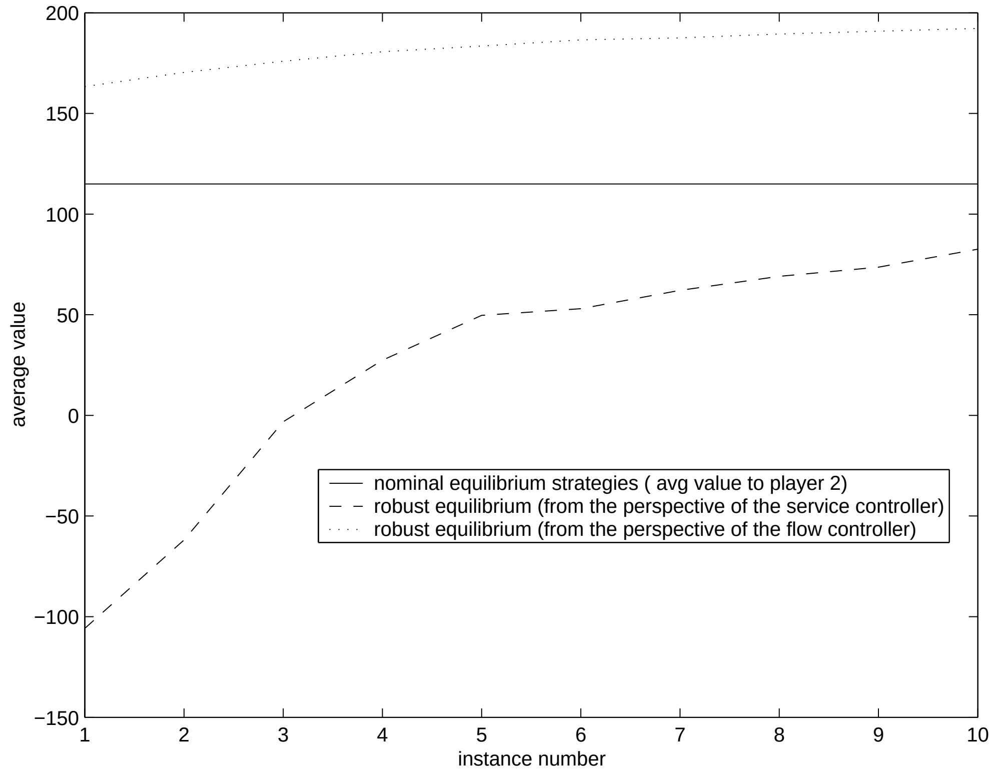

Figure 4 average value of the game

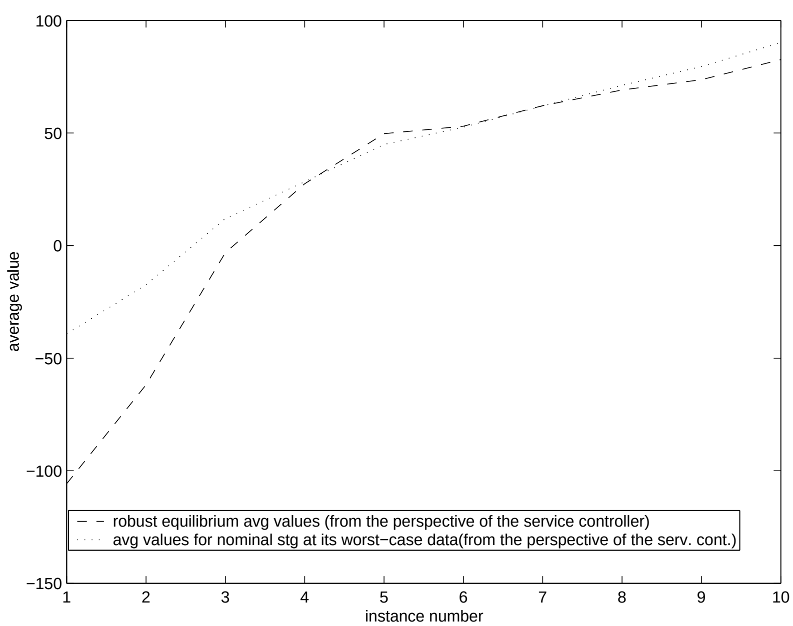

Figure 5 robust equilibrium average values vs average values for nominal strategies at their worst-case data

to their own worst-case transition data. Considering that the opponent is playing with nominal strategies and their respective worst-case data, a player could be willing to deviate from using nominal strategies unless they face an equilibrium. Here, the dotted line is not associated with an

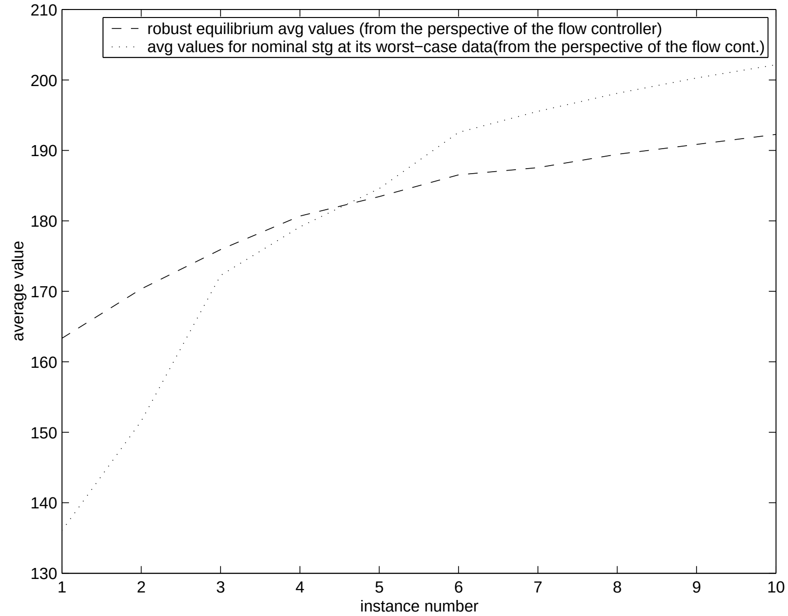

Figure 6 robust equilibrium average values vs average values for nominal strategies at their worst-case data

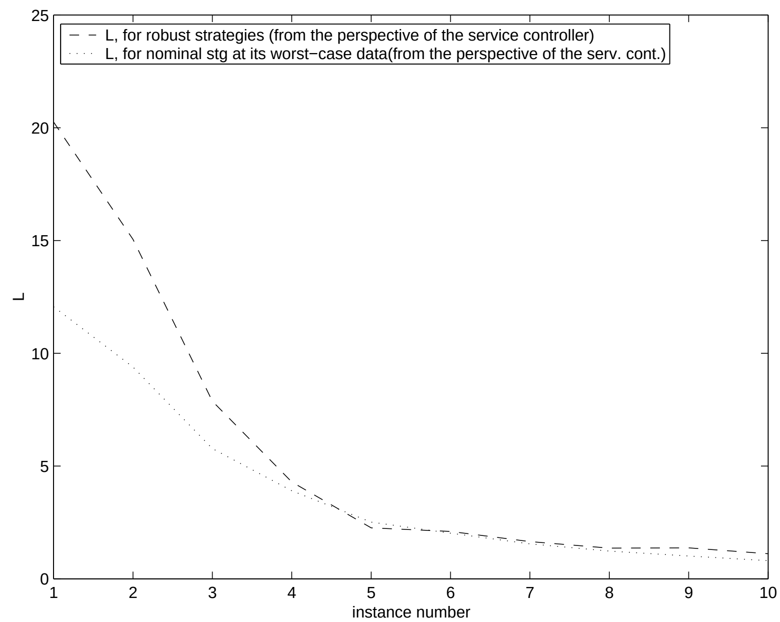

Figure 7 L values for robust equilibrium vs for nominal strategies at their worst-case data

equilibrium point, hence is not a stable point from which the players would be willing to deviate. This phenomenon becomes prevalent for the flow controller when the uncertainty in the service interval rates are relatively less (up to instance 5), as seen in Figure 6. This implies that for our example, as the uncertainty gets larger, using robust strategies is beneficial for the flow controller, as far as he/she is concerned with the average value of the game. Figures 7 and 8 indicate that, using robust strategies usually leads to a larger number of customers in the system. However, we just observed that robust strategies yield better (less) average values for the flow controller when the interval length for the service rates increases. Since having more customers in the system is

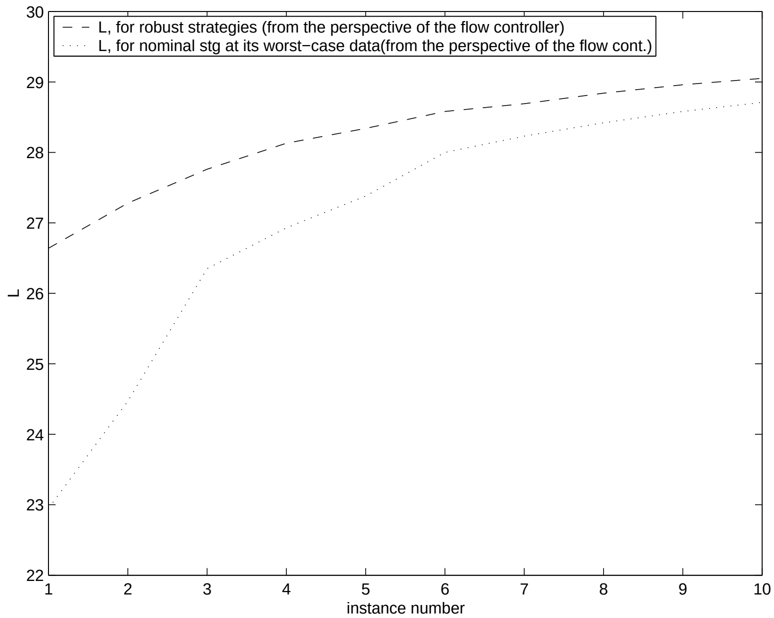

Figure 8 L values for robust equilibrium vs for nominal strategies at their worst-case data

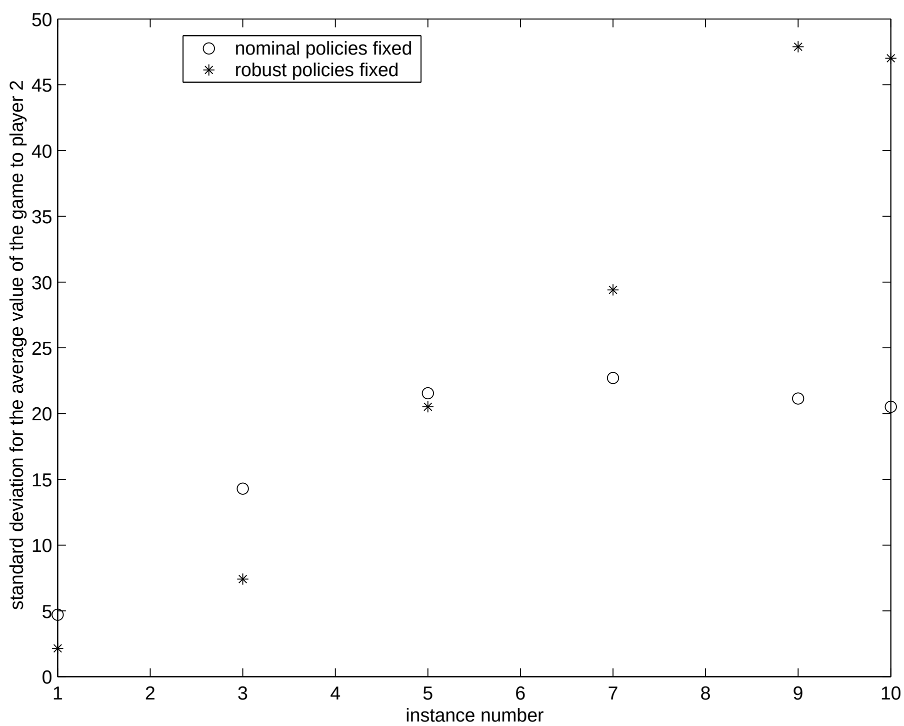

Figure 9 Simulation results for average values of the game to player 1

costly for the flow controller, this may seem unusual. However, it is not, and the explanation is as follows: Our computational results indicate that nominal startegies at their worst-case data scenario and robust strategies yield very similar steady state probabilities. However, the value (cost) of the game to the flow controller starting in any given state given by the former is always greater than the one given by the latter strategies. Therefore, although robust strategies yield slighltly higher number of customers in the system, they provide less average values.

Finally, we fixed the robust policies for both players and simulated the transition data for the instance numbers 1,3,5,7,9, and 10, and calculated the corresponding values by solving equations 14. Our sample size is 50. For each sample transition data, we also set the strategies for players to their nominal strategies and re-calculated the corresponding values by solving Equations 14 one more time.

$$\omega_s^i = \sum_{a_s \in A_s} \prod_{m=1}^{N} x^m_{s,a^m_s} \{C_{sa_s} + \beta \sum_{k=1}^{M} P_{sa_s k} \omega_k^i\}. \tag{14}$$

For each instance, we calculate the standard deviation of its sample, when robust and nominal strategies are fixed, and when transition data is simulated from a uniform distribution over the corresponding probability transition interval of the instance. Figure 9 depicts that for instances 1,3, and 5, the robust equilibrium strategies yield less standard deviation than that of the nominal equilibrium strategies. However, as the uncertainty increases the robust equilibrium strategies yield higher standard deviation values. The reason for this is that when the intervals of probability transition data enlarges, both players adopt overly conservative strategies, since they see the system in a very pessimistic way. For instance, when players use robust optimization, for the instances 7,9, and 10, the flow controller assumes that the service controller is going to serve at lower service rates, and the service controller assumes that the flow controller is going to admit less customers into the system. Hence, to protect themselves, players play with the overly conservative slow service and arrival rates. As a result, the service controller perceives that the number of customers in the system will be very low, and the flow controller perceives that it will be very high. When the transition data is simulated uniformly from a large interval, we observe substantial discrepancies between the realized (observed) average service and flow rates, and the assumed rates that the players intended to play with. Consequently, for the large uncertainty scenarios, we observe higher standard deviations when players play robustly.

6. Concluding Remarks and Future Research

In this paper, we consider n-player, non-zero sum discounted stochastic games in which none of the players knows the true transition probabilities and/or payoffs of the game and each player adopts a robust optimization approach to data ambiguity. We offer an alternative equilibrium concept for stochastic games with incomplete information. We propose a distribution-free model that relaxes the former approaches' assumptions on the player who has incomplete information and on whether the transitions are controlled by a single player. Our approach lends itself to computational results via a feasibility formulation for an equilibrium of a discounted robust stochastic game. We finally illustrate the use of discounted robust stochastic games in a queueing control context.

We determined several properties of discounted robust stochastic games in this research:

- An equilibrium exists even if there exists players who do not adopt a robust optimization approach. This stems from the fact that when there are no uncertainty sets for the data of a stochastic game, best response functions are already continuous, as shown in Fink (1964). Hence, we can construct a correspondence that satisfies Kakutani's theorem and that includes players who may play non-robustly.
- In our existence proof we have assumed that the players commonly know the uncertainty set of payoffs C_s at each state. The results of this existence theorem hold, even if each player has different sets for uncertainty.
- The zero-sum property is most likely to vanish for stochastic games in which the payoff uncertainty is a common set for all players, and in which there exists a player who plays robustly.

• If there is ambiguity in any data of the game, the players' approach to this ambiguity may differ.

• If the stochastic game is a two person zero-sum game but the transition data is ambigious, then the equilibrium values for the players do not negate each other. This implies that although such games are zero-sum, formulations for zero-sum stochastic games could not be used for analyses and properties that pertain to zero-sum games cannot be expected to hold in the presence of ambiguity.

• As dilemmas are common in game theory, it is possible to find dilemmas in discounted robust stochastic games. As in normal form games, discounted robust stochastic games could encompass different equilibrium points with different values and probability perspectives for players. In this sense, an equilibrium may not imply the best of what the players could do. For instance, the equilibrium values obtained by the minizing player in a two player discounted stochastic game could be less than that of its nominal counterpart. An equilibrium, in fact, means that given the probability perspectives adopted by the players, no player is unilaterally willing to deviate from his/her own strategy (Nash equilibrium). Consequently, based on the uncertainty sets used, it could be possible to have robust equilibrium points that does not have desirable properties. This issue is in parallel to discussions on the drawbacks of Nash equilibrium. Hence, the question of stability and perfection (see van Damme (1991)) also arises in the context of discounted robust stochastic games.

There are several new research directions that emerges as a result of the current research. An interesting one from a practical point of view is to explore equilibrium points in the flow control model of section 5, assuming that the flow controller does not use robust optimization. Another line of promising research would be to extend the ideas in this paper to limiting average stochastic games.

Appendix

THEOREM 6 **(Banach's Theorem and Proofs from Fink (1964))**. *Let (W,ρ) be a complete metric space and let $\gamma: W \to W$ be a contraction mapping. Then there exists a unique fixed point of the function γ.*

LEMMA 4 **(Fink (1964))**. *$\alpha^i_{s,\omega^i}(x_s^{-i})$ is continuous on X_s^{-i}. Furthermore, the set $\{\alpha^i_{s,\omega^i} | \omega^i \text{ is bounded}\}$ is equicontinuous.*

Proof.
Let

$$\begin{aligned}\alpha^i_{s,\omega^i}(x_s^{-i}) &= \psi^i_s(C^i_s(x_s^{-i},u_s^{*i}),P^i_s(x_s^{-i},u_s^{*i});x_s^{-i},u_s^{*i};\omega^i),\\ \alpha^i_{s,\omega^i}(y_s^{-i}) &= \psi^i_s(C^i_s(y_s^{-i},z_s^{*i}),P^i_s(y_s^{-i},z_s^{*i});y_s^{-i},z_s^{*i};\omega^i).\end{aligned}$$

Furthermore,

$$\begin{aligned}\alpha^i_{s,\omega^i}(y_s^{-i}) - \alpha^i_{s,\omega^i}(x_s^{-i}) &\le \psi^i_s(C^i_s(y_s^{-i},u_s^{*i}),P^i_s(y_s^{-i},u_s^{*i});y_s^{-i},u_s^{*i};\omega^i)\\ &\quad - \psi^i_s(C^i_s(x_s^{-i},u_s^{*i}),P^i_s(x_s^{-i},u_s^{*i});x_s^{-i},u_s^{*i};\omega^i),\\ \alpha^i_{s,\omega^i}(x_s^{-i}) - \alpha^i_{s,\omega^i}(y_s^{-i}) &\le \psi^i_s(C^i_s(x_s^{-i},z_s^{*i}),P^i_s(x_s^{-i},z_s^{*i});x_s^{-i},z_s^{*i};\omega^i)\\ &\quad - \psi^i_s(C^i_s(y_s^{-i},z_s^{*i}),P^i_s(y_s^{-i},z_s^{*i});y_s^{-i},z_s^{*i};\omega^i).\end{aligned}$$

If ω^i is restrained to be in a bounded region, then the right hand sides can be made uniformly small because of the uniform continuity of ψ^i_s on compact sets. □

LEMMA 5 **(Fink(1964))**. *If $x^{-i,n} \to x^{-i}$ and $\tau^i_s(x^{-i,n}) \to \omega^i_s$ as $n \to \infty$, then $\tau^i_s(x^{-i}) = \omega^i_s$.*

Proof.
We have

$$|\omega_s^i - \gamma^i_{s,x_s^{-i}}(\omega^i)| \leq |\omega_s^i - \tau_s^i(x^{-i,n})| + |\tau_s^i(x^{-i,n}) - \gamma^i_{s,x_s^{-i}}(\tau^i(x^{-i,n}))| + |\gamma^i_{s,x_s^{-i}}(\tau^i(x^{-i,n})) - \gamma^i_{s,x_s^{-i}}(\omega^i)|$$

Now, by assumption, as $n \to \infty$ $|\omega_s^i - \tau_s^i(x^{-i,n})| \to 0$ and $|\gamma^i_{s,x_s^{-i}}(\tau^i(x^{-i,n})) - \gamma^i_{s,x_s^{-i}}(\omega^i)| \to 0$. Note that $|\tau_s^i(x^{-i,n}) - \gamma^i_{s,x_s^{-i}}(\tau^i(x^{-i,n}))| = |\alpha^i_{s,\tau^i(x^{-i,n})}(x_s^{-i,n}) - \alpha^i_{s,\tau^i(x^{-i,n})}(x_s^{-i})| \to 0$ as $n \to \infty$ by Lemma 3 in Fink (1964). Hence, $|\omega_s^i - \gamma^i_{s,x_s^{-i}}(\omega_s^i)| \to 0$ as $n \to \infty$. □

Acknowledgments
I would like to acknowledge the financial support provided by the Center for Risk and Economic Analysis of Terrorism Events (CREATE) based in the Epstein Department of Industrial and Systems Engineering at the University of Southern California. I also would like to acknowledge and thank Michelle Aghassi for the insights she provided on the existence proof for robust one-shot n-player normal form games.

References

Aghassi, M., D. Bertsimas. 2006. Robust game theory. *Mathematical Programming (Series B)* **107** 231–273.

Altman, E. 1994a. Flow control using the theory of zero-sum markov games. *IEEE Trans. Automatic Control* **39**(4) 814–818.

Altman, E. 1994b. Monotonicity of optimal policies in a zero sum game: A flow control model. *Advances in Dynamic Games and Applications*, vol. 1. Birkhauser.

Altman, E., A. Hordijk. 1995. Zero-sum markov games and worst-case optimal control of queueing systems. *QUESTA* **21** 415–447.

Altman, E., N. Shimkin. 1998. Individually optimal dynamic routing in a processor sharing system: Stochastic game analysis. *Operations Research* **46**(6) 776–784.

Aumann, R. J., M. B. Maschler. 1968. Repeated games of incomplete information: The zero-sum extensive case. Tech. rep., U.S. Arms Control and Disarmament Agency. ST-143, Washington D.C., Chapter III, 37116.

Bagnell, J., A. Ng, J. Schneider. 2001. Solving uncertain markov decision problems. Tech. rep., Robotics Institute, Carnegie Mellon University. Tecnical Report CMU-RI-TR-01-25, Pittsburgh, PA.

Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Mathematics of Operations Research* **23**(4).

Bertsimas, D., M. Sim. 2004. The price of robustness. *Operations Research* **52** 35–53.

Filar, J., K. Vrieze. 1997. *Competitive Markov Decision Processes*. Springer-Verlag.

Fink, A. M. 1964. Equilibrium in a stochastic n-person game. *Journal of Science in Hiroshima University, Series A-I* **28** 89–93.

Gilboa, I., D. Schmeidler. 1989. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics* **18** 141–153.

Givan, R., S. Leach, T. Dean. 1997. Bounded parameter markov decision processes. *Fourth European Conf. Planning*. 234–246.

Harsanyi, J. C. 1967,1968. Games with incomplete information played by bayesian players, i-iii. *Management Science* **14**.

Heyman, D. P. 1968. Optimal operating policies for m/g/1 queueing systems. *Operations Research* **16** 362–282.

Iyengar, G. 2005. Robust dynamic programming. *Mathematics of Operations Research* **30**(2) 1–21.

Kakutani, S. 1941. A generalization of brouwer's fixed point theorem. *Duke Mathematical Journal* **8** 457–459.

Lo, K. C. 1996. Equilibrium in beliefs under uncertainty. *Journal of Economic Theory* **71** 443–484.

Marinacci, M. 2000. Ambiguous games. *Games and Economic Behavior* **31** 191–219.

Nilim, A., L. El Ghaoui. 2005. Robust control of markov decision processes with uncertain transition matrices. *Operations Research* **53**(5) 780–798.

Rosenberg, D., E. Solan, N. Vieille. 2004. Stochastic games with a single controller and incomplete information. *SIAM Journal of Control and Optimization* **43**(1).

Satia, J. K., R. L. Lave. 1973. Markov decision processes with uncertain transition probabilities. *Operations Research* **21**(3) 728–740.

Shapiro, A., A. J. Kleywegt. 2002. Minimax analysis of stochastic problems. *Optim. Methods Software* **17**(1) 523–592.

Shapley, L. S. 1953. Stochastic games. *Proceedings of the National Academy of Science*, vol. 39. 1095–1100.

Sobel, M. J. 1969. Optimal average-cost policy for a queue with start-up and shut-down costs. *Operations Research* **17** 145–162.

Sorin, S. 1984. Big match with lack of information on one side i. *International journal of Game Theory* **13** 201–255.

Sorin, S. 1985. Big match with lack of information on one side ii. *International journal of Game Theory* **14** 173–204.

Stidham, S., R. R. Weber. 1989. Monotonic and insensitive optimal policies for control of queues with undiscounted costs. *Operations Research* **87**(4) 611–625.

van Damme, E. 1991. *Stability and Perfection of Nash Equilibria*. Springer-Verlag.

White, C. C., H. K. Eldeib. 1994. Markov decision processes with imprecise transition probabilities. *Operations Research* **42**(4) 739–749.

Yechiali, U. 1971. On optimal balking rules and toll charges in a gi/m/1 queueing process. *Operations Research* **19** 349–370.