Graduate Theses and Dissertations   Graduate School

2009

Generation capacity expansion in restructured energy markets

Vishnuteja Nanduri
University of South Florida

Follow this and additional works at: http://scholarcommons.usf.edu/etd

Part of the American Studies Commons

Generation Capacity Expansion in Restructured Energy Markets

by

Vishnuteja Nanduri

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Major Professor: Tapas K. Das, Ph.D.
José L. Zayas-Castro, Ph.D.
Alex Savachkin, Ph.D.
Kandethody Ramachandran, Ph.D.
Ralph E. Fehr, III, Ph.D.
Audun Botterud, Ph.D.

Date of Approval:
May 12, 2009

Keywords: Deregulated Electricity Markets, Generation Expansion Planning,
Matrix Games, Reinforcement Learning, and Conditional Value-at-Risk

DEDICATION

To Amma, Daddy, Anna, Vinnu, and Chaitu.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GENERATION CAPACITY EXPANSION IN RESTRUCTURED ENERGY MARKETS

Vishnuteja Nanduri

ABSTRACT

With a significant number of states in the U.S. and countries around the world trading electricity in restructured markets, a sizeable proportion of capacity expansion in the future will have to take place in market-based environments. However, since a majority of the literature on capacity expansion is focused on regulated market structures, there is a critical need for comprehensive capacity expansion models targeting restructured markets. In this research, we develop a two-level game-theoretic model, and a novel solution algorithm that incorporates risk due to volatilities in profit (via CVaR), to obtain multi-period, multi-player capacity expansion plans.

To solve the matrix games that arise in the generation expansion planning (GEP) model, we first develop a novel value function approximation based reinforcement learning (RL) algorithm. Currently there exist only mathematical programming based solution approaches for two player games and the N-player extensions in literature still have several unresolved computational issues. Therefore, there is a critical void in literature for finding solutions of N-player matrix games. The RL-based approach we develop in this research presents itself as a viable computational alternative. The solution approach for matrix games will also serve a much broader purpose of being able to solve a larger class of problems known as stochastic games.

This RL-based algorithm is used in our two-tier game-theoretic approach for obtaining generation expansion strategies. Our unique contributions to the GEP literature include the explicit consideration of risk due to volatilities in profit and individual risk preference of generators. We also consider transmission constraints, multi-year planning horizon, and multiple generation technologies. The applicability of the two-tier model is demonstrated using a sample power network from PowerWorld software. A detailed analysis of the model is performed, which examines the results with respect to the nature of Nash equilibrium solutions obtained, nodal prices, factors affecting nodal prices, potential for market power, and variations in risk preferences of investors. Future research directions include the incorporation of comprehensive cap-and-trade and renewable portfolio standards components in the GEP model.

PREFACE

First and foremost, I would like to thank my family for their constant support, love, and encouragement throughout my life. Thank you for standing by me while I successfully completed this long, arduous, and fruitful journey. I consider myself extremely blessed to have such an amazing family that is so supportive and encouraging, every minute of everyday. Mom and Dad, I am here because of your blessings, they sustain me. Anna and Vinnu, I cannot thank you enough for all the non-stop encouragement and love and for and keeping my spirits upbeat.

A debt of gratitude is owed to Dr. Tapas Das, my Guru, mentor, friend, and family, without whom this work would not be possible. He is a treasure chest of knowledge, and a constant source of inspiration. His hard work, persistent work ethic, and drive for excellence, have been a true source of motivation. I will fondly remember our thoughtful discussions about research, days and weeks of paper and proposal writing, and our many digressions into other topics of life, family, and beyond, over the past 6 years. I look forward to a lifetime of friendship, research collaborations, more joint publications, and last but not least, I will look to him for continued inspiration.

The time I spent at University of South Florida, and specially the Department of Industrial & Management Systems Engineering (IMSE), will remain some of the most precious moments of my life. It has been my home away from home and my family away from family. I want to thank Dr. José Zayas-Castro, the Chair of IMSE for everything he has done for me. His leadership in the department has been a great source of inspiration. His take on several issues such as research, education,

camaraderie, and his steadfast focus on the well being of students and faculty are tremendous. I have learned a great deal from him and I look forward to our continued friendship and collaborations. IMSE would not be the same without the dedicated work of Ms. Gloria Latter and Ms. Jackie Stephens. Thank you for everything from the bottom of my heart.

I am profoundly thankful to Dr. Das and Dr. Okogbaa for giving me the opportunity to serve as the Project Manager of the NSF funded GK-12 program-STARS (NSF Grant DGE 0638709). The GK-12 program and continued interactions with Dr. Das and Dr. Okogbaa and the managerial team of Wilkistar Otieno and Diana Prieto have helped my personal and professional growth tremendously. I wish to thank Dr. Das for the research assistantship via NSF ECCS grant # 0400268 and the IMSE department for the teaching assistantships, without which I would have never been able to come to USF and fulfill this dream.

My sincere and heartfelt thanks to Dr. Ramachandran for his guidance on several aspects of the reinforcement learning algorithm. I will continue to use his classnotes for the stochastic processes courses that I plan to teach in the future. I would also like to thank Dr. Ralph Fehr, Dr. Audun Botterud, and Dr. Alex Savachkin, for giving me some valuable suggestions that have improved the quality of the dissertation.

Thanks are due to my lab-mates and my closest friends during my doctoral studies at USF: Wilkistar Otieno, Patricio Rocha, and Diana Prieto. I will forever remember our lengthy impromptu discussions in the graduate lab about all topics under the sun and all the fun pot-lucks and get-togethers at each others homes, all of which have made my graduate education unforgettable. Thanks to all the INFORMS student chapter committee members, with whom I have served. I have learned a great deal from my constant interactions with all the STARS GK-12 Fellows. It was a wonderful experience. Thanks to those students that worked before with Dr. Das: Dr. Rajesh

CHAPTER 1

INTRODUCTION

Motivated by the success of deregulation in industries such as telecommunications, airlines, and transportation, the electric power industry restructuring was introduced in many parts of the U.S. as well as in many countries around the world. Electricity market restructuring has spurred a significant amount of research to model and subsequently improve our understanding of how various segments of the market perform and interact with one other. Due to the interactions of political, socioeconomic, and technological forces, the deregulated electric power industries both in the United States and abroad have undergone many structural transformations. Though significant differences exist in the working of markets around the world, the common goals of restructuring are the reduction of prices for the end-user, ushering in technological innovation, and increase of social welfare.

Despite some major initial setbacks in California, successful deregulated markets like Pennsylvania-New Jersey-Maryland (PJM) interconnection, New York Independent System Operator (NYISO), Electric Reliability Council of Texas (ERCOT), and several markets around the world, have reinvigorated the policy makers. Currently, over a fourth of the states across the U.S., and several countries around the world, notably UK, Nordic countries, and Australia trade electricity in a deregulated environment. Several insightful monographs ([1, 2, 3]) that deal with power system economics and operation of restructured markets exist in the literature. Recently, survey papers were presented to the power market literature by Ventosa et al. [4],

1

Day et al. [5], Boucher and Smeers [6], and Nanduri and Das [7]. The survey in [4] consists of an excellent overview of recent market modeling trends, and [5] and [6] discuss market equilibrium formulations respectively. However, the above monographs and survey papers do not shed light on the model solution approaches, and certain important issues like electricity auctions and solution approaches used to obtain optimal bidding strategies and Nash equilibria. These topics are discussed in detail in [7].

1.1 Generation Expansion Planning

According to the National Energy Policy (NEP) developed in 2001 and the Annual Energy Outlook 2007, energy demand in the U.S. is slated to increase sharply over the next two decades [8, 9]. It is stated in the NEP that the United States will need about 393,000 MW of new generating capacity by 2020 to meet this growing demand. With about fifteen States in the U.S. currently trading electricity in restructured markets, a significant proportion of the aforementioned capacity expansion will have to take place in a market based environment. Current literature is rich with research examining capacity expansion under the regulated market paradigm. However, there is a critical need for developing comprehensive capacity expansion models in restructured market settings. This research aims to address this need.

1.1.1 Regulated Settings v/s Restructured Settings

GEP in traditional settings is formulated as a least cost optimization problem that minimizes production and capital costs. GEP in restructured settings, on the other hand, needs to be modeled as a non-cooperative profit maximization problem. This is because GEP in restructured settings has multiple competing decision makers, as opposed to a single decision maker in traditional settings. Therefore, the investment

decisions made by a generator in restructured markets affect not only his/her profits but the other generators' profit as well, and hence the need to model it as a non-cooperative game.

Generation expansion planning (GEP) in a restructured market is the challenge of determining *which* type, *where*, and at *what* time periods new generation capacities are likely to be installed by the *competing generators* in response to: expected demand growth, changes in network conditions, and market design incentives. This research addresses the above challenge by developing a comprehensive matrix game model that subsumes electric power market features like multiple competing generators, a multi-year planning horizon, transmission constraints, and demand stochasticity. The model also explicitly considers risk due to volatilities in profit using a conditional value-at-risk measure as well as using individual generator risk preferences. The model has a two-tier matrix game construct that iteratively builds multi-year, multi-player expansion strategies for the competing generators. The expansion strategies from the model are obtained using a reinforcement learning based value function approximation algorithm for solving matrix games, which we present in Chapter 5 (see [10]).

1.2   Research Objectives

The objectives of this research are the following. Each of these broad research objectives are addressed in various chapters of the dissertation.

1. Develop a comprehensive matrix game model that addresses the challenge of generation capacity expansion in restructured electric power markets

2. Develop a solution algorithm to solve the matrix games embedded in the two-tier model

3. Perform a detailed empirical analysis of the matrix game solution algorithm

4. Formulate an overall solution framework, which uses the matrix game solution algorithm, to solve the GEP problem

5. Demonstrate the applicability of the GEP model using sample power networks

1.3 Research Contributions

This research makes some significant contributions in the advancement of the state-of-the-art both in generation capacity expansion planning as well as in solution approaches to multiplayer matrix games. Our novel two-tier matrix game model for generation expansion planning in restructured power market settings is the first of its kind. The two-tier model considers investment competition at the upper tier and the embedded supply function competition at the lower tier. The use of a reinforcement learning algorithm, as presented in Chapter 5 and in [10], shows promise in solving matrix games of relatively higher dimensionality. The contributions also include the incorporation of generator risk preferences and a measure of conditional value-at-risk (CVaR), which makes the investment decisions more robust. The model and its solution methodology are demonstrated on a sample network with five buses, seven transmission lines, three generators, and four loads. The simultaneous consideration of several important elements in expansion planning, such as, transmission constraints, risk, demand variations, multi-period planning horizon, and multiple generation technologies is not found in the existing literature.

The novel value function approximation based reinforcement learning algorithm for obtaining NE of n-player matrix games is a significant contribution to the literature. Extensive numerical experimentation is presented in Chapter 5, which demonstrates the ability of the learning algorithm to obtain Nash equilibrium. This section

includes sixteen matrix games with up to four players and sixty four actions for each player, followed by an example of a restructured power network with competing generators. The numerical results indicate that the learning based approach presented in this research holds significant promise in its ability to obtain NE for large n-player matrix games. To our knowledge, the algorithm is the first of its kind that harnesses the power of stochastic value approximation method that has been successfully used in solving large scale Markov and semi-Markov decision process problems with single decision makers ([11, 12, 13]). A formal proof establishing the convergence of the algorithm to Nash equilibrium solutions is not fully developed yet, and is currently being investigated. However, as discussed in Chapter 6, the empirical evidence clearly indicates the algorithms' ability to converge to NE solutions.

1.4 Dissertation Outline

A brief overview of power market equilibria and generation expansion planning literature can be found in Chapter 2. Some fundamental concepts of game theory and solution approaches to game theory problems as found in power market literature are presented in Chapter 3. The comprehensive two-tier matrix game model addressing GEP in restructured settings is developed in Chapter 4. To solve these embedded matrix games, we develop a value function approximation based reinforcement learning algorithm, which is presented in detail in Chapter 5. An empirical analysis of the performance of the RL algorithm is presented in Chapter 6. The solution framework used to solve the two-tier GEP model is presented in Chapter 7. Chapter 8 consists of a demonstration of the applicability of the model via a sample problem. The chapter contains detailed examination of model results as to the nature of expansion plans, generator payoffs, and nodal prices, for given demand growth. A regression model is developed to identify the factors affecting nodal prices post-expansion. This chapter

also examines consumer surpluses under strategic bidding, and choice of generator expansion plans under varying risk preferences. Concluding remarks based on this work are presented in Chapter 9 and some future research directions such as the inclusion of cap-and-trade programs and renewable portfolio standards for CO_2 emission control are reserved for Chapter 10.

CHAPTER 2

LITERATURE REVIEW

2.1 Generation Capacity Expansion Planning

Until the late nineties, a significant number of papers appeared in the literature examining the generation expansion planning (GEP) process in regulated electricity markets. Some of the key contributions are [14, 15, 16, 17, 18, 19]. An excellent review paper by Zhu and Chow [20] discusses both mathematical programming based and heuristic based techniques used to solve GEP problems in regulated settings. Most of the papers listed above formulated the GEP problem with the objective of *minimizing production and capital costs.* On the other hand, GEP in restructured markets needs to be modeled as a noncooperative game, where the generators *compete to maximize their profits.* In both cases, however, the constraints have to include capacity, transmission, energy balance, investment, and system reliability. While commercial software like Wien Automatic System Planning (WASP [21]) and Electric Generation Expansion Analysis System (EGEAS [22]) exist to address GEP in regulated markets, GEP research in deregulated markets is still in its early stages. Hence, there is a need and room for more research to fully explore and understand the GEP problem in the current competitive environment. The current competitive environment has introduced some rather new challenges in the already complex GEP problem. First of which is the modeling of competitive behavior of generators. Second, the generators investing in the market have to consider individual risks due to volatilities in

profits. Third, emissions and environmental regulations are set to become critically important in the coming years, due to which, comprehensive models incorporating cap-and-trade/carbon trading mechanisms will become necessary.

In the next few paragraphs, we focus our attention on some of the recent research contributions addressing GEP in restructured markets. The papers that we have chosen to review here, help to highlight the similarities that our model shares with the literature, as well as the distinctions. Chuang et al. [23] presented one of the initial GEP models in a restructured setting. They model GEP as a Cournot game by making the following main assumptions: generators compete only in quantities, new entries do not occur in the middle of the game, and all generators make invest-ment decisions simultaneously. While we also make the latter two assumptions, we use supply function competition instead of Cournot competition to more accurately represent power market bidding. Chuang et al. compute the price and quantity al-locations of generators using the California ISO/Power Exchange (PX) system, that buys and sells energy through auctions. Finally, the solution of the GEP Cournot game is obtained using a simple iterative search procedure. We use an optimal power flow formulation to obtain price-quantity allocations and then compute the generator profits. Thereafter, a value function approximation based learning algorithm is used to find the solution of the game.

Murphy and Smeers [24] present three different GEP models. The first model, which considers perfect competition, is developed to serve as a base line case for com-parison against the other two models. The second model is an open loop Cournot model where investment decisions and power dispatch occur simultaneously. The third model is a 2-stage equilibrium problem with investments in stage 1 and power dispatch in stage 2. This 2-stage model is an extension of the MPEC type problems, which while realistic, are often extremely difficult to solve and are fraught with conver-

Table 2.1. Important Modeling Attributes from GEP Literature

Authors	Model	Risk	Emissions	Transmission Constraints	System Reliability	Multiyear Horizon	Demand Variations	Technology
Chuang et al. [23]	Cournot Game	No	No	No	Yes	No	No	Yes
Murphy and Smeers [24]	Cournot Game	No	No	No	No	No	No	Yes
Kaymaz et al. [25]	Cournot Game	No	No	Yes	Yes	No	No	No
Jirutitijaroen and Singh [26]	Optimization	No	No	Yes	Yes	No	No	No
Kim et al. [27]	Cournot Game	No	No	No	Yes	Yes	No	Yes
Ng et al. [28]	Cournot Game	No	No	Yes	Yes	No	No	Yes
Ehrenmann and Smeers [29]	Cournot Game	No	No	No	No	Yes	No	No
Ehrenmann and Smeers[29]	Stackelberg Game	No	No	No	No	Yes	No	No
Botterud et al. [30]	optimization	Yes	No	No	No	No	No	No

gence related challenges [25]. In this research we adopt the strategy of simultaneous investment and power dispatch, similar to the second model in [24].

One of the important features missing from the models in [23] and [24] is the consideration of transmission constraints. Kaymaz et al. [25] include transmission constraints and extend Hobbs's LCP formulation [31] for power markets by incorporating GEP-related decision variables in the objective function. Their model results, as expected, show that transmission constraints affect the capacity expansion decisions. They also show that transmission constraints adversely affect the consumer benefits, which often tend to reduce due to high electricity prices and congestion rents. In this research we also consider transmission constraints. For the sake of brevity, we summarize other relevant GEP literature and their key modeling attributes in Table 2.1.

2.2 Nash Equilibria of Multiplayer Games

Globalization has played a significant role over the last decade in transforming the marketplace into one where most goods and services are transacted through multi-party competition. Consequently, the study of game theoretic concepts and the de-

velopment of effective methods for solving multiplayer games have gained increasing attention in the research literature. Games occur in two primary forms: matrix games and stochastic games. An n-player matrix game is characterized by n different reward matrices (one for each player) and a set of action combinations characterizing the equilibria (Nash-equilibria, in particular). Nash [32] defined equilibrium to be an action combination from which no single player could unilaterally deviate to increase profit. Stochastic games are comprised of finite or infinite horizon stochastic processes with finite states and state transition probability structure, in which the players seek equilibrium actions for every state so as to maximize their rewards from the overall game. Therefore, stochastic games are construed as sequence of matrix games (one for each state) connected with transition probabilities. Further classification of games arises from the nature of reward structure: zero sum games and nonzero (general) sum games. Rewards of stochastic games are classified as discounted reward, average reward, and total reward.

Though the fundamentals of game theory are fairly well established ([32]), the computational difficulties associated with finding Nash equilibria have constrained the scope of the research literature largely to the study of bimatrix games with limited action choices. Even in the absence of sufficient tools to appropriately analyze stochastic or matrix games, a majority of the marketplaces have evolved to incorporate transactions through competition. Therefore, to ensure healthy growth of the current competition based economy, it is imperative to develop computationally feasible tools to solve large scale stochastic and matrix games. In recent years, researchers have been able to characterize equivalent matrix games for both discounted reward and average reward stochastic games ([33, 34, 35, 36]). They also harnessed the advances in reinforcement learning based techniques to construct these equivalent matrix games ([33, 34]). However, obtaining the Nash equilibrium for these equivalent

matrix games has remained an open research issue, which is one of the foci of this research.

As discussed in [37], the appropriate method of computing Nash equilibria of a matrix game depends on whether it is required to find one or all equilibrium points, the number of players in the game, and the importance of the *value* of the Nash equilibrium. No computationally viable method addressing all of the above is available in existing literature. Nash equilibria of n-player matrix games can be obtained by solving a nonlinear complementarity problem (NCP), which for a 2-player matrix game becomes a linear complementarity problem (LCP) ([37]). Lemke and Howson [38] developed an efficient algorithm for obtaining Nash equilibria for bimatrix games by solving the associated LCP. Their algorithm was extended for finding Nash equilibria of n-person matrix games in [39] and [40]. However, these algorithms still have unresolved computational challenges. Mathiesen [41] proposed a method of solving NCP for n-player matrix games through a sequence of LCP approximations. A survey by [42] summarizes these and other developments on this topic. It may be noted that these methods are not guaranteed to obtain global convergence and often depend on the choice of the starting point. To our knowledge, the only openly available software that attempts to solve multiplayer matrix games is GAMBIT ([43]). However, as observed by Lee and Baldick [44], this software takes an unusually long computation time as the number of players and their action choices increase.

Game theoretic models have been studied extensively in examining market competition in the energy and transmission segments of restructured power markets (as in Pennsylvania-Jersey-Maryland, New York, New England, and Texas). These games are characterized by multidimensional bid vectors with continuous parameters. Upon suitable discretization of these bid vectors, many of these games can be formulated as matrix games. The degree of discretization dictates both the computational burden

11

and the probability of identifying the Nash equilibria. Almost all of the literature studying power market games is devoted to optimization based approaches, such as mathematical programming ([45, 46, 47]), co-evolutionary programming ([48]), and exhaustive search ([49]). Even in a limited number of studies, where such games are formulated as matrix games, numerical examples are converted to bimatrix games and are solved using linear programming and LCP approaches ([44, 50, 51]).

Mathematical programming approach to finding NE of matrix games has two primary variants: mathematical program with equilibrium constraints (MPEC, [52]), and equilibrium problem with equilibrium constraints (EPEC, [53]). MPEC is a generalization of bilevel programming, which in turn is a special case of hierarchical mathematical programming (with two or more levels of optimization). MPECs resemble Stackelberg (leader-follower) games, which form a special case of the Nash game. In a Nash game each player possesses the same amount of information about competing players, whereas, in Stackelberg type games, a leader can anticipate the reactions of the other players, and thus possesses more information in the game. The leader in a Stackelberg game chooses a strategy from his/her strategy set, and the followers choose a response based on the leaders actions ([52]), while in a Nash game all players choose actions simultaneously. When multiple players face optimization problems in the form of MPECs, EPEC models have been used to simultaneously find the equilibria of the MPECs ([47, 53, 54, 55]). MPEC, LCP, and EPEC problems are discussed briefly in the next chapter along with some game theory fundamentals.

CHAPTER 3
BASIC CONCEPTS

The objective of this chapter is to cover some fundamentals of game theory, reinforcement learning, and discuss equilibrium strategies from the perspective of power market operations. We present a detailed review of some papers that develop methodologies to obtain power market equilibria.

3.1 Basic Concepts of Game Theory

Game theory examines the behavior of *rational players* in interaction with other *rational* players. Players are considered to be *rational* if they maximize their objective functions given their beliefs about the environment. In a game theoretic setting, players act in an environment where other players' decisions influence their payoffs. The concept of strategy as a complete plan of action provides an approach for modeling behavior that takes informational as well as dynamic characteristics of the environment into account.

3.1.1 Zero Sum and Non-Zero Sum Games

Games can be classified based on payoff structure as zero sum games and non-zero sum games. A *two player zero-sum game* is a game in strategic form such that

$$p_1(s_1, s_2) + p_2(s_1, s_2) = 0, \qquad \forall \ s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2 \tag{3.1}$$

where p_1, p_2 are the payoff functions of two players and $\mathcal{S}_1$ and $\mathcal{S}_2$ are the pure strategy sets of the two players.

From the above definition, it is seen that zero sum games are strictly competitive which means that what one player gains the other loses. In non-zero sum games some outcomes are more favorable to some players than others. Some outcomes may even yield a positive payoff and others a negative payoff for every player. This introduces a certain common interest among players to attain such more favorable outcomes even if they are not the most favorable outcomes for everyone. Such games are non-strictly competitive since they have both competitive and cooperative elements.

3.1.2   Pure and Mixed Strategy

The concept of strategy is fundamental to game-theoretic analysis as it provides a complete plan to the player for how to play the game. When players play each strategy with probability one, then the players are said to have a pure strategy. A mixed strategy simply means that the players randomly choose a pure strategy. Thus a mixed strategy is a probability distribution on the set of pure strategies. The set of mixed strategies always includes all pure strategies because a pure strategy can be considered as a special case of a mixed strategy in which the respective pure strategy is played with probability one and any other pure strategy with probability zero.

3.2   Equilibrium in Power Market Games

Participants of the energy market attempt to maximize their benefits by seeking optimal bidding strategies. A generic version of the bidding strategy formulation problem in a power network can be given as follows. Let $\mathcal{B}$ denote the set of buses in the network, and $\mathcal{B}_s \subset \mathcal{B}$ denote the subset of supply buses (nodes). Let the number of generators at a supply bus $i \in \mathcal{B}_s$ be denoted by N_i, and M denote the number

14

of loads in the network. Let $\mathcal{G}_i = \{1, 2, \cdots, N_i\}$ and $\mathcal{L} = \{1, 2, \cdots, M\}$ denote the set of generators at a supply bus i and the set of loads in the network respectively. Let $N = \sum N_i$, and $\mathcal{G} = \cup \mathcal{G}_i$. To keep the exposition simple, we consider only generator side bidding in the market.

Let the state of the network at time t (X_t) be the vector of realized loads (demands) q_t^s and prices p_t^s. Hence, $X_t = \{q_t^s, p_t^s\}$, where $q_t^s = (q_t^1, q_t^2, \cdots, q_t^{|\mathcal{B}|})$ and q_t^s denotes the realized hourly load quantity vector at the sth bus, $s \in \mathcal{B}$. Also, $p_t^s = (p_t^1, p_t^2, \cdots, p_t^{|\mathcal{B}|})$, where p_t^s represents the realized hourly price vector at bus $s \in \mathcal{B}$.

Let the bid decision vector at the tth time be given by $D_t = \{\mathcal{D}_t^l : l \in \mathcal{G}\}$, where $\mathcal{D}_t^l$ is the decision vector of generator l and $\mathcal{D}_t^l \in \{D^l\}$ and D^l denotes the set of all bid parameters vectors for generator l. These bid parameters depend on the nature of bids, for example, polynomial functions and piecewise linear functions, and determine the offer prices corresponding to the generation quantities. The bidding process involves selection of bid parameters by the generators, who seek to maximize their individual profits for the forecasted state of the network X_t. The profits corresponding to a set of bids submitted at any time t by the generators are obtained by solving the optimal power flow (OPF) model. The profit maximization problem for generator j, as commonly presented in the literature, can be stated as a bi-level problem as follows. Choose $\mathcal{D}_t^j$, so as to maximize profit $g(f_t^j, P_t^j)$, subject to choice of other bidders $\mathcal{D}_t^l : l \in \mathcal{G} \setminus j$ and the OPF Problem and its constraints. Where, f_t^j and P_t^j are the nodal clearing price (cost of power generation) and quantity allocation for generator j as determined by the OPF model, which is provided next.

OPF models are formulated either to maximize social welfare or to minimize the total cost of meeting the power demand of a network. The OPF model simultaneously satisfies several system related constraints such as demand and supply constraints, voltage constraints, thermal limit constraints, and the constraints of power flow. Sev-

eral papers presented to the literature utilize a DC version of the OPF model to curtail the computational complexity involved in solving an AC-OPF model. We adopt a similar approach in our work. However, for the sake of completeness, we provide below a generic mathematical formulation of the cost minimization version of the AC-OPF model.

Let f_t^j denote the cost of active power generation by supplier j at a decision epoch. Also, let P_t^j and Q_t^j denote the active and the reactive power generation quantities respectively.

$$\min \sum_{j \in \mathcal{B}_s} f_t^j(P_t^j) \tag{3.2}$$

$$\text{subject to: } \sum_{j \in \mathcal{B}_s} P_t^j - l - l(V, \theta) = 0, \tag{3.3}$$

$$\sum_{j \in \mathcal{B}_s} Q_t^j - \tilde{l} - \tilde{l}(V, \theta) = 0, \tag{3.4}$$

$$S_{y,z} \leq S_{y,z}^{max} \forall \ y \neq z \in \{\mathcal{B}\} \tag{3.5}$$

$$V_w^{min} \leq V_w \leq V_w^{max}, \quad \forall \ w \in \{\mathcal{B}\}, \ \mathcal{B} = \{\text{set of buses}\}. \tag{3.6}$$

$$P_{min}^j \leq P_t^j \leq P_{max}^j, \quad \forall \ j \in \{\mathcal{B}_s\} \tag{3.7}$$

$$Q_{min}^j \leq Q_t^j \leq Q_{max}^j, \quad \forall \ j \in \{\mathcal{B}_s\} \tag{3.8}$$

Constraint 3.3 in the OPF model ensures that all the active demand (l) and the active transmission losses ($l(V, \theta)$) are met by the generators selected for dispatch at any given time (active power balance equation). The constraint 3.4 ensures that all the reactive demand ($\tilde{l}$) and the reactive transmission losses ($\tilde{l}(V, \theta)$) are met by generators selected for dispatch (reactive power balance equation). The term $S_{y,z}$ in equation 3.5 denotes the flow limit for the power transmitted from Bus y to Bus z. Constraint 3.5 ensures that the maximum flow limit constraints in both directions are not violated. The constraint 3.6 is used to maintain the voltage limits for each Bus.

Constraints 3.7 and 3.8 are used to maintain active and reactive power generation limits.

3.3 Solution Strategies

Note that the bi-level bidding strategy problem is presented above from the perspective of profit maximization of generator j. But, the requirement of the knowledge of bid choices of the other players, as stated in the constraint set, makes the bi-level problem unsolvable in a deregulated market, where bid choices are not known a priori. Thus, the optimal generator bids should be derived from the Nash equilibrium strategies of the game. However, nonavailability of computationally viable tools to solve for Nash equilibria of multiplayer games had motivated researchers to look for alternative approaches to obtain optimal bidding strategies. For the purpose of examining the existing literature, we classify these contributions into two major categories: approaches that *optimize individual strategies* for given strategies of other players, and approaches that *seek equilibrium strategies*.

3.3.1 Optimization of Individual Bidding Strategies

Several different optimization approaches have been used for this task including genetic algorithms ([56], [57], [58]), evolutionary programming ([59]), Monte Carlo simulation ([60]), dynamic programming ([61], [62]), and mathematical program with equilibrium constraints ([52, 46]). In what follows, we review the key contributions and limitations of the above papers.

The work presented in [56] offers a *genetic algorithm* (GA) approach to optimizing profits of individual generators having multiple generating units. Solution of individual generator profits are obtained by assuming that the bids of other players are known in the form of probability distribution functions. GA is used as a means to

17

navigate through the large actions spaces D_j of the individual generators $j \in \mathcal{G}$ while considering randomized bidding behavior of the other players. The solutions thus obtained do not have any equilibrium properties, since in a noncooperative bidding environment, no rational generator can be expected to behave randomly guided by a probability density function. As a result, the expected generator profits calculated by the algorithm are unlikely to be ever realized.

Attaviriyanapap et al. [59] present an *evolutionary programming* approach to finding bidding parameters that maximize individual generators profits. The authors attempt to obtain optimal bidding strategies of a supplier who owns multiple generating units. The clearing price f_j^t is obtained using a PX-type market settlement (simple matching of supply and demand curves) for 24 hours of the day. The role of EP in this paper is to simply search through the decision space for profitable bids. Due to lack of consideration of OPF and transmission constraints (like the Equations 3.3 - 3.8) the use of such models in real power markets is ineffective.

Wen and David use a *Monte Carlo (MC) simulation method* to obtain optimal generator bidding strategies in [60]. In [60], the authors consider rivals bids ($\mathcal{D}_l^t : l \in \{\mathcal{G} \setminus j\}$) to be available in the form of probability density functions and subsequently use MC simulation to obtain random samples from these bid pdf's. These samples are then considered to be fixed in the overall generator bidding strategy problem. Then, an elementary search technique known as *golden section* method used in finding the profit maximizing bid. However, it may be remarked here that the assumption of probabilistic estimation of rivals bids affects the ability of this approach to attain true optimality.

Rajaraman and Alvarado [61] present a deterministic nested *dynamic programming* (DP) approach of finding optimal bidding strategies for multi period power market problems. DP-based approaches are suitable for small scale problems where

Table 3.1. Some Important Modeling Attributes from Bidding Strategy Literature

Solution Methodology	Overall Problem Structure	# of buses	Market Clearing	Type of Bids
Genetic algorithms [56]	Two level optimization	9-bus	DC-OPF	linear supply functions
Genetic algorithms [57]	Traditional optimization	24-Bus	PX-type	linear supply functions
MPEC [46]	Bi-level optimization	30-Bus	DC-OPF	linear supply functions
Evolutionary programming [59]	Traditional optimization	10-Bus	PX-type	linear supply functions
Monte Carlo simulation [60]	Stochastic optimization	6-Bus	DC-OPF	linear supply functions
Dynamic programming [62]	Two level optimization	5-Bus	PX-type	step function bid curve

decisions from one period affect the decisions and profits in subsequent periods (day ahead auction markets). The authors in [61] present several cases with consideration of hydro and thermal generators as well as cases with price making and price taking generators. However, their study does not consider multiple competing generators or transmission constraints. Also, the authors assume that the transition probability matrices (TPMs) are readily available. However, it is well known that even for problems of relatively small sizes, determination of TPMs becomes almost impossible. As a result of such computational and modeling limitations, the approach presented in [61] cannot be applied to large transmission constrained networks having multiple competing generators. Nevertheless, the DP model may serve as a guidance tool for individual generators in determining profitable bidding strategies, for very small networks with limited state spaces.

Hobbs et al. [46], present a *mathematical program with equilibrium constraints* (MPEC) approach to finding optimal bidding strategies of generators in a power network. The authors assume that while making their own bid all generators have complete information about rival players' bids. A bilevel optimization model is formulated, where a generator's profit maximization problem at the first level is subjected to the OPF constraints at the second level. As part of the MPEC procedure, the

OPF constraints are then replaced with equivalent KKT conditions resulting in a linear complementarity problem framework (LCP). This 2-level problem, known as MPEC, has a maximization problem in the first level and equilibrium constraints in the second level. Such problem structures have been gaining significant attention lately due to their widespread applicability in a variety of fields such as chemical engineering, transportation science, and power system economics. For this reason, we chose to present a generic formulation of an MPEC problem based on [63].

$$Max_{x,y,z}\Pi(x,y,z)$$

$$\text{Subject to:} 0 \leq F(x,y,z) \perp x \geq 0,$$

$$G(x,y,z) = 0,$$

$$z \in S,$$

$$x,y,z \in \Re,$$

(3.9)

where z represents first level variables and x and y represent second level variables, which must satisfy an LCP with fixed values of z from the first level. In general, $0 \leq x \perp y \geq 0$ is read as $x \geq 0$, $y \geq 0$, and $xy = 0$. In the power market context, the first level variables are generator bids (similar to $\mathcal{D}_t^l$) which serve as fixed parameters in the second level OPF problem. The above MPEC problem is a non-convex optimization problem, which has to be solved using special solution algorithms such as the penalty interior point (PIP) method. Details of the PIP algorithm can be found in [46]. Table 3.1 presents some important attributes of bidding strategy formulation problems available in literature.

3.3.2 Approaches Seeking Equilibrium Strategies

In a competitive power network with multiple participants, Nash equilibria (NE) is that combination of strategies from which no market participant has the incentive to unilaterally deviate. This can be mathematically stated as:

$$g(x_j^*, x_{-j}^*) \geq g(x_j, x_{-j}^*) \quad \forall j \tag{3.10}$$

where, x_j^* is the optimal bid of a participant j, and x_{-j}^* are the optimal bids for all other participants. As alluded to earlier, due to nonavailability of computationally viable approaches to find NE strategies, many researchers have approached the problem from two different viewpoints: 1) individual generators' profit maximization perspective (discussed earlier), and 2) methodologies that solve for equilibria of Nash games by making assumptions about the competitive bidding behavior of generators (explained next). Some of these assumptions are Nash-Cournot, Nash-Bertrand, and Nash-supply function, where all players bid simultaneously. These assumptions are explained next followed by a detailed discussion of the equilibrium seeking methodologies.

1. Nash-Cournot Competition: Under the Cournot assumption the generators compete only with quantities. Each generator assumes that the opponents quantity is fixed and then makes his/her own quantity decision. Then the game is solved for a Nash-Cournot equilibrium, where no generator gains by unilaterally deviating from his/her bid quantity.

2. Nash-Bertrand Competition: Under the Bertrand assumption the generators compete with prices. Each generator assumes that the opponents price is fixed

and then makes his/her own price bid. The NE obtained under such competition is termed as Bertrand-Nash equilibrium.

3. Nash-Supply Function Competition: Supply functions are price-quantity curves submitted by generators to the ISO. Supply function competition is often argued to represent the working of ISO-type power markets more closely than Cournot and Bertrand type competitions. The resulting equilibria are known as Nash-supply function equilibria.

4. Stackelberg Competition: Unlike in the above three Nash games, in certain oligopolistic situations, it is assumed that one of the players has more information than the rest. Such an assumption leads to the so-called Stackelberg game. In a Stackelberg game, a "leader" makes a decision first, and then the "followers" make their decision knowing the leader's decision. Such competition has been shown to be useful in modeling oligopolistic markets with a large dominating firm and a few smaller competing firms.

Even though the above assumptions have been extensively used in bidding strategy literature, it may be noted that the premise of complete information about rivals bids before making one's own bidding decision is not representative of non-cooperative power market games. In the remainder of this section we briefly discuss some approaches to find NE bidding strategies of power market games: linear complementarity Problems (LCP), equilibrium problem with equilibrium constraints, and reinforcement learning (RL) based approach.

3.3.2.1   LCP

A general formulation for linear complementarity problems (LCP) from [64] is given here. The objective is to find variables w and z where $w = (w_1, \cdots, w_n)^T, z =$

$(z_1, \cdots, z_n)^T$ satisfy $w - mz = q$, and $w \geq 0, z \geq 0$ and $w_i z_i = 0$ $\forall i$. Hobbs [31] uses such a framework to identify market equilibria in a POOLCO setting. He defines market equilibrium as those set of prices, supply, demand, and line flows that simultaneously satisfy each market participants first order conditions for maximizing profit while matching network demand and supply.

The LCP framework from [31] is presented here for exposition. For a constrained optimization problem, as the one given below,

$$Max \ \ F(x, y),$$
$$Subject\ to:$$
$$G(x, y) = 0,$$
$$H(x, y) \leq 0,$$
$$x \geq 0,$$

(3.11)

the KKT conditions can be written as follows:

$$x : \partial F/\partial x - \lambda \partial G/\partial x - \mu \partial H/\partial x \leq 0; \ \ x \geq 0,$$
$$x(\partial F/\partial x - \lambda \partial G/\partial x - \mu \partial H/\partial x) = 0,$$
$$y : \partial F/\partial y - \lambda \partial G/\partial y - \mu \partial H/\partial y \leq 0,$$
$$\lambda : G(x, y) = 0,$$
$$\mu : H(x, y) \leq 0, \ \ \mu \geq 0, and \ \ \mu H(x, y) = 0.$$

(3.12)

The equations associated with the non-negative variables are known as *complementarity conditions*, and λ and μ are the dual variables pertaining to the constraints G and H [31]. Hobbs develops such KKT conditions and combines them with the market clearing conditions. The first order KKT optimality conditions together with the market clearing conditions form the LCP. An equivalent quadratic program can then be written for the LCP and solved using standard solvers available in GAMS software. Another paper which discusses power market games, [65], utilizes the well established Lemke-Howson algorithm of solving LCPs. In [65], the LCP is formulated from a bimatrix power market game. It may be noted that, while LCP's have been shown (both theoretically and computationally) to obtain NE of 2-player games, nonlinear complementarity problem (NCP) frameworks have only been theoretically presented to solve games with more than two players. The proposed approaches of solving multiplayer games, such as ([39, 40]), still have unresolved computational challenges.

3.3.2.2 EPEC

The MPEC optimization approach presented earlier can be extended to a game theoretic setting with multiple competing players, known as equilibrium problem with equilibrium constraints (EPEC). In EPEC, each player is solving an MPEC problem subject to a set of common OPF constraints. We adopt the same notation used in the MPEC problem discussed earlier. Let all K players have the first level decision variables z_k, $k = 1 \cdots K$. The EPEC problem can now be stated as follows [63].

$$z_k^* \quad solves \quad Max_{x,y,z_k} \Pi^k(x, y, z_k, z_{-k}^*)$$

$$Subject\ to:$$

$$0 \le F(x, y, z_k, z_{-k}^*) \perp x \ge 0,$$

$$G(x, y, z_k, z_{-k}^*) = 0,$$

$$z_k \in S_k, and$$

$$x, y, z \in \Re.$$

$$(3.13)$$

The variables z^*_{-k} represent optimal and fixed values of opponents. According to [66], there are two general methods to solve the EPEC problem: obtain the optimality conditions (KKTs) for all the MPEC problems and solve them together as a complementarity problem, or iteratively solve each of the MPECs using standard MPEC algorithms (like PIP) until the equilibrium solution of the EPEC game is obtained. The EPEC problem is extremely complicated and moreover does not guarantee an NE solution. If a solution does exit, it is called a subgame perfect Nash equilibrium. Some good applications of EPEC models have been presented in ([53, 47, 54, 55]).

3.3.2.3 RL Based Approach

Value function approximation based Reinforcement Learning (RL) approach, which we develop in this research, to finding NE differs significantly from the mathematical programming approaches like EPEC, NCP, and LCP. Unlike in the mathematical programming approaches, where one assumes complete knowledge of rivals bids, in our approach, all players compete simultaneously without knowledge of other players actions. Such a framework, we believe, represents the true noncooperative game amongst power market participants. In Chapter 5, we use the well established *value approximation mechanism* which was previously successfully employed in solving large scale, Markov and semi-Markov decision process problems with a single player, [67], to develop a reinforcement learning based algorithm that solves for NE of multiplayer noncooperative games. We next present some basics of the reinforcement learning approach.

3.4 Brief Overview of Reinforcement Learning

The theory of RL is founded on two important principles: Bellman's equation and the theory of stochastic approximation ([68, 69]). Any learning model contains four basic elements:

1. System environment (*simulation model*)

2. Learning agents (*market participants*)

3. Set of actions for each agent (*action spaces*)

4. System response (*participant rewards*)

Consider a system with three competing market participants. At a decision making epoch when the system is in state s, the three learning agents that mimic the market participants select an action vector $(a = (a^1, a^2, a^3) \; \epsilon \; \mathcal{A})$. These actions and the system environment (model) collectively lead the system to the next decision making state (say $s\prime$). As a consequence of the action vector (a) and the resulting state transition from s to $s\prime$, the agents get their rewards $(r^1(s, a, s\prime), \; r^2(s, a, s\prime))$, and $r^3(s, a, s\prime))$ from the system environment. Using these rewards, the learning agents update their knowledge base (R-values, also called reinforcement value) for the most recent state-action combination encountered (s, a). The updating of the R-values is carried out slowly using a small value for the *learning rate*. This completes a learning step. At this time the agents select their next actions based on the R-values for the current state $s\prime$ and the corresponding action choices. The policy of selecting an action based on the R- values is often violated by adopting a random choice, which is known as *exploration*, since this allows the agents to explore other possibilities. The probability of taking an exploratory action is called the *exploration rate*. Both learning and exploration rates are decayed during the iterative learning process. This

process repeats and the agent performances continue to improve until the process converges to the optimal solution. For a detailed description of RL, its applications, and recent advances, the readers are referred to the texts by Gosavi [70], and Sutton and Barto [71]. In the next chapter, we present a two-tier matrix game-theoretic model to obtain generation expansion plans for competing generators.

CHAPTER 4

GENERATION EXPANSION PLANNING MODEL

4.1   Two-Tier Matrix Game Model for GEP

The generation expansion planning model that we propose in this research consists of two tiers, as shown in Figure 4.1. The top tier of the model represents the investment competition amongst generators. This competitive decision making scenario is modeled as a matrix game and is henceforth referred to as *investment game*. The bottom tier, on the other hand, represents the competition amongst generators to supply electricity into the network. This scenario is also modeled as a matrix game and is referred to as *supply function game*. It is called a supply function matrix game due to the fact that the generators are assumed to compete with supply functions (Cournot or Bertrand competitions can be used as well). Each strategy combination of the investment game represents a possible generation capacity expansion alternative. Therefore, for each such alternative, there exists a corresponding supply function game, which when solved allows the examination of the profitability of each expansion alternative. We next present the investment matrix game model (top tier) and supply function matrix game model (bottom tier) in detail. We also explain how the payoff matrices of these matrix games are calculated, and how the two tiers interact with each other in order to result in a multi-year, multi-player generation expansion strategy.
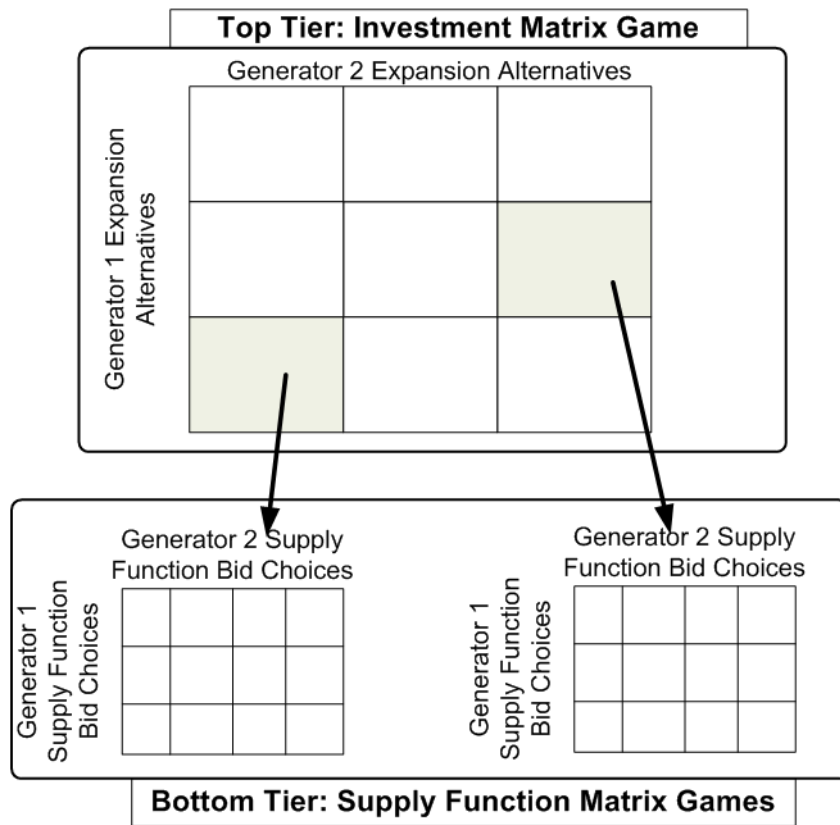
Figure 4.1. Schematic of the Two-Tier GEP Model for a Two Generator Scenario

4.1.1 Top Tier: Investment Game

The investment matrix game is defined by a tuple $< N, A^1, \ldots, A^N, R^1, \ldots, R^N >$. The elements of the tuple are as follows.

1. N denotes the number of generators.

2. A^k denotes the set of expansion alternatives available to generator k.

3. $R^k : A^1 \times \ldots \times A^N \to R$ is the payoff function for generator k, where an element $r^k(a^1, \ldots, a^N)$ of R^k is the risk constrained payoff (explained later) of generator k.

R^k for all k, can be written in the form of N-dimensional matrices representing the investment matrix game as follows

$$R^k = \left[r^k(a^1, a^2, \cdots, a^N) \right] \cdot {}_{a^1=1,\ldots,a^N=1}^{|A^1|,\ldots,|A^N|} \tag{4.1}$$

The generators select expansion alternatives from the set of available choices with the goal of maximizing their payoffs which depend on all other generators' selections. The concept of Nash equilibrium is used to describe a strategy as being the most rational behavior by the generators acting to maximize their payoffs. So, for the investment matrix game, a pure strategy Nash equilibrium is a collection of expansion alternatives $\mathbf{a}^* = (a_*^1, \cdots, a_*^N)$, for which $r^k(a_*^k, a_*^{-k}) \geq r^k(a^k, a_*^{-k})$, $\forall a^k \in A^k$, and $k = 1, 2, \cdots, N$, where a^k indicates the selection of a non-Nash equilibrium alternative by the kth generator and a_*^{-k} indicates the Nash equilibrium choice of all the other generators. We developed a risk constrained profit calculation model for obtaining the payoff matrices R^k. This model is presented next.

4.1.1.1 Risk Constrained Profit Calculation Model

In electric power markets, the amount of revenue earned by a generator, depends on the interaction of several factors such as strategic bidding behavior of the competing generators, transmission constraints, system contingencies (line failures, generator outages), fuel price volatilities, and demand variations. For instance, in a transmission constrained power network, the generators may be able to use location in the network to their advantage to bid strategically and make higher profits. On the other hand, unforseen outages and fuel price volatilities may adversely affect the profits. Conditional value at risk (CVaR), also known as expected shortfall (ES), is a risk metric that can be used by generators to capture such variabilities for profit calculations. As noted in the literature ([72, 73]), CVaR is gaining popularity in the finance and insurance industries as a viable risk metric. Before we discuss CVaR, it is important to describe how it is an extension of the traditionally used metric called value-at-risk (VaR). In addition to our work, to our knowledge, the only other paper in open GEP literature that uses CVaR to aid GEP investment decisions is [74]. The following discussion about VaR and CVaR is based on [75].

Let Z be the random variable which indicates the return on an investment. Let $\alpha = A\% \in (0,1)$ represent a percentage of worst case scenarios of the return on the investment. Then the VaR with respect to the $z^{(\alpha)}$ quantile of the of worst case scenarios is given as,

$$VaR^{(\alpha)}(Z) = -\sup\{z|P[Z \leq z] \leq \alpha\}. \tag{4.2}$$

However, VaR has two fundamental deficiencies: a) it is the threshold of losses in the worst case scenarios, and does not provide any information about those losses that may be significantly greater than VaR, and b) it does not satisfy the property

31

of subadditivity, which states that the global risk of a portfolio is always less than or equal to the sum of the risks of the individual assets (for more details see [72, 73]).

These weaknesses of VaR motivated researchers to develop a new, yet, related metric called conditional value-at-risk (or expected shortfall, ES). In [75], CVaR has been shown to address both the above weaknesses of VaR. While VaR is the minimum of the A% worst case scenarios (losses), CVaR is the average of the A% worst case scenarios. In other words CVaR *is the expected value of losses given that the losses are greater than VaR*. In [75] an estimator for this measure was defined as follows:

$$CVaR_n^\alpha(Z) = -\frac{\sum_{i=1}^w Z_{i:n}}{w},\tag{4.3}$$

where, $Z_{i:n}$ are the order statistics of the return on investment random variable, $w = \lfloor n\alpha \rfloor = max[m|m \le n\alpha, m \in N]$, n is the total number of scenarios, and $\alpha \in (0,1)$ is a probability value.

We use the equilibrium profit from the bottom tier supply function matrix game to calculate the risk constrained profit for generator i as follows.

$$\Pi_i = \Omega_i^* - \zeta_i(CVaR),\tag{4.4}$$

where Ω_i^* is the equilibrium profit for generator i from the supply function game. The term ζ_i denotes the risk preference of generator i varying between 0 and 1, and CVaR is as described in Equation 4.3.

4.1.2 Bottom Tier: Supply Function Game

The supply function game is denoted by the tuple: $< N, \tilde{A}^1, \ldots, \tilde{A}^n, \tilde{R}^1, \ldots, \tilde{R}^n >$. The elements of the tuple are as follows.

1. N denotes the number of generators.

2. $\tilde{A}^k$ denotes the set of supply function bid choices available to generator k.

3. $\tilde{R}^k : \tilde{A}^1 \times \ldots \times \tilde{A}^N \to R$ is the payoff function for generator k, where an element $\tilde{r}^k(b^1, \ldots, b^N)$ of $\tilde{R}^k$ is the profit of generator k when the generators choose supply function bids b^1 through b^N.

$\tilde{R}^k$ for all k, can be written in the form of N-dimensional matrices representing the supply function matrix game as follows

$$\tilde{R}^k = \left[\tilde{r}^k(b^1, b^2, \cdots, b^N)\right]_{\cdot b^1 = 1, \ldots, b^N = 1}^{|\tilde{A}^1|, \ldots, |\tilde{A}^N|} \tag{4.5}$$

The generators select bids from the set of available supply function bid choices with the goal of maximizing their payoffs which depend on all other generators' bids. The pure strategy Nash equilibrium for the supply function game is defined as that bid choice profile $\mathbf{b}^* = (b_*^1, \cdots, b_*^N)$, for which $\tilde{r}^k(b_*^k, b_*^{-k}) \geq \tilde{r}^k(b^k, b_*^{-k})$, $\forall b^k \in \tilde{A}^k$, and $k = 1, 2, \cdots, N$.

The generator profits $\tilde{r}^k(b^1, b^2, \cdots, b^N)$ constituting the supply function game are calculated as follows [76].

$$\tilde{r}^k(b^1, b^2, \cdots, b^N) = 1/2 \left[(p_i - x_i) + (p_i - (x_i + y_i \ q_i))\right] q_i, \tag{4.6}$$

where, p_i and q_i are the optimal price and quantity allocations for bid choices b^1, b^2, $\cdots$, b^N. These optimal price and quantity allocations are obtained by solving a linearized DC-OPF model, which is presented next. Note that Ω_i^* in Equation 4.4 is the equilibrium profit of the supply function game, obtained as

$$\Omega_i^* = 1/2 \left[(p_i^* - x_i) + (p_i^* - (x_i + y_i \ q_i^*))\right] q_i^*, \tag{4.7}$$

where x_i, y_i are cost function parameters, and p_i^*, q_i^* are the equilibrium price and quantity allocations of generator i.

4.1.2.1 Optimal Power Flow Model

The optimal power flow model used in this research is adopted from [76]. The independent system operator (ISO) receives supply and demand functions from the market participants and then solves a social welfare maximization problem. The OPF computes optimal price and quantity allocations at each bus of the network, while satisfying system security and transmission related constraints. The DC-OPF model used here is rather simple and allows for easier economic interpretations than nonlinear AC-OPF models.

We assume that generators submit linear supply functions to the ISO. The supply functions have the following general form:

$$p_i = x_i + y_i q_i, \quad \forall i \in \mathcal{G}, \tag{4.8}$$

where, $\mathcal{G}$ is the set of generators, p_i (\$/MWH) and q_i (MWH) are the price and quantity respectively, and x_i, y_i are the intercept and slope of the linear supply function.

We assume that consumers submit decreasing linear demand functions to the ISO. The demand functions have the following general form:

$$p_j = x_j - y_j d_j, \quad \forall j \in \mathcal{C}, \tag{4.9}$$

where, $\mathcal{C}$ is the set of consumers, p_j (\$/MWH) and d_j (MWH) are the price and quantity respectively, and x_j, y_j are the intercept and slope of the demand function.

As stated earlier, this DC-OPF formulation is adopted from Berry et al. [76]. $B_h[p_h]$ is the total benefit to the consumers and $C_h[P_h]$ is the total cost to the generators (see [76] for details). Q_h is the total amount of power supplied by all generators and D_h is the total amount of power demanded by all consumers, at bus h. R_{hk} is the reactance on the path from bus h to k, t_{hk} is the power flowing from bus h to k, $q_i[p_h]$ is the power supplied by supplier i at the price p_h, and $d_j[p_h]$ is the quantity of power demanded at price p_h. Assuming that supply and demand bids submitted by the generators and consumers are linear, this becomes an optimization problem with a quadratic objective function subject to linear constraints 4.11 - 4.16. Constraints 4.13 and 4.14 help to satisfy Kirchhoff's current and voltage laws respectively, while constraints 4.15 and 4.16 are used to satisfy transmission limits.

$$Max \ \ TW[P] = \sum_h B_h[p_h] - \sum_n C_h[p_h] \tag{4.10}$$

$$Subject \ to \ constraints :$$

$$Q_h - \sum_{i\in i(h)} q_i[p_h] = 0 \quad \forall \ \ nodes \ \ h \tag{4.11}$$

$$D_h - \sum_{j\in j(h)} d_j[p_h] = 0 \quad \forall \ \ nodes \ \ h \tag{4.12}$$

$$Q_h - D_h - \sum_{k\in k(h)} (t_{hk} - t_{kh}) = 0 \quad \forall \ \ nodes \ \ h \tag{4.13}$$

$$\sum_{hk\in A(v)} R_{hk}(t_{hk} - t_{kh}) = 0 \quad \forall \ \ voltage \ \ loops \ \ v \tag{4.14}$$

$$t_{hk} \leq T_{hk} \quad \forall \ \ arcs \ \ hk \tag{4.15}$$

$$t_{hk} \geq 0 \quad \forall \ \ arcs \ \ hk \tag{4.16}$$

The payoffs for each generator calculated from the solution of the OPF model are used to populate the N-dimensional payoff matrices for the supply function game. Then, the reinforcement learning algorithm, we develop in Chapter 5 (see [10]), is used to obtain the equilibrium bids and corresponding price and quantity allocations.

These allocations are used to compute the risk constrained profits, which, as explained before, form the N-dimensional payoff matrices of the investment game. Then, the reinforcement learning algorithm (Chapter 5) is used to obtain the equilibrium expansion plan for a given year for all generators. This process is repeated one year at a time to obtain the multi-year, multi-player, generation expansion strategy.

In the next chapter we show how the matrix games (such as those encountered in both tiers) are the fundamental building blocks of a much larger class of problems known as stochastic games. We then develop a value function approximation based learning algorithm to solve these matrix games. Later, the solutions obtained by the algorithm are benchmarked against those obtained by a commercial matrix game solver.

CHAPTER 5

REINFORCEMENT LEARNING BASED SOLUTION ALGORITHM FOR MULTIPLAYER MATRIX GAMES

5.1   Matrix Games

A matrix game can be defined by a tuple $< n, A^1, \ldots, A^n, \tilde{R}^1, \ldots, \tilde{R}^n >$. The elements of the tuple are as follows.

1. n denotes the number of players.

2. A^k denotes the set of actions available to player k.

3. $r^k : A^1 \times \ldots \times A^n \to R$ is the payoff function for player k, where an element $r^k(a^1, \ldots, a^n)$ is the payoff to player k when the players choose actions $\mathbf{a} = (a^1, \cdots, a^n)$.

$\tilde{R}^k$ for all k, can be written as an n-dimensional matrix as follows

$$\tilde{R}^k = \left[r^k(a^1, a^2, \cdots, a^n) \right]_{a^1=1,\ldots,a^n=1}^{a^1=|A^1|,\ldots,a^n=|A^n|} . \tag{5.1}$$

The players select actions from the set of available actions with the goal of maximizing their payoffs which depends on all the players' actions. The concept of *Nash equilibrium* is used to describe the strategy as being the most rational behavior by the players acting to maximize their payoffs. So for a matrix game, a pure strategy Nash equilibrium is an action profile $\mathbf{a}^* = (a_*^1, \cdots, a_*^n)$, for which $r^k(a_*^k, a_*^{-k}) \geq r^k(a^k, a_*^{-k})$, $\forall a^k \in A^k$, and $k = 1, 2, \cdots, n$. The equilibrium values denoted by $Val[\cdot]$ for player

k with payoff matrices $\tilde{R}^k$ is obtained as $Val[\tilde{R}^k] = r^k(a^1_*, \cdots, a^n_*)$. The appealing feature of the Nash equilibrium is that any unilateral deviation from it by any player is not worthwhile. A mixed strategy Nash equilibrium for matrix games is a vector $(\pi^1_*, \cdots, \pi^n_*)$, for which we can write

$$\sum_{a^1=1}^{|A^1|} \cdots \sum_{a^n=1}^{|A^n|} \pi^k_*(a^k)\pi^{-k}_*(a^{-k})r^k(a^k, a^{-k}) \geq \sum_{a^1=1}^{|A^1|} \cdots \sum_{a^n=1}^{|A^n|} \pi^k(a^k)\pi^{-k}_*(a^{-k})r^k(a^k, a^{-k}),$$

$$(5.2)$$

where $\pi^{-k}_*(a^{-k}) = \pi^1_*(a^1) \cdots \pi^{k-1}_*(a^{k-1}).\pi^{k+1}_*(a^{k+1}) \cdots \pi^n_*(a^n)$.

A matrix game may not have a pure strategy Nash equilibrium, but it always has a mixed strategy Nash equilibrium ([32]). There exist methods for solving Nash equilibrium of finite nonzero-sum matrix games ([37, 40, 43]). Since in matrix games, there are no transition probability functions, matrix games are static. Also matrix games can be viewed as recursive stochastic games with a single state. On the other hand, stochastic games can be viewed as extensions of matrix games from a single state to a multi-state environment.

A general sum stochastic game has equivalent matrix games. Therefore, once the equivalent matrix games are established, solution of a stochastic game reduces to solving the set of matrix games (one for each state). Hence, matrix games play a very critical role for solving this broad class of problems. The intent of the following section is to provide a brief overview of the main results from the recent literature concerning the existence of equivalent matrix games for both discounted reward (DR) and average reward (AR) stochastic games.

5.1.1 Equivalent Matrix Games for Discounted Reward Stochastic Games

A stochastic game can be defined by a tuple $< n, S, A^1, ..., A^n, P, \tilde{R}^1, ..., \tilde{R}^n >$, which differs from matrix games by having the following additional elements:

1. S: a finite set of states (s) of the environment, and

2. P: the set of transition probability matrices, where $p(s' \mid s, \mathbf{a})$ is the transition probability of reaching state s' as a result of a joint action $\mathbf{a}$ by all of the n players.

In a stochastic game, the transition probabilities and the reward functions depend on the choices made by all agents. Thus, from the perspective of an agent, the game environment is nonstationary during its evolution phase. However, for irreducible stochastic games, optimal strategies constitute stationary policies and hence it is sufficient to consider only the stationary strategies ([36]). We define $\pi^k(s)$ as the mixed strategy at state s for agent i, which is the probability distribution over available action set, $A^k(s)$, of player k. Thus $\pi^k(s) = \{\pi^k(s, a) : a \in A^k(s)\}$, where $\pi^k(s, a)$ denotes the probability of player k choosing action a in state s, and $\sum_{a \in A^k(s)} \pi^k(s, a) = 1$. Then $\pi = (\pi^1, ..., \pi^n)$ denotes a joint mixed strategy, also called a policy. A pure action $a \in A^k(s, a)$ can be treated as a mixed strategy π^k for which $\pi^k(a) = 1$. Let the cardinality of $A^k(s)$ be denoted by $m^k(s)$.

Under policy π, the transition probability can be given as

$$p(s' \mid s, \pi) = \sum_{\mathbf{a^1}=1}^{\mathbf{m^1(s)}} \cdots \sum_{\mathbf{a^n}=1}^{\mathbf{m^n(s)}} \mathbf{p(s' \mid s, a^1, ..., a^n)} \pi^\mathbf{n}(\mathbf{s, a^n}) \cdots \pi^\mathbf{1}(\mathbf{s, a^1}). \qquad (5.3)$$

The immediate expected reward of player k induced by a mixed strategy π in a state s is given by

$$r^k(s, \pi) = \sum_{\mathbf{a^1=1}}^{\mathbf{m^1(s)}} ... \sum_{\mathbf{a^n=1}}^{\mathbf{m^n(s)}} \mathbf{r^k(s, a^1, ..., a^n)}\pi^{\mathbf{n}}(\mathbf{s, a^n})...\pi^{\mathbf{1}}(\mathbf{s, a^1}). \tag{5.4}$$

Then the overall discounted value of a policy π to player k starting in state s can be given as

$$V_\beta^k(s, \pi) = \sum_{t=0}^{\infty} \beta^t E_s(r_t^k) = \sum_{t=0}^{\infty} \beta^t \sum_{s' \in S} p^t(s' \mid s, \pi) r^k(s', \pi), \tag{5.5}$$

where $p^t(.)$ denotes an element of the tth power of the transition probability matrix P.

The discounted reward given in (5.5) can be rewritten in component notation in terms of expected immediate reward and the expected discounted value of the next state as follows

$$V_\beta^k(s, \pi) = \mathbf{r^k(s, \pi)} + \beta \sum_{\mathbf{s' \in S}} \mathbf{p(s' \mid s, \pi)} \mathbf{V_\beta^k(s', \pi)}, \tag{5.6}$$

from which the definition of Nash equilibrium can be given as

$$r^k(s, \pi_*) + \beta \sum_{s' \in S} p(s' \mid s, \pi) V_\beta^k(s', \pi_*) \geq \quad r^k(s, \pi_*^{-k}, \pi^k) + \beta \sum_{s' \in S} p(s' \mid s, \pi_*^{-k}, \pi^k)$$

$$V_\beta^k(s', \pi_*^{-\mathbf{k}}, \pi^{\mathbf{k}}). \tag{5.7}$$

Directly solving for Nash equilibrium using the inequality (5.7) is difficult, even when the reward functions and transition probabilities are available. Filar and Vrieze [36] combined the theories of discounted Markov decision processes and Matrix games to develop an auxiliary bi-matrix game for two player discounted stochastic

games. The above technique is extended in [34] to n-player games for constructing n-dimensional equivalent auxiliary matrices $Q^k(.)$ for all players $k = 1, ..., n$.

The elements of the $Q^k(.)$ matrices are payoffs for all possible pure action sets $\mathbf{a}$, which take into account both the immediate reward and the future opportunities. For $s \in S$, the matrix with size $m^1(s) \times m^2(s) \times ... \times m^n(s)$ for the kth player is:

$$Q^k(s) = \left[r^k(s, a^1, ..., a^n) + \beta \sum_{s' \in S} p(s' \mid s, a^1, ..., a^n) V_\beta^k(s', \pi_*) \right]_{a^1=1,...,a^n=1}^{a^1=m^1(s),...,a^n=m^n(s)} \tag{5.8}$$

where $V_\beta^k(s', \pi_*)$ is the equilibrium value for the stochastic game starting at state s' for player k. Note that this auxiliary matrix, $Q^k(.)$ captures the information from the matrix game resulting from the pure strategies as well as the equilibrium payoff of the stochastic game. This enables the establishment of the connection between the matrix games and discounted reward stochastic games as given by the following result of [34].

In the following, Theorem 1, items 1 and 2 are equivalent.

1. π_* is an equilibrium point in the discounted reward stochastic game with equilibrium payoffs $(V_\beta^1(\pi_*), ..., V_\beta^n(\pi_*))$.

2. For each $s \in S$, the strategy $\pi_*(s)$ constitutes an equilibrium point in the static n-dimensional matrix game $(Q^1(s), ..., Q^n(s))$ with equilibrium payoffs $(Val[Q^1(s), \pi_*], \cdots, Val[Q^n(s), \pi_*])$. The entry of $Q^k(s)$ corresponding to actions $\mathbf{a} = (\mathbf{a^1}, \cdots, \mathbf{a^n})$ is given by $Q^k(s, \mathbf{a}) = \mathbf{r^i}(\mathbf{s}, \mathbf{a}) + \beta \sum_{s' \in \mathbf{S}} \mathbf{p}(s' \mid \mathbf{s}, \mathbf{a}) \mathbf{V_\beta^i}(\mathbf{s'}, \pi_*)$, for $i = 1, ..., n$, where $\mathbf{a} \in \prod_{i=1}^{n} \mathbf{A^i}(\mathbf{s})$.

We note that, the entries in this matrix game (5.8) have similar structure to the Bellman's optimality equation for discounted MDP. Well known algorithms to solve Bellman's discounted optimality equation are value iteration and policy iteration.

An extension of the value iteration and redefinition of the value operator to solve stochastic games was presented in [77]. There exist learning algorithms that attempt to learn the entries of the $Q^k(\cdot)$ matrices. The matrices are updated during each stage and are expected to converge to their optimal forms. Minmax Q-learning algorithm for discounted zero-sum games is presented in [78]. A Nash Q-learning for discounted general-sum games is presented in [79]. Both Minmax Q-learning and Nash-Q learning algorithms are extensions of the model-free reinforcement Q-learning [80, 71]. A summary of the available stochastic game algorithms can be found in [81]. One assumption that is inherent in the above literature is that once the equivalent matrices $Q^k(\cdot)$ are constructed, they can be solved using existing methods. However, the existing methods for obtaining NE value $(Val[Q^k(s), \pi_*])$ of n-player $(n > 2)$ matrix games are fraught with computational and convergence related challenges ([39, 40]). Development of a computationally viable method of finding the NE value of a matrix game $(Val(Q_t^k(s), (\pi_t)_*))$ is still an open challenge and is addressed in this research.

5.1.2 Equivalent Matrix Games for Average Reward Stochastic Games

Let $V_\alpha^k(\pi_*)$ denote the *gain equilibrium value*, and $h^k(\pi_*)$ denote the *bias equilibrium value* of an average reward stochastic game. The above equilibrium values can be defined as

$$V_\alpha^k(s, \pi^*) = \limsup_{\mathbf{T} \to \infty} \frac{\mathbf{1}}{\mathbf{T}} \sum_{\mathbf{t=0}}^{\mathbf{T-1}} \mathbf{p^t}(\mathbf{s'} \mid \mathbf{s}, \pi^*) \mathbf{r^k}(\mathbf{s'}, \pi^*). \tag{5.9}$$

and

$$h^k(s, \pi^*) = \lim_{T \to \infty} E_s \sum_{t=0}^{T-1} [r_t^k - g^k(\pi^*)], \tag{5.10}$$

where $g^k(\pi^*)$ is long-run expected *average-reward*, which can be given by

$$g^k(\pi^*) = \limsup_{T \to \infty} E(\frac{1}{T} \sum_{t=0}^{T-1} r_t^k). \tag{5.11}$$

Similar to the discounted games, for n-player average reward games, it is shown in [33] that n-dimensional equivalent auxiliary matrices $R^k(.)$ for all players $k = 1, ..., n$ can be constructed. The elements of these matrices are payoffs for all possible pure action sets $\mathbf{a}$, which take into account both the immediate reward and the future opportunities. For $s \in S$ the matrix with size $m^1(s) \times m^2(s) \times .. \times m^n(s)$ for the kth player can be given by

$$
R^k(s) = \left[r^k(s, a^1, ..., a^n) - V_\alpha^k(\pi_*) + \sum_{s' \in S} p(s' \mid s, a^1, ..., a^n) h^k(s', \pi_*) \right]_{a^1=1,...,a^n=1}^{a^1=m^1(s),...,a^n=m^n(s)}
$$
$$(5.12)$$

The following theorem establishes the connection between average reward irreducible stochastic games and the average reward matrix games ([33]).

In the following, Theorem 2, items 1 and 2 are equivalent.

1. π_* is an equilibrium point in the average reward irreducible stochastic game with bias equilibrium value $h^k(\pi_*)$ and gain equilibrium value $V_\alpha^k(\pi_*)$ for $k = 1, 2, \cdots, n$.

2. For each fixed $s \in S$, the strategy set $\pi_*(s)$ constitutes an equilibrium point in the static n-dimensional equivalent matrix game $(R^1(s), \cdots, R^n(s))$ with bias equilibrium value $h^k(s, \pi_*)$ and gain equilibrium value $Val[R^k(s), \pi_*]$ for $k = 1, \cdots, n$.

So far, we have defined matrix games and presented a summary of the available results from [34] and [33]. These results show that for both discounted and average reward stochastic games, there exist equivalent matrix games, the solutions of which provide the equilibrium strategies and values. Clearly, computationally feasible solution methodologies for matrix games play a fundamental role in solving a large

class of stochastic games. In what follows, we present a new algorithm that uses a reinforcement learning approach to solve *matrix games.*

5.2 Finding NE of Matrix Games

In this section we present a new approach to obtain Nash equilibrium of n-player matrix games. Let $R^k(\mathbf{a})$ denote the reward matrix of the kth player of which $r^k(a^1, \cdots, a^n)$ are the matrix elements. Define the value of an action a^k to player k as

$$Val[R^k(a^k)] = \sum_{\{a^1, \cdots, a^n \backslash a^k\}} p(a^{-k}, a^k) r^k(a^1, \cdots, a^k, \cdots, a^n), \qquad (5.13)$$

where $p(a^{-k}, a^k)$ denotes the probability of choice of an action combination a^{-k} by all the players while player k chose action a^k. In decision making problems with a single player (MDPs and SMDPs), there exist optimal *values* for each state-action pair, which determine the optimal action in each state ([68]). Drawing an analogy, for matrix games that have multiple players and a single state, we conjecture that there exist optimal *values* for all actions of the players that can yield pure and mixed NE strategies. However, the probabilities $(p(a^{-k}, a^k))$ needed to compute these *values* are impossible to obtain for real life problems without prior knowledge of players' behavior. Therefore, we employ a learning approach to estimate the *values* of the actions as follows. We rewrite (5.13) as

$$Val[R^k_{t+1}(a^k)] = (1 - \gamma_t)[R^k_t(a^k)] + \gamma_t \left[r^k(a^1, \cdots, a^k, \cdots, a^n) \right]. \qquad (5.14)$$

The algorithm presented below utilizes the value learning scheme (5.14) to derive pure and mixed NE strategies for n-player matrix games.

5.2.1 A Value Iteration Algorithm for n-Player Matrix Games

We assume that the game has n-players and each player k has a set of A_k action choices. Hence, n different reward matrices of size $|A_1| \times |A_2| \times \cdots \times |A_n|$ are available.

1. Eliminate rows and columns of the matrices associated with the dominated strategies. A dominated strategy is one that will never be adopted by a rational player irrespective of the choices of other players. A strategy $a \in A_k$ for player k is said to be dominated if $r(k, a, a^{-k}) \leq r(k, \bar{a}, a^{-k})$, where $\bar{a} \in A_k \backslash a$ and a^{-k} denotes the actions of all other players.

2. Let iteration count $t = 0$. Initialize the R-values for all player and action combinations $R(k, a)$ to an identical small positive value (say, 0.001). Also initialize the learning parameter γ_0, exploration parameter ϕ_0, and parameters γ_τ, ϕ_τ needed to obtain suitable decay rates of learning and exploration. Let *Maxsteps* denote the maximum iteration count.

3. If $t \leq$ *Maxsteps*, continue learning of the R-values through the following steps.

   (a) Greedy action selection for pure strategy Nash equilibrium: Each player k, with probability $(1 - \phi_t)$, chooses a greedy action for which $R^k(a) \geq R(k, \bar{a})$. A tie is broken arbitrarily. With probability ϕ_t, the player chooses an exploratory action from the remaining elements of A_k (excluding the greedy action), where each exploratory action is chosen with equal probability. Probabilistic action selection for mixed strategy Nash equilibrium: Compute the probabilities for the action choices using the ratio of R-values at iteration t as follows. For each player k, the probability of choosing the action $a \in A_k$ is given by $\frac{R(k,a)}{\sum_{b \in A_k} R(k,b)}$.

(b) *R*-Value Updating: Update the specific *R*-values for each player k corresponding to the chosen action a using the learning scheme given below.

$$R_{t+1}(k, a) \leftarrow (1 - \gamma_t)R_t(k, a) + \gamma_t \left(r(k, \mathbf{a}) \right), \qquad (5.15)$$

where $\mathbf{a}$ denotes the action combination chosen by players.

(c) Set $t \leftarrow t + 1$.

(d) Update the learning parameters γ_t and exploration parameter ϕ_t following the decay scheme given by Darken et al. in ([82]):

$$\Theta_t = \left(\frac{\Theta_0}{1 + u} \right), \quad \text{where} \quad u = \left(\frac{t^2}{\Theta_\tau + t} \right), \qquad (5.16)$$

where Θ_0 denotes the initial value of a learning/exploration rate, and Θ_τ is a large value (e.g., 10^6) chosen to obtain a suitable decay rate for the learning/exploration parameters. Exploration rate generally has a large starting value (e.g., 0.8) and a quicker decay, whereas learning rate has a small starting value (e.g., 0.01) and very slow decay rate. Exact choice of these values depends on the application ([11, 12]).

(e) If $t <$ MaxSteps, go to Step 3(a), else go to Step 4.

4. Equilibrium Strategy Determination: For each player k, the pure strategy is action a for which $R(k, a) \geq \max_{b \in A_k} R(k, b)$. The pure strategies for all players combined constitute the pure strategy equilibrium. For each player k, the mixed strategy is to select each action $a \in A_k$ with probability $\frac{R(k,a)}{\sum_{b \in A_k} R(k,b)}$.

CHAPTER 6

EMPIRICAL ANALYSIS AND PRACTICAL APPLICATION

6.1 Numerical Evaluation of the Learning Algorithm

In this chapter we present results from an extensive comparative numerical study conducted with an objective of establishing the ability of the RL algorithm to obtain Nash equilibrium for n-player matrix games. For this purpose, sixteen matrix game examples with known Nash equilibria were solved by using both an openly available software (GAMBIT) and the RL algorithm. To demonstrate the practical applicability of the RL algorithm, we also solved a matrix game that models strategic bidding in a restructured electric power market.

6.1.1 Matrix Games with Known Equilibria

Matrix games that were studied consisted of up to four players and sixty four different action choices. Ten out of these sixteen examples have pure strategy Nash equilibria, which were solved using the variant of the RL algorithm that seeks a pure strategy. The remaining six games were solved using the mixed strategy version of the RL algorithm.

Table 6.1 summarizes the matrix games specifying the number of players and their available action choices. Some of these problems are adopted from GAMBIT library of matrix games, for which the file names used in GAMBIT are used as identifiers. The Nash equilibrium solutions obtained by both GAMBIT and RL algorithm are

Table 6.1. Sample Matrix Games with Pure Strategy Nash Equilibria

Matrix Games Studied for Pure Strategy Nash Equilibrium	
Game 1 		1	2
1	3, 3	0, 5
2	5, 0	1, 1	 (2 Players: 2 x 2)	**Game 4: GAMBIT © wink3** (2 Players: 3 x 3) **Game 5: GAMBIT © perfectl** (2 Players: 3 x 3)
Game 2 		1	2
1	2, 2	0, 3
2	3, 0	1, 1	 (2 Players: 2 x 2)	**Game 6: GAMBIT © 2x2x2** (3 Players: 2 x 2 x 2) **Game 7: GAMBIT © 8x2x2** (3 Players: 8 x 2 x 2) **Game 8: GAMBIT © Palf** (2 Players: 3 x 3)
Game 3 		1	2	3
1	1, 4	2, 2	2, 3
2	3, 1	1, 5	4, 1
3	2, 0	3, 4	1, 2 | (2 Players: 3 x 3)	**Game 9: GAMBIT © 2x2x2x2** (4 Players: 2 x 2 x 2 x 2) **Game 10: GAMBIT © 8x8** (2 Players: 8 x 8)

Table 6.2. Pure Strategy Nash Equilibrium Results

Matrix Game #	GAMBIT solution		Learning Algorithm Solution		
	Equilibria	Equilibrium Value	Equilibrium	Equilibrium Value	Convergence time in Seconds
Game 1	(2, 2)	(1, 1)	(2, 2)	(1, 1)	10
Game 2	(2, 2)	(1, 1)	(2, 2)	(1, 1)	14
Game 3	(3, 2)	(3, 4)	(3, 2)	(3, 4)	27
Game 4	**(1, 2)** (2, 1)	**(3, 4)** (1, 2)	(1, 2)	(3, 4)	27
Game 5	(1, 1) **(2, 2)** (3, 3)	(3, 1) **(2, 2)** (3, 1)	(2, 2)	(2, 2)	40
Game 6	**(1, 1, 1)** (2, 2, 1) (1, 2, 2) (2, 1, 2)	**(9, 12, 8)** (9, 2, 8) (3, 6, 4) (3, 6, 4)	(1, 1, 1)	(9, 12, 8)	42
Game 7	**(1, 2, 1)** (2, 1, 7)	**(7.0, 4.6, 7.5)** (5.7, 4.5, 7.5)	(2, 1, 7)	(5.7, 4.5, 7.5)	11
Game 8	(3, 3)	(3, 2)	(3, 3)	(3, 2)	10
Game 9	(1, 1, 2, 1) **(2, 2, 1, 1)**	(2.3, 3.2, 3.0, 2.4) **(4.5, 4.9, 5.7, 4.5)**	(2, 2, 1, 1)	(4.5, 4.9, 5.7, 4.5)	39
Game 10	**(4, 6)** (6, 3) (7, 2)	**(7.5, 7.9)** (4.9, 5.7) (5.6, 5.6)	(4, 6)	(7.5, 7.9)	12

48

summarized in Table 6.2. The following observations can be made from the results. For all ten games, the RL algorithm found a Nash equilibrium which coincided with a GAMBIT solution. It may be noted that GAMBIT obtained multiple pure strategy NE for six out of the ten games. For each of these games (except in Game #7), RL algorithm chose the equilibrium with the highest player rewards. Though a formal mathematical proof will be required to support this observation, we believe that, since the RL algorithm learns the values for the actions and chooses actions based on these values, the solution tends to converge to the NE with the highest player rewards.

Table 6.2 also presents the convergence time of the RL algorithm which was run for 10,000 iterations for all the games on a computer with a 1.6 GHz Pentium M processor. However, an accurate assessment of the convergence time will require further optimization of the learning parameters of the algorithm, which could be problem dependent. For example, many of the games that are presented in the table converged much sooner than 10,000 iterations. Hence, the convergence times presented here are intended only to provide a general idea of the computational efforts required by the algorithm.

Table 6.3 presents the comparison of mixed strategies obtained by GAMBIT and the RL algorithm for six matrix games. Though GAMBIT found multiple mixed NE for most of these problems, for fairness of comparison, only those NE with maximum player rewards obtained by GAMBIT are presented in the table. As evident from the table, though the mixed strategies obtained by the RL algorithm are different from the NE obtained by GAMBIT, player rewards from the RL algorithm in almost all of the games are comparable. It can also be seen from the table that even when the mixed strategy version of the RL algorithm is implemented, it yields a pure strategy (if one exists, as in Games 4 and 5). It may be noted that for Games 4 and 5, GAMBIT also finds the pure strategies. However, in this table we present only mixed

Table 6.3. Mixed Strategy Equilibrium Results

	Mixed Strategy Equilibrium						
	Gambit Solution				**Learning Approach**		
Game 1		0.33	0.33	0.33		0.55	0	0.45
	0.33	0, 0	5, 4	4, 5	0	0, 0	5, 4	4, 5
	0.33	4, 5	0, 0	5, 4	1	4, 5	0, 0	5, 4
	0.33	5, 4	4, 5	0, 0	0	5, 4	4, 5	0, 0
	Value to player 1= 2.94				Value to player 1= 4.45			
	Value to player 2= 2.94				Value to player 2= 4.55			
Game 2		0	1	0		0.5	0.48	0.2
	0.5	1, 3	3, 4	4, 2	0.5	1, 3	3, 4	4, 2
	0.5	1, 2	3, 1	2, 1	0.5	1, 2	3, 1	2, 1
	0	0, 4	2, 3	3, 3	0	0, 4	2, 3	3, 3
	Value to player 1= 3				Value to player 1= 2.54			
	Value to player 2= 2.5				Value to player 2= 2.75			
Game 3		0.57	0.43	0.5		0.5	0	0.5
	0.5	3, 1	0, 0	0, 1	0	3, 1	0, 0	0, 1
	0.5	1.5, 1	2, 2	1.5, 1	0.5	1.5, 1	2, 2	1.5, 1
	0	0, 1	0, 0	3, 1	0.5	0, 1	0, 0	3, 1
	Value to player 1= 1.7				Value to player 1= 1.5			
	Value to player 2= 1				Value to player 2= 1			
	In the Games 4 and 5 Mixed Strategy Learning Algorithm obtains a pure strategy solution							
Game 4		0.82	0.18			0	1	
	0.67	0, 0	9, 3		1	0, 0	9, 3	
	0.33	2, 6	0, 0		0	2, 6	0, 0	
	Value to player 1= 1.64				Value to player 1= 9			
	Value to player 2= 2				Value to player 2= 3			
Game 5		0.833	0.167			1	0	
	0.833	70, 70	10, 60		1	70, 70	10, 60	
	0.167	60, 10	60, 60		0	60, 10	60, 60	
	Value to player 1= 60				Value to player 1= 70			
	Value to player 2= 60				Value to player 2= 70			
Game 6	GAMBIT © bayes2a				GAMBIT © bayes2a			
	2 Players 64 actions each				2 Players 64 actions each			
	Value to player 1: 8.9				Value to player 1: 8.3			
	Value to player 2: 6.6				Value to Player 2: 8.1			

strategy results obtained by both GAMBIT and the RL algorithm. In Game 6, where the two players have 64 actions each, the mixed strategies for both players have large support sets and thus could not be presented in the table. Therefore, we chose to present only the player rewards as means for comparison. In the next subsection, we

50

present a matrix game example from a real life marketplace that is settled through multiparty competition on a periodic basis.

6.1.2 A Power Market Matrix Game

In restructured electric power markets, like in PJM (Pennsylvania- Jersey- Maryland), New York, New England, and Texas, power is traded in long term bilateral market, day ahead market, and spot market. The generators and retailers compete in the market by strategically bidding for price and quantity of power traded in order to maximize profits. The market is settled by an independent system operator, who matches the supply and demand and satisfies the network constraints while maximizing social welfare (total benefit minus total cost). This settlement yields price and quantity allocations at all the network nodes. The generators strategize to raise their prices above the marginal (base) costs, while the retailers' strategies are aimed at maintaining prices close to the marginal costs. The ability of the generators to maintain prices above the marginal costs for a sustained period of time is defined as *market power*. A market is said to be competitive when the prices are at or near the marginal costs, which is one of the primary objectives of a restructured electricity market design. A day ahead power market can be modeled as a repeated n-player matrix game, of which the reward matrices can be constructed using the producer surplus (for generators) and consumer surplus (for retailers).

We consider a four bus (two generators and two retailers) power network as shown in Figure 6.1, which was studied in ([76]). The supply function bids of the generators at nodes A and B and the demand functions of the retailers at nodes C and D are as follows: $p_{S_1} = a_1 + m_1 q_1$, $p_{S_2} = a_2 + m_2 q_2$, $p_{D_1} = 100 - 0.52 d_1$, $p_{D_2} = 100 - 0.65 d_2$, where q_1 and q_2 are the quantities (in megawatt-hour, MWh) produced by generators S_1 and S_2 respectively, and d_1 and d_2 are the quantities demanded by the retailers D_1

and D_2 respectively. The supply function has two strategic bid parameters (intercept a in \$/MWh and slope m) that the generators manipulate to maximize their profits. Demand side bidding by the retailers is not considered and hence the demand function parameters are maintained constant at their base values. As in ([76]), the reactances are considered to be the same on all lines.
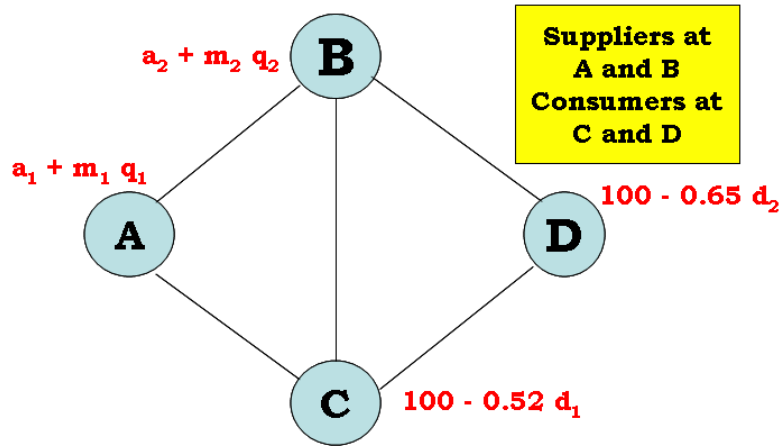
Figure 6.1. 4-Bus Power Network

In [76], the effects of strategic bidding are studied by imposing transmission constraints on lines AC and BD (one at a time) resulting in network congestion. Nash equilibria for both slope-only and intercept-only bidding scenarios for each of the transmission constrained cases (AC and BD) are separately examined.

Berry et al. [76] used an iterative algorithm to obtain NE of the above game. The algorithm involves solving the ISO's problem for a series of bid options of a generator, while holding the bids of the other generator constant. The bid option that produces maximum profit is then fixed, and the same procedure is repeated for the other generator. This process is repeated until neither generator has an alternative bid to further improve profit. The matrix game approach developed in this research differs

from the above approach in that all generators select actions simultaneously without any knowledge of the others actions.

In order to apply the learning algorithm, as a first step, the reward matrices for the generators are constructed. To accomplish this, the feasible range of the bid parameters are suitably discretized (which dictate the size of the reward matrices), and the rewards for each combination of the generators bids are calculated. It may be noted that generator reward is a function of the nodal prices and quantities, which are obtained by solving a social welfare maximization problem. Details of the mathematical formulation can be found in ([76]). The feasible ranges of slope and intercept parameters are discretized to 250 values giving matrix sizes of 250×250. In particular, the slope parameter ranged from 0.35 to 2.85 for S_1 and 0.45 to 2.95 for S_2, both in steps of 0.01. The intercept bid parameter for both generators S_1 and S_2 ranged from 10 $/MWh to 260 $/MWh with a step length of 1 unit. The solution of the social welfare problem and calculation of the generator rewards for all the above bid combinations are accomplished using GAMS software. The results from ([76]) and those from the learning algorithm are presented in Table 6.4. It can be seen from the table that the learning algorithm obtains better or comparable profits for both generators in all cases.

We also extend the numerical experimentation by allowing generators to bid for both slope and intercept together, instead of bidding for one parameter at a time as in ([76]). The bid parameters in this experiment are discretized as follows. The slope is varied in twenty five steps of 0.1 for both generators ranging from 0.35 to 2.85 for S_1 and 0.45 to 2.95 for S_2. The intercept is varied in twenty five steps of 3 ranging from 10 $/MWh to 85 $/MWh. Hence, each generator has $25 \times 25 = 625$ action choices and the resulting reward matrices are of size 625×625. The RL algorithm is run for 500,000 iterations, which took 770 seconds on a computer with a 2 GHz

Pentium IV processor. As shown in Table 6.4, in the AC-congestion case, bidding in both slopes and intercepts lead to similar profits as in the cases of one parameter at a time bidding. Whereas, in the case of BD-congestion, the profits obtained by the players through joint bidding is much higher than bidding one parameter at a time.

Table 6.4. Results from the Study of 4-Bus Power Network

	Bidding with Slopes while Intercepts held constant for both players at 10, 10		Biding with Intercepts while Slopes held constant for both players at 0.35, 0.45		
	Berry et al.	Learning Algorithm	Berry et al.	Learning Algorithm	
Profit Gen A ($)	1343	1457	1087	1136	Line AC
Profit Gen B ($)	3345	3284	3373	3338	constrained at
NE Bids	(2.42, 0.72)	(2.26, 0.78)	(60, 25)	(59, 26)	30 MW
Profit Gen A ($)	2478	2640	2115	2156	Line BD
Profit Gen B ($)	1374	1422	972	1044	constrained at
NE Bids	(0.77, 1.2)	(0.8, 1.36)	(22, 28)	(28, 29)	30 MW
	Bidding with both Slopes and Intercepts NE obtained using Learning Algorithm				
	Line AC constrained at 30 MW		Line BD constrained at 30 MW		
Profit Gen A ($)	1518		2851		
Profit Gen B ($)	3266		1680		
NE Bids	Gen A: (61, 0.35), Gen B: (37, 0.45)		Gen A: (27, 0.69), Gen B: (34, 0.93)		

6.2   Some Remarks

Though the internet era has provided the technological infrastructure necessary to invigorate market competition, lack of commensurate advancements in computational algorithms to solve multiplayer games has been a limiting factor in examining the market behavior. Meteoric rise in computing power via tera and peta scale computing (made possible by efficient harnessing of cluster computing) has created an opportunity to break through perceived computational barriers of state space explo-

sion. This research presents a new computational approach to find Nash equilibrium of multiplayer matrix games. The approach is founded on the value function learning strategy that is being successfully used in solving large scale decision making problems modeled as Markov and semi-Markov decision processes. In the wake of recent studies that link a large class of stochastic games to matrix games ([34, 33]), our solution approach stands to impact a broad range of decision making problems.

The comparative numerical results presented for a large number of matrix games help to demonstrate the validity of our conjecture (in Chapter 5, Section 5.2) on value function guided NE determination. Though one might think that games generally involve a larger number of players than what is considered in the example problems, in real life, applications of matrix games tend to have a limited number of players. This oligopolistic structure of most contemporary markets naturally occurs due to extensive market segmentation. Some examples of such oligopolistic markets include retail sales, home and auto insurance, mortgage lending, service industries like airlines, hotels, and entertainments.

In Chapter 5, we developed a solution algorithm to solve multi-player matrix games and in the current chapter we benchmarked the solutions obtained from the RL algorithm with those obtained from GAMBIT. In the following chapter, we present a detailed solution framework for the overall two-tier model to obtain multiyear, multiplayer GEP strategies. The algorithm utilizes the RL based solution algorithm, developed here, to solve the matrix games embedded within the two-tier GEP model.

CHAPTER 7

SOLUTION FRAMEWORK FOR TWO-TIER GEP MODEL

7.1 Solution Algorithm for the Two-Tier GEP Model

The following step by step algorithm is used to solve the two-tier matrix game model for generation expansion planning. A schematic representation of the algorithm is presented in Figure 7.1.

1. At the start of every year, potential investors (generators) assess the future demand projections, profits from previous years, network conditions, and market design incentives to develop a set of feasible generation expansion investment alternatives (Box 1).

2. Let $a^i : i = 1, \cdots, N$ denote the number of investment alternatives available to generator i. Then, the investment matrix game A is an N-dimensional matrix of size $a^1 \times a^2 \times \cdots \times a^N$ (Box 2).

3. For each element of matrix game A, there is a corresponding supply function (SF) matrix game of size $\prod_{i=1}^{N} b^i$, where b^i denotes the number of supply function bids of generator i (Box 3).

4. Profits for each element of the SF games $(\tilde{r}^k(b^1, b^2, \cdots, b^N))$ are obtained after solving the corresponding DC-OPF (Box 4). See Equation 4.6.

5. Once the profits for each element of the SF games are obtained, a value approximation based reinforcement learning algorithm (Chapter 5) is used to find the

equilibrium profits (Ω_i^*) for the generators (Box 5). See Equation 4.7 for the formula used to compute Ω_i^*.

6. Subsequently, these equilibrium profits and respective equilibrium bids are utilized to compute the risk constrained profits (RCP, Π_i) via a conditional value-at-risk measure (Box 6). See Table 7.1 for the steps involved in computing the CVaR and see Equation 4.4 for the formula to compute Π_i.

7. These risk constrained profit values constitute the payoff matrices for investment matrix game A. Finally, the reinforcement learning algorithm, developed in (Chapter 5), is used on matrix game A to obtain the equilibrium solution. This solution is the risk constrained generation expansion strategy for the year under consideration (Box 7).

8. This procedure (Steps 1-7) is repeated one year at a time, until the generation expansion strategy for the entire planning horizon is obtained for each generator.
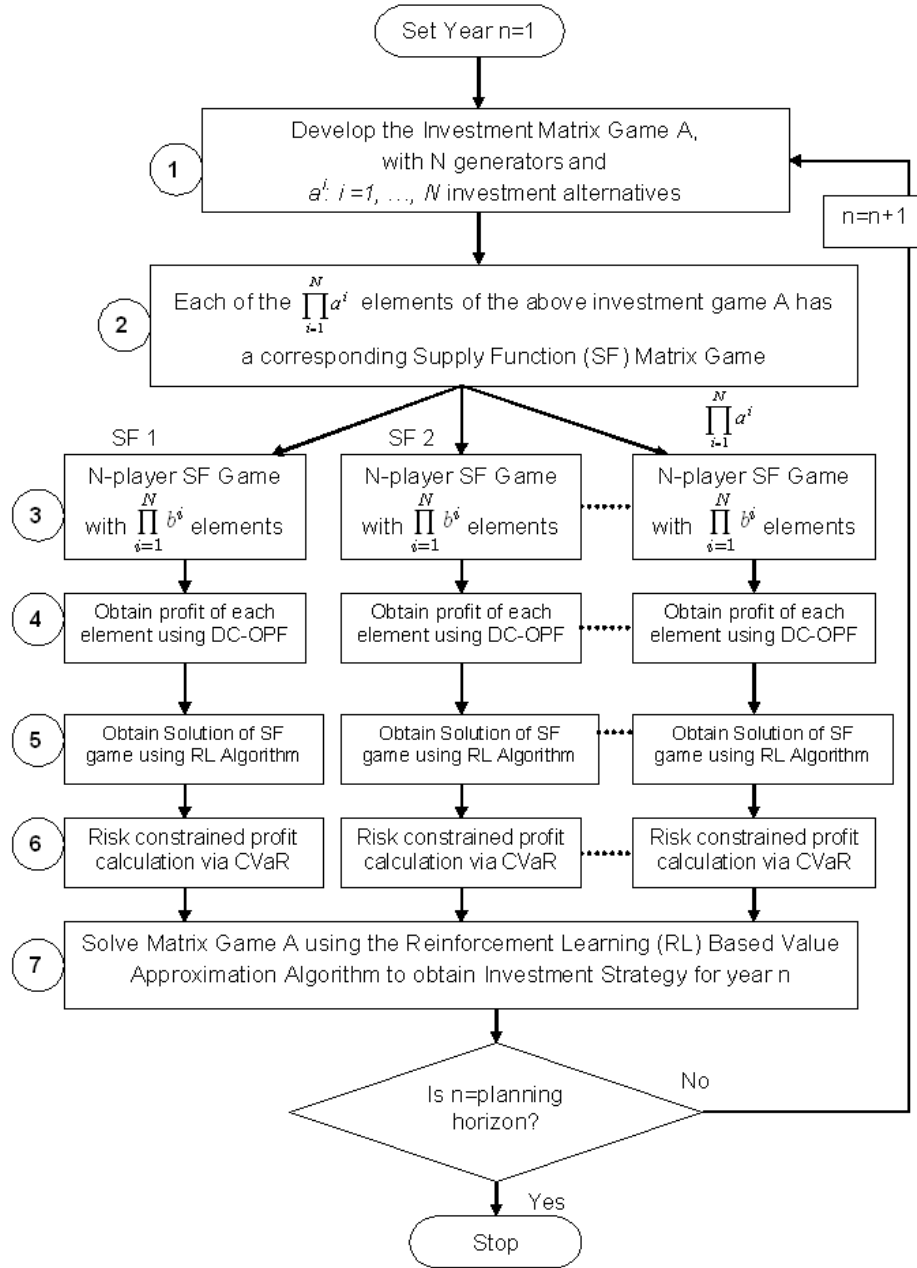
Figure 7.1. Schematic for Two-Tier GEP Model Solution Algorithm

Table 7.1. Steps for Calculating CVaR

Step 1	Define contingency scenarios $j = 1, \cdots, n$ where $n \in N$ is the total number of scenarios.
Step 2	Use the equilibrium price-quantity bids from every SF game and solve the DC-OPF problem for all contingency scenarios.
Step 3	Obtain profits β_j^i for each contingency scenario j and generator i.
Step 4	$\hat{\beta}_j^i$ is computed as the difference between the equilibrium profits of the SF game and the profits (β_j^i) from contingency scenarios. For example, if the equilibrium profit from the SF game is \$100 and the profit made due to a contingency scenario is \$70, the $\hat{\beta}_j^i$ is -\$30.
Step 5	Using $\hat{\beta}_j^i$ values from all the scenarios, compute the order statistics $\hat{\beta}_{(1:n)}^i, \leq, \cdots, \leq, \hat{\beta}_{(n:n)}^i; \forall i$.
Step 6	For a pre-defined value of $\alpha \in (0,1)$, calculate $CVaR_n^{(\alpha)}$, using Equation 4.3.
Step 7	Depending upon the risk preference of each generator ζ_i, compute the risk constrained profit using Equation 4.4.
Step 8	Repeat steps 1-7 for all $\prod_{i=1}^{N} a^i$ SF games.

CHAPTER 8

NUMERICAL ANALYSIS

8.1 Numerical Experimentation and Analysis

To demonstrate how the two-tier model works, we chose a 5-bus network from Powerworld software package [83]. The 5-bus network shown in Figure 8.1, consists of four loads and seven transmission lines and is currently served by three generators. There are two main reasons for choosing this particular network. First, we believe that the number of buses and generators are adequate to demonstrate the applicability of the two-tier model. Second, the relatively small size of the network allows for detailed numerical experimentation and analysis. We next present the important features of the sample problem.

Table 8.1 shows demand curve parameters for a four year planning horizon. The intercept parameter (in column 2) is considered the same for all four years, while the slope parameter is reduced each year indicating a growth in demand. Consistent with industry standards and energy literature, we make appropriate assumptions about marginal supply functions for coal, natural gas, nuclear, and petroleum fired plants, as shown in Table 8.2.

To keep the problem exposition simple, we assume that Generator 1 (located at Bus 1) and Generator 2 (located at Bus 4) compete against each other, while Generator 3 (located at Bus 2) acts as the price-taker. In other words, Generators 1 and 2 submit strategic bids aimed at maximizing individual profits, while Generator

Table 8.1. 4-year Demand Projections

Bus #	Demand Intercept All Years	Demand slope Year 1	Demand slope Year 2	Demand slope Year 3	Demand slope Year 4
1	0	0	0	0	0
2	100	-2	-1.5	-1	-0.5
3	100	-2	-1.5	-1	-0.5
4	100	-2	-1.5	-1	-0.5
5	100	-2	-1.5	-1	-0.5

Table 8.2. Supply Function Parameters of Generators

Gens	Plant Type	Bus #	Linear Supply Function Parameters	
			Base Slope	Base Intercept
Gen 1	Natural Gas Plant	1	0.75	1
	Coal Plant	3	0.3	3
Gen 2	Natural Gas Plant	4	0.5	2
	Fossil Fuel Plant	5	0.75	5
Gen 3	Nuclear Plant	2	0.25	10

3 simply accepts the price set by the market. In order to maintain the dimensionality of the sample network, we assume that generators do not bid strategically for newly built plants, i.e., new generating plants, if and when added at Buses 3 and 5, act as price-takers. While bidding, generators 1 and 2 hold their intercepts constant at their base values (as shown in Table 8.2) and bid strategically only with respect to slopes. We allow generators to bid in increments of 0.1 from their base values up to 10 steps, which means that the supply function matrix game has a size of 10×10. It may be noted that the reinforcement learning based solution algorithm is capable of handling much larger games (see [10]). That is, we could allow generators to bid strategically in both slope and intercept. Limiting the supply function game to 10×10 is done only for the ease of exposition.
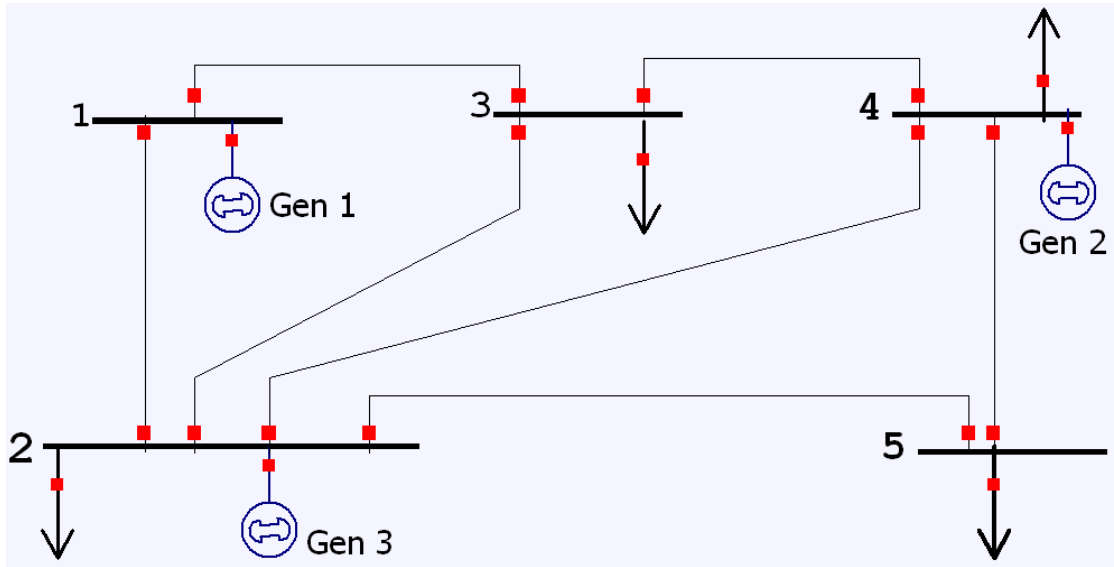
Figure 8.1. Five Bus Electric Power Network

Table 8.3. Line Failure Scenarios used in Computing CVaR

Scenarios	Line Failures	Scenarios	Line Failures
1	1-2	6	5-2
2	4-5	7	2-4
3	3-2	8	1-2 & 4-5
4	1-3	9	3-4 & 1-2
5	3-4	10	5-2 & 3-4

To demonstrate computation of CVaR, we subjected the sample network to signifi-cant variability through hundred different demand and line failure scenarios. Columns 2 and 4 of Table 8.3 show the lines that are assumed to fail. The ten demand vari-ability scenarios are shown in Table 8.4. Using the procedure shown in Table 7.1, we compute the CVaR and subsequently use it to obtain the risk constrained profit.

Table 8.4. Demand Variations used in Computing CVaR

Bus #	Demand Function		Demand Function		Demand Function		Demand Function		Demand Function	
	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
	Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5	
1	0	0	0	0	0	0	0	0	0	0
2	100	-2	50	-0.5	50	-0.5	200	-2.5	150	-0.5
3	100	-2	150	-3	100	-2	100	-2.5	100	-3.5
4	100	-0.75	100	-0.75	100	-0.75	100	-0.75	150	-0.75
5	50	-3	50	-3	100	-0.75	50	-3	50	-3.5
	Scenario 6		Scenario 7		Scenario 8		Scenario 9		Scenario 10	
1	0	0	0	0	0	0	0	0	0	0
2	200	-1.5	10	-2.75	100	-2.25	75	-1.25	125	-4.25
3	150	-2	150	-2.75	10	-2.25	100	-2.25	150	-3.25
4	200	-1.75	200	-0.75	10	-0.75	125	-0.75	175	-1.75
5	50	-4.5	50	-3	250	-0.5	50	-3.25	50	-1.5

8.1.1 Computational Results

The objectives of the numerical experimentation are three-fold: to demonstrate the ability of the model to obtain multi-year, multi-player, generation expansion plans, to present a statistical analysis of the impact of demand variations and transmission constraints on the nodal prices in the network before and after expansion, and to demonstrate how risk preferences of generators affect the choice of expansion plans.

Table 8.5 presents results of the GEP model for a four year planning horizon for the demand projection scenario depicted in Table 8.1. Generators are assumed to be highly risk prone ($\zeta = 0.1$) during the four year period. As presented in the Table 8.5 (Year 1, top segment), the current plants in the network are: a 50 MW natural gas plant at Bus 1 owned by Generator 1, a 50 MW natural gas plant at Bus 4 owned by Generator 2, and a 100 MW nuclear plant at Bus 2 owned by Generator 3. As shown in the table, Generator 1 has the following investment alternatives: do nothing (or post pone expansion), expand capacity of the existing natural gas plant at Bus 1 from 50 MW to 100 MW, or build a 50 MW coal plant at Bus 3. Similarly, Generator

Table 8.5. GEP Decisions for Demands from Table 8.1 ($\zeta = 0.1$)

	Year 1					
Gens	Current Plant Type	Bus #	Current Capacity (MW)	Proposed Investment Alternatives	Proposed Capacity (MW)	Proposed Bus Location
Gen 1	Natural Gas	1	50	Do Nothing		
				Expand Natural Gas	100	1
				Build Coal Plant	**50**	**3**
Gen 2	Natural Gas	4	50	Do Nothing		
				Expand Natural Gas	150	4
				Build Petroleum Plant	**50**	**5**
Gen 3	Nuclear	2	100			
	Year 2					
Gens	Current Plant Type	Bus #	Current Capacity (MW)	Proposed Investment Alternatives	Proposed Capacity (MW)	Proposed Bus Location
Gen 1	Natural Gas	1	50	Do Nothing		
	Coal Plant	3	50	Expand Natural Gas	100	1
				Expand Coal Plant	**100**	**3**
Gen 2	Natural Gas	4	50	Do Nothing		
	Petroleum	5	50	**Expand Natural Gas**	**150**	**4**
				Expand Petroleum Plant	100	5
Gen 3	Nuclear	2	100			
	Year 3					
Gens	Current Plant Type	Bus #	Current Capacity (MW)	Proposed Investment Alternatives	Proposed Capacity (MW)	Proposed Bus Location
Gen 1	Natural Gas	1	50	**Do Nothing**		
	Coal Plant	3	100	Expand Natural Gas	100	1
				Expand Coal Plant	150	3
Gen 2	Natural Gas	4	150	Do Nothing		
	Petroleum	5	50	Expand Natural Gas	200	4
				Expand Petroleum Plant	**100**	**5**
Gen 3	Nuclear	2	100			
	Year 4					
Gens	Current Plant Type	Bus #	Current Capacity (MW)	Proposed Investment Alternatives	Proposed Capacity (MW)	Proposed Bus Location
Gen 1	Natural Gas	1	50	Do Nothing		
	Coal Plant	3	100	Expand Natural Gas	100	1
				Expand Coal Plant	**150**	**3**
Gen 2	Natural Gas	4	150	Do Nothing		
	Petroleum	5	100	Expand Natural Gas	200	4
				Expand Petroleum Plant	**150**	**5**
Gen 3	Nuclear	2	100			

2 has the following investment alternatives: do nothing, expand capacity of existing natural gas plant at Bus 4 from 50 MW to 150 MW, or build a petroleum fired plant at Bus 5 with capacity of 50 MW.

The investment alternatives chosen here are somewhat arbitrary. In reality, feasible expansion plans can be developed by considering one year of operation at a time. Operating profits for potential expansion actions can then be calculated from previous years dispatch and LMP results. Investment costs can be annualized based on a risk-adjusted discount rate and subtracted from the operating profits. This gives a ranking of technologies from which a selection of possible expansion actions could be made. An expansion action can consist of a single plant or a portfolio of new plants at different locations in the grid. A user-defined limit could be imposed on the number of possible actions for each generator to reduce the dimensionality of the problem. Note that we do not consider transmission investments as potential expansion actions in the model. However, investments in new transmission can be specified as an exogenous input to the model. Future retirements of existing generating plants and transmission lines can also be included as a user-defined input. However, we do not perform an investment analysis in this work. Our goal here is to demonstrate how multi-year, multi-player generation expansion plans can be obtained by a set of competing generators with given investment alternatives.

In the numerical example, the three investment alternatives for both Generators 1 and 2 give rise to an investment matrix game with nine (3×3) elements, each of which is a potential expansion alternative. Each year, all the expansion alternatives are analyzed as supply function matrix games and profits are calculated using OPF and CVaR models. Thereafter, the equilibrium expansion plans are obtained using the reinforcement learning algorithm. This procedure is repeated for each year of the planning horizon.

The equilibrium expansion plans obtained by the reinforcement learning algorithm for year 1 are: Generator 1 builds a coal plant at Bus 3 and Generator 2 builds the petroleum fired plant at bus 5. This expansion plan is shown in a gray shade in year 1 of the Table 8.5, indicating that it is a pure strategy Nash equilibrium solution. The equilibrium solution of the two-tier model from year 1 is assumed to be part of the existing network for the subsequent years.

The network configuration for year 2 is shown in Table 8.5. It shows both the current plants in the network as well as the equilibrium plans chosen in year 1. The expansion alternatives considered for Generator 1 for year 2 are: do nothing, expand the natural gas plant at Bus 1 to 100 MW, and expand the newly built coal plant further up to 100 MW. Generator 2 considers the following investment alternatives: do nothing, expand natural gas plant at Bus 4 to 100 MW, or expand the petroleum fired plant at Bus 5 to 100 MW. The two-tier model is solved again for year 2 giving an equilibrium expansion plan for both generators. The equilibrium expansion plan for year 2 is: Generator 1 expands the coal plant at Bus 3 and Generator 2 expands the natural gas plant at Bus 4 to 150 MW. This is shown in a gray shade in the Table. Similarly, the equilibrium expansion strategies are successively computed for years 3 and 4. We present the nodal prices (p1 through p5) and quantity allocations (q1 through q5) obtained from the two-tier model for the four year planning horizon in Table 8.6.

The payoff matrices for the investment game for all the four years of expansion are shown in Table 8.7, where the two elements in each cell represent the payoffs for generators 1 and 2 respectively. Each element of the payoff matrix is the solution of a supply function matrix game. The payoff matrices for the supply function matrix games are not shown here for the sake of brevity. It may be noted from the payoff matrices that the reinforcement learning algorithm finds the pure strategy Nash

66

Table 8.6. Price and Quantity Allocations for the Four Year Planning Horizon

Year	p1 ($/MWH)	p2 ($/MWH)	p3 ($/MWH)	p4 ($/MWH)	p5 ($/MWH)	q1 (MWH)	q2 (MWH)	q3 (MWH)	q4 (MWH)	q5 (MWH)
1	21	21	18	21	21	19	45	50	24	22
2	24	24	24	24	24	20	54	69	36	25
3	32	32	32	32	32	20	87	96	33	36
4	55	35	48	55	55	37	100	150	59	67

equilibrium (NE) for each year of the expansion. NE is defined as that combination of strategies from which no player will gain by unilaterally deviating. These pure strategy NE solutions are highlighted in a gray shade in each of the four years in the table.

8.1.1.1   Mixed Strategies and Multiple Equilibria

It is well known that matrix games may not always have a pure strategy NE but will always have a mixed strategy NE. However, for problems such as GEP, a mixed strategy solution is impractical from an application standpoint. Therefore, when there is no pure strategy NE solution, the generators should consider other good "out-of-equilibrium" [84] pure strategies generated by our RL algorithm. In some other cases, a matrix game may have multiple pure strategy NE. In these cases, as shown in Chapter 5 ([10]), our value based reinforcement learning algorithm identifies the NE with the best value.

8.1.1.2   Generator Profits and Consumer Surpluses

Generators can examine their profits under perfect competition (where generators bid at marginal costs) and under imperfect competition (strategic bidding), to gauge

Table 8.7. Generator Payoff Matrices and Game Solutions for the Four Year Planning Horizon ($\zeta = 0.1$)

Year 1 & $\zeta = 0.1$			
Gen 1	**Gen 2**		
	Do Nothing	**Expand Natural Gas**	**Build Petroleum**
Do Nothing	420, 576	514, 625	347, 691
Expand Natural Gas	442, 592	481, 603	339, 727
Build Coal Plant	604, 358	659, 422	**561, 440**
Year 2 & $\zeta = 0.1$			
Gen 1	**Gen 2**		
	Do Nothing	**Expand Natural Gas**	**Expand Petroleum**
Do Nothing	692, 697	748, 758	698, 721
Expand Natural Gas	709, 739	749, 794	692, 697
Expand Coal Plant	908, 609	**958, 649**	940, 608
Year 3 & $\zeta = 0.1$			
Gen 1	**Gen 2**		
	Do Nothing	**Expand Natural Gas**	**Expand Petroleum**
Do Nothing	1625, 1090	1553, 1089	**1728, 1114**
Expand Natural Gas	1680, 1117	1752, 1086	1620, 1146
Expand Coal Plant	1628, 1092	1624, 1121	1688, 1093
Year 4 & $\zeta = 0.1$			
Gen 1	**Gen 2**		
	Do Nothing	**Expand Natural Gas**	**Expand Petroleum**
Do Nothing	3364, 5218	3504, 5182	3608, 5029
Expand Natural Gas	3609, 4912	3559, 4631	3462, 4752
Expand Coal Plant	4554, 3338	4495, 3538	**4452, 3591**
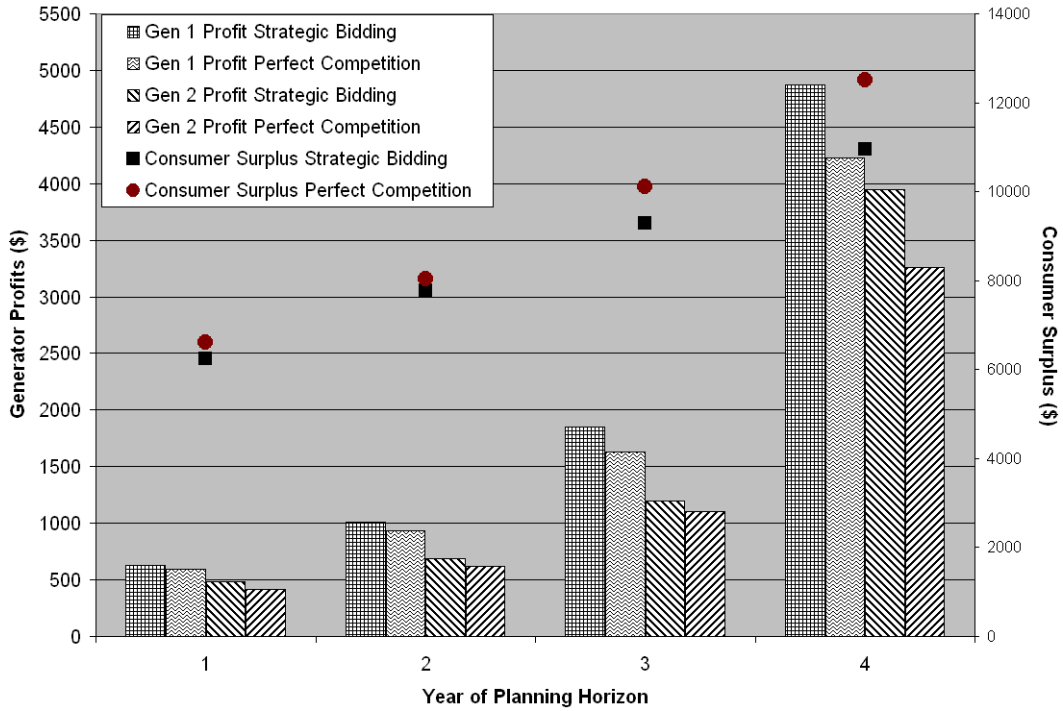
Figure 8.2. Profits and Consumer Surpluses in Strategic Bidding and Perfect Competition

the amount of additional profit they can make. The Figure 8.2 presents a plot of generator profits (primary Y-axis) and consumer surplus (secondary Y-axis) against the planning horizon. It can be seen from the graph that the total profits made by both generators under perfect competition are lesser than those under strategic bidding. The difference in profit grew larger with increase in demand. This is expected, since higher demand provides more opportunities for strategic bidding by generators. The graph also shows consumer surplus along the 4-year planning horizon. By observing the changes in consumer surplus, generators can assess how much of the overall surpluses they are able to transfer to themselves. However, generators need to be wary of bidding too high in the market, since higher bids may lead to higher profits, resulting in eroding consumer surpluses, which may invite a potential regulatory intervention.

69

8.1.2 Nodal Price Sensitivity Analysis

The objective of the following statistical analysis is to assess the effect that line capacity and the slope of the demand curve have on the price at each bus post-expansion. Line capacity was deemed an important factor to include in this analysis due to its direct impact on transmission congestion and, as a result, on the price. The slope of the consumer demand curve is included in the analysis because its price elasticity is expected to have an impact on the type of expansion plan chosen by a generator.

Understanding the influence that line capacity and demand exert over post expansion prices is beneficial for all stakeholders in the electricity market. It enables generators to forecast how expansion plans will impact nodal prices. Likewise, the ISO can assess the impact of line capacity restrictions and demand variations on the nodal prices after a potential expansion decision. Finally, the consumers can also benefit by examining what type of demand curve variations can help them to possibly attain lower post-expansion prices.

For the sake of simplicity and to aid the visualization of potential effects, line capacity was varied among three levels and only considered in lines $1-3, 3-4, 4-5$ (see Figure 8.1). The rationale behind selecting these lines lies in the fact that nodes 1, 3, 4, and 5 are considered for potential expansions. The demand slope is varied among 5 values at nodes 2, 3, 4, and 5 (the same demand slope is used at all nodes for each run). This yields a two factor mixed factorial experiment with the factors at three and five levels respectively. Table 8.8 presents the levels for demand and line capacity. After observing the prices before and after expansion for fifteen runs (3 levels of line capacity $\times$ 5 levels of slope of the demand curve) the difference in nodal prices (for each bus) was captured as the response variable. An analysis of variance

Table 8.8. Line Capacity and Consumer Demand Slope Levels

Slope of Demand Curve				
-3	-2.5	-2	-1.5	-1
Line Capacity in MW (Congestion)				
20		15		10

Table 8.9. F Statistic Values in 2-Factor Design

	F dem	F linecap	F int
ΔP1	0.71	3.49**	4.92
ΔP2	4.55*	1.91	0.38
ΔP3	14.52*	6.40*	0.22
ΔP4	2.80	2.74	1.46
ΔP5	0.46	2.75	6.04*

(ANOVA) for the factorial design was carried out whose results are presented in Table 8.9. Note that in Table 8.9 those values accompanied by $*$ were found to be significant at the level of 0.05, whereas those with $**$ were significant at 0.1. Based on results from Table 8.9, demand does not appear to play a significant role in bus 1 and neither does line capacity in bus 2. Demand and line capacity appear to have a significant effect on the price differential observed in bus 3, whereas the interaction between them was the only significant factor at bus 5.

In accordance with the results from the factorial design, regression models were fitted to measure the effect of each factor on the price differential (ΔP = nodal price after expansion - nodal price before expansion) at each bus. Table 8.10, presents the coefficients obtained for a model with the general form,

$$y = \beta_0 + \beta_1 x_{dem} + \beta_2 x_{linecap} + \beta_3 x_{dem} x_{linecap}. \tag{8.1}$$

Since line capacity was the only significant factor affecting ΔP at bus 1 (see Table 8.9), the model at bus 1 was fitted considering only line capacity. The negative sign

71

Table 8.10. Linear Regression Model Coefficients

	β0	β1	β2	β3
ΔP1	13.77	-	-1.02	-
ΔP2	0.77	2.90	-	-
ΔP3	-82.66	-18.92	1.49	-
ΔP4	-	-	-	-
ΔP5	-67.48	-24.07	3.51	1.3944

in the line capacity coefficient implies that the higher the line capacity the higher the reduction in price after the expansion plans are implemented. This result fits within the framework of basic microeconomic theory. Reduced line capacity causes congestion, generating less resource availability and thereby, increased prices. If line capacity is reduced to 13.5 MW (54% of full capacity) the price differential is zero. Smaller values of line capacity will increase the price post-expansion at that bus.

Only the slope of the demand was considered in the regression model for bus 2. The resulting regression coefficient was positive implying that the steeper the demand curve the higher the reduction of price at the bus. Flat demand curves (i.e. values for the slope less than 0.26) will generate increments in price. A steep demand curve corresponds to a comparatively more inelastic demand, hence capacity expansion will have high impact on price reduction.

The models obtained for nodes 3 and 5 are more complex because they involve both factors and in the case of bus 5 an interaction term is also present. These predictive models are useful in examining network behavior. For example, the equation for bus 3: $y = -82.66 - 18.92\, x_{dem} + 1.49\, x_{linecap}$, can be used to show that for a line capacity of 25MW, a negative price differential will occur only as long as the absolute value of the slope of the demand curve is smaller than 2.4. That is, for less elastic consumer demand (absolute value of slope higher than 2.4) the price differential will become positive (increasing nodal price). Even though conducting a bus-by-bus analysis, as

presented above, may become cumbersome for very large size networks, it can be accomplished for those buses deemed "critical" in the network.

8.1.3 Impact of Risk Preference on GEP

We examined how GEP outcomes vary with respect to generator risk preferences ζ_i (see Equation 4.4). For a given power network, this could serve to establish the ζ_i values at which the generators begin to switch their investment choices. Figure 8.3 shows the different expansion plans chosen for different values of the risk preferences for demand corresponding to year 1 of Table 8.1. Note that, higher values of risk preference ζ_i indicates higher level of risk aversion. Results show that though generator 1's expansion plan does not change, the investment decisions at a network level vary with increasing risk aversion.

In the analysis above, the same ζ_i value was assumed for both generators. However, we can also use our model to observe how different risk preferences of generators can have an impact on the expansion plans. Generators or ISO can also examine the changes in GEP pattern when the risk preferences change over the years. For example, a generator who may have recently invested in a plant, might be averse to risk in the subsequent few years. Such variations in risk preferences can be considered by our model.
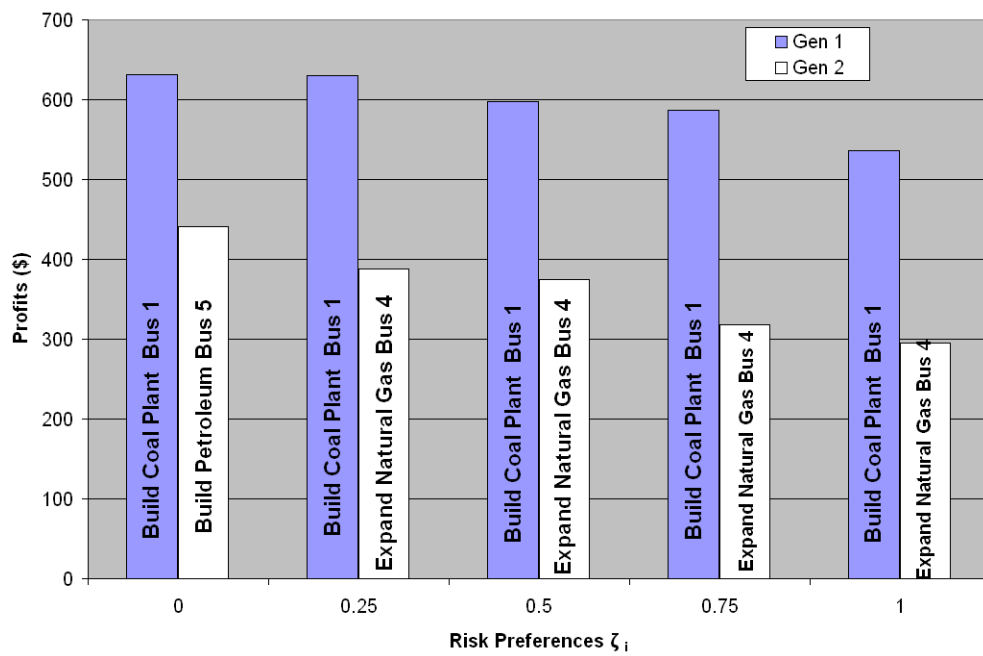
Figure 8.3. Profits and Expansion Plans versus Risk Preferences

CHAPTER 9
CONCLUDING REMARKS

Electricity market restructuring gave rise to a new realm of issues that needed to be addressed, that were not seen in the research presented in the era of regulated markets. A great amount of research has been devoted to issues like market design, market power assessment, financial transmission rights, capacity markets, and ancillary markets, however, the research in generation expansion planning in restructured markets has significant room for exploration. This research presented a novel methodology to aid this exploration.

9.1   Advances Made by this Research in GEP

As explained earlier, currently, over fifteen states in the US have restructured energy markets, where generators are required to compete in an open market to supply power into the electricity grid. The Energy Information Administration (EIA) forecasts that several power plants are to be constructed in these restructured markets, leading to investments of billions of dollars in the next two decades to satisfy the rising demand for electricity. This issue, referred to as generation capacity expansion, has been very well studied in regulated markets. However, models developed under regulated settings had an optimization structure considering only a single decision maker, and are rendered obsolete under the new restructured energy market paradigm. In the restructured markets, this issue, as presented in this dissertation, needs to be modeled as a game-theoretic problem since it requires the simultaneous consideration

of several competing decision makers. One of the primary research contributions of this dissertation is the development of a comprehensive game-theoretic model which considers several important restructured market features like competitive behavior of the generators, transmission line constraints, voltage & current related constraints, multiple year planning horizon, power demand stochasticity, and risk due to volatilities in profit. Such a comprehensive game-theoretic model with all the aforementioned features is a significant advancement in the area of generation capacity expansion in restructured electricity markets.

Our model examines the non-cooperative competition of generators at two tiers. At one tier the generation investment game is examined and at another tier the supply function game at the power network operational level is examined. We present a novel solution algorithm for the two-tier model that shows how these tiers interact to obtain a multi-year generation expansion plan. Using a sample electric power network the applicability of the methodology is demonstrated. There are some important features that can be included in the model to enhance its applicability. Features such as reliability and capacity markets can be incorporated in the DC-OPF model, using the strategy presented in [23, 85]. Furthermore, the OPF model that we adopt in this research is a linearized DC-OPF version (ala Hobbs et al. [46, 76]), and was chosen only to simplify the computation of supply function equilibria. The advantages of using such linearized supply functions were discussed in great detail in [86]. One could also replace the DC-OPF with AC-OPF model and step-function bidding, as in our previous work ([87]).

The model presented in this research can be beneficial for all power market constituents: generators, consumers, and ISO. The matrix game approach we adopt in this research allows generators to assess the profitability of several investment alternatives by incorporating risk preferences and CVaR. We believe that the consideration

of risk preferences and CVaR makes the investment decisions generated by our model more robust. Consumers can use our model to study how different expansion plans adopted by generators will affect nodal prices, while the ISO can test different market designs aimed at maximizing social welfare.

9.2 Practical Applications of Matrix Games

The multiplayer electric power market problem we solved in Chapter 5 as well as the multiplayer GEP problem we solved in Chapter 8 serve as excellent examples of real life application of matrix games. Outcomes of such games determine the nature of hourly and daily power transactions as well as the ability of a market to meet demand growth over several years. Hence, the ability to accurately obtain NE for matrix games allows for better assessment of market performance and efficient market design, which translate to stable power market operations with limited price spikes. Solutions of relatively large matrix games (of size 625×625) resulting from the sample power network problem indicate the algorithms potential to tackle real life power networks, which can be magnitudes larger in size. Though the numerical results are promising and encourage further exploration of our algorithms performance, a theoretical proof of convergence and optimality is required. We believe that such a proof can be constructed following the logic used in ([33]), and we are currently working on developing such a proof.

Work in progress and future research directions include the theoretical convergence analysis of the reinforcement learning algorithm. An interesting line of research will be the application of the multiplayer matrix game approach in other fields that have oligopolistic competition. Some examples of such oligopolistic markets include retail sales, home and auto insurance, mortgage lending, service industries like airlines, hotels, and entertainments.

CHAPTER 10

FUTURE RESEARCH DIRECTIONS

The restructuring of the electricity industry has given rise to a new realm of exciting as well as extremely challenging modeling issues at the intersection of operations research, public-policy, economics, and risk management. Such modeling issues, due to their inherent complexities with several dynamic elements can be best modeled and solved using computational optimization approaches. Computational modeling affords the distinct advantage of being able to handle extremely complex systems with data-rich environments. Energy markets are noted by National Science Foundation and in literature as complex systems that need to be modeled using computational approaches to be able to comprehensively capture all the stochastic dynamics. Moreover, computational models help in furthering our fundamental understanding of the complex interactions of multilevel, multi-scale systems such as energy markets, which may not be possible via traditional modeling approaches

The research presented in this dissertation will help in jointly addressing two emerging areas of critical national importance: generation capacity expansion in restructured (deregulated) energy markets to meet the growing energy demands (discussed so far), and environmental emission control via carbon cap-and-trade (CT) programs and renewable portfolio standards (RPS), aimed at reducing the negative impact of electric power generation on climate change. The objective of this chapter is to briefly discuss how these areas are closely intertwined and explain the need for

the development of a comprehensive stochastic optimization model to jointly address these issues.

10.1 Cap-and-Trade Programs and Renewable Portfolio Standards

Electricity generators in the US are the single largest industrial contributors of CO_2 and are one of the leading causes of the climate change crisis. Due to the need for a fierce urgency in reversing detrimental human effects on the earth's climate, countries around the world as well as some regions in the US have enacted carbon cap-and-trade (CT) programs for emissions reductions.

Cap-and-trade systems have historically been used as an effective market mechanism to limit the emission of pollutants like NO_X and SO_2 [88]. Since the inception of such a program in the U.S. in 1995, there have been significant emissions reductions and far ranging environmental as well as human health benefits, at a lower than expected compliance cost. Economists and policy analysts believe that implementation of a similar cap-and-trade system for CO_2 should be a central element of any emissions control policy [89]. A cap-and-trade system would establish CO_2 emission limits either at an upstream level for producers of fuels, or at a downstream level for industrial consumers of fuels including electricity generators. Regulated entities will buy allowances that will permit them to release a certain amount of CO_2 within a specified period of time. If the emissions exceed allowances, entities need to purchase more allowances or pay a penalty in terms of increased price of allowance purchase for the next period. On the other hand, if an entity does not use all of its allowances, they can be banked for future or sold in an open secondary market [89, 90]. By gradually lowering the cap on total emissions, regulated entities will be forced to invest in cleaner sources of energy and *greener* technologies. Different variations of the cap-and-trade system have been operational in Europe [91].

Recently, the first CO_2 CT program of US, the Regional Greenhouse Gas Initiative (RGGI), became fully functioning in 10 Northeastern and Mid-Atlantic States. Also, the Western Climate Initiative and the Midwestern Greenhouse Gas Accord, comprising over a dozen states, are set to commence early next decade. With the implementation of such initiatives, per World Resources Institute (WRI), almost 50% of the population of US will reside in states with CT programs, out of which several markets are restructured.

Renewable portfolio standards (RPS), per Energy Efficiency and Renewable Energy division of Department of Energy, is a state policy that requires electricity providers to obtain a minimum percentage of their power from renewable energy resources by a certain date [92]. With the enforcement of renewable portfolio standards (RPS), electric utilities in over 24 states in the coming decade are required to produce a significant percentage of electricity using renewable energy sources like biofuels, solar, wind, and geothermal. Since carbon-based fuels are the current primary sources for electricity generation, the implementation of CT and RPS is expected to trigger a fundamental transformation in the technologies used to produce electricity in the coming decades.

10.2 Joint Model for GEP and Emissions Control

Since electricity generation and climate change crisis are interrelated, the generation capacity expansion planning problem and environmental emission control are closely connected. To address this critical societal challenge, similar to the model presented in this dissertation, a comprehensive stochastic optimization (game-theoretic) model needs to be developed, which will capture at multiple interconnected tiers: the competition among generators for capacity investments, the competition for allowances in the CT markets, and the optimization of electric power flow while meeting

RPS. Solving such complex models is almost impossible with traditional optimization approaches. Therefore, to address the stochastic dynamics involved in solving this problem, a simulation-based optimization computational solution methodology similar to the one developed in this research must be formulated.

10.3 Further Policy Analysis and Planning Applications

Several issues presented below can be addressed based on the models presented in this work.

1. Examining the effect of allowance prices on electricity market prices,

2. CO_2 levels over a long term planning horizon,

3. effects of different allowance allocation methods: auctioning, grandfathering, or hybrid models,

4. examining effects of different risk attitudes of investors on expansions,

5. examination of portfolio of generation plants over a long term horizon,

6. effect of RPS on generation expansion on a state by state basis, and

7. effect of large-scale introduction of *microgrids* as a potential expansion alternative.

REFERENCES

[1] S. Stoft. *Power System Economics.* IEEE Press, New Jersey, 2002.

[2] A. J. Wood and B. F. Wollenberg. *Power Generation Operation and Control.* John Wiley Sons, Inc, New York, 1994.

[3] K. Bhattacharya, M. H. J. Bollen, and J. E. Daalder. *Operation of restructured power systems.* Kluwer academic publisher, Boston, 2001.

[4] M. Ventosa, A. Baillo, A. Ramos, and M. Rivier. Electricity market modeling trends. *Energy Policy*, 33:897–913, 2005.

[5] C. J. Day, B. F. Hobbs, and J. S. Pang. Oligopolistic competition in power networks: A conjectured supply function approach. *IEEE Transactions on Power Systems*, 17(3):597–607, 2002.

[6] J. Boucher and Y. Smeers. Alternative models of restructured electricity systems, part1: No market power. *Operations Research*, 49(6):821–838, 2001.

[7] V. Nanduri and T. K. Das. A survey of critical research areas in the energy segment of restructured electricity markets. *Electrical Power and Energy Systems*, 31(5):181–191, 2009.

[8] National Energy Policy. Report of the national energy policy development group. Technical report, 2001.

[9] Annual Energy Outlook. Annual energy outlook, with projections to 2030. Technical report, Energy Information Administration, 2007.

[10] V. Nanduri and T. K. Das. A reinforcement learning approach for obtaining the Nash equilibrium of multiplayer matrix games. *IIE Transactions*, 41:158–167, 2009.

[11] T. K. Das, A. Gosavi, S. Mahadevan, and N. Marchalleck. Solving semi-Markov decision problems using average reward reinforcement learning. *Management Science*, 45(4):560–574, 1999.

[12] A. Gosavi, N. Bandla, and T. K. Das. A reinforcement learning approach to airline seat allocation for multiple fare classes with overbooking. *IIE transactions, Special issue on advances on large-scale optimization for logistics, production and manufacturing systems*, 34(9):729–742, 2002.

[13] A. Gosavi. A convergent reinforcement learning algorithm for solving semi-Markov decision problems under average cost for an infinite time horizon. *European Journal of Operational Research*, 155:654–674, 2003.

[14] P. Masse and R. Gilbrat. Application of linear programming to investments in electric power industry. *Management Sciences*, 3(2):149–166, 1957.

[15] J. A. Bloom. Long-range generation planning using decomposition and probabilistic simulation. *IEEE Transactions on Power Systems*, 101(4):797–802, 1982.

[16] B. Mo, J. Hegge, and I. Wangensteen. Stochastic generation expansion planning by means of stochastic dynamic programming. *IEEE Transactions on Power Systems*, 6(2):662–668, May 1991.

[17] A. B. Borison, P. A. Morris, and S. Oren. A state-of-the-world decomposition approach to dynamics and uncertainity in electric utility generation expansion planning. *Operations Research*, 32(5):1052–1068, 1984.

[18] J. J. C. Meza, M. B. Yilidirim, and A. S. M. Masud. A model for the multi-period multiobjective power generation expansion problem. *IEEE Transactions on Power Systems*, 22(2):871–878, May 2007.

[19] B. F. Hobbs. Optimization methods for electric utility resource planning. *European Journal of Operational Research*, 83:1–20, 1995.

[20] J. Zhu and M-. Y. Chow. A review of emerging techniques on generation expansion planning. *IEEE Transactions on Power Systems*, 12(4):1722–1728, November 1997.

[21] S. T. Jenkins and D. S. Joy. Wien automatic system planning package (WASP)-an electric utility optial generation expansion planning computer code. Technical report, Oak Ridge National Laboratory, ORNL 4945, 1974.

[22] Electric Power Research Institute. Electric generation expansion analysis system (EGEAS). Technical report, EPRI, Palo Alto, CA, 1982.

[23] A. S. Chuang, F. Wu, and P. Varaiya. A game-theoretic model for generation expansion planning: Problem formulation and numerical comparisons. *IEEE Transactions on Power Systems*, 16(4):885–891, 2001.

[24] F. Murphy and Y. Smeers. Generation capacity expansion in imperfectly competitive restructured electricity markets. *Operations Research*, 53(4):646–661, July-August 2005.

[25] P. Kaymaz, J. Valenzula, and C. S Park. Transmission congestion and competition on power generation expansion. *IEEE Transactions on Power Systems*, 22(1):156–163, 2007.

[26] P. Jirutitijaroen and C. Singh. Reliability and cost tradeoff in multiarea power system generation expansion using dynamic programming and global decomposition. *IEEE Transactions on Power Systems*, 21(3):1432–1441, August 2006.

[27] J-. H. Kim, J-. B. Park, J-. K. Park, and S-. K. Joo. A market-based analysis on the generation expansion planning strategies. *ISAP Conference Proceedings*, pages 458–463, 2005.

[28] S. K. K. Ng, C. W. Lee, and J. Zhong. A game theoretic approach to study strategic interaction between transmission and generation expansion planning. *IEEE Conference Proceedings*, 2006.

[29] A. Ehrenmann and Y. Smeers. Capacity expansion in non regulated electricity markets. *Proceedings of the 39th Hawaii International Conference in System Sciences*, 2006.

[30] A. Botterud, M. Ilic, and I. Wangensteen. Optimal investments in power generation under centralized and decentralized decision making. *IEEE Transactions on Power Systems*, 20(1):254–263, Feb 2005.

[31] B. F. Hobbs. Linear complementarity models of nash-cournot competition in bilateral and POOLCO power markets. *IEEE Transactions on Power Systems*, 16(2):194–202, 2001.

[32] J. Nash. Non-cooperative games. *The Annals of mathematics.*, 54(2):286–295, 1951.

[33] J. Li, T. K. Das, and K. M. Ramachandran. Learning Nash equilibrium policies in average reward irreducible stochastic games. *Working paper, Department of IMSE, USF*, 2009.

[34] J. Hu and M. P. Wellman. Multiagent reinforcement learning: Theoretical framework and an algorithm. In *Proc. 15th International Conf. on Machine Learning*, Madison, WI, 1998.

[35] V. S. Borkar. Reinforcement learning in Markovian evolutionary games. *Advances in Complex Systems*, 5(1):55–72, 2002.

[36] J. Filar and K. Vrieze. *Competitive Markov Decision Processes.* Springer Verlag, New York, 1997.

[37] R. McKelvey and A. McLennan. Computation of equilibria in finite games. In J. Rust H. Amman, D. Kendrick, editor, *Handbook of Computational Economics*, pages 87–142. Elsevier, 1996.

[38] C. E. Lemke and J. T. Howson Jr. Equilibrium points of bimatrix games. *SIAM J. of App. Math.*, 12:413–423, 1964.

[39] J. Rosenmuller. On the generalization of the lemke-howson algorithm to noncooperative n-person games. *SIAM J. of App. Math.*, 21(1):73–79, 1971.

[40] R. Wilson. Computing equilibria of n-person games. *SIAM Applied Math*, 21:80–87, 1971.

[41] L. Mathiesen. Computational experience in solving equilibrium models by a sequence of linear complementarity problems. *Operations Research*, 33(6):1225–1250, 1985.

[42] P. T. Harker and J. S. Pang. Finite dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming*, 48:161–220, 1990.

[43] R. D. McKelvey, A. M. McLennan, and T. L. Turocy. GAMBIT: Software tools for game theory, version 0.2007.01.30. 2007.

[44] K. H. Lee and R. Baldick. Solving three-player games by the matrix approach with an application to an electric power market. *IEEE Transactions on Power Systems*, 18(4):166–172, 2003.

[45] B. F. Hobbs. Linear complementarity models of Nash-Cournot competition in bilateral and POOLCO power markets. *IEEE Transactions On Power Systems*, 16(2):194–202, May 2001.

[46] B. F. Hobbs, B. M. Carolyn, and J. S. Pang. Strategic gamming analysis for electric power systems: An MPEC approach. *IEEE Tran. Power Syst.*, 15(2):638–645, 2000.

[47] J. Yao, S. Oren, and I. Adler. Computing Cournot equilibria in two settlement electricity markets with transmission constraints. *Proceedings of 37th Hawaii international conference on system sciences*, pages 1–9, 2003.

[48] T. C. Price. Using co-evolutionary programming to simulate strategic behavior in markets. *J. Evol. Econ.*, 7:219–254, 1997.

[49] L. B. Cunningham, R. Baldick, and M. L. Baughman. An empirical study of applied game theory: Transmission constrained cournot behavior. *IEEE Transactions on Power Systems*, 17(1):166–172, 2002.

[50] R. W. Ferrero, M. Shahidehpour, and V. C. Ramesh. Transaction analysis in deregulated power systems using game theory. *IEEE Transactions on Power Systems*, 12(3):1340–1347, 1999.

[51] S. Stoft. Using game theory to study market power in simple networks. *IEEE Tutorial on Game Theory in Electric Power Markets*, pages 33 – 40, 1999.

[52] Z. Luo, J. S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Boston, 1996.

[53] C. L. Su. Equilibrium problems with equilibrium constraints: Stationarities, algorithms, and applications. *Dissertation from the Department of Management Sciences, Stanford University*, 2005.

[54] D. Ralph and Y. Smeers. Epecs as models for electricity markets. *Power Systems Conference and Exposition, Atlanta, GA*, 2006.

[55] J. B. Cardell, C. C. Hitt, and W. W. Hogan. Market power and strategic interaction in electricity networks. *Resource and Energy Economics*, 19:109–137, 1997.

[56] V. Gountis and A. G. Bakirtzis. Bidding strategies for electricity producers in a competitive electricity marketplace. *IEEE Transactions on Power Systems*, 19(4):356–365, 2004.

[57] C. W. Richter Jr. and G. Sheble. Genetic algorithm evolution of bidding strategies for the competitive market place. *IEEE Transactions on Power Systems*, 13(1):256–261, 1998.

[58] C. W. Richter Jr., G. Sheble, and D. Ashlock. Comprehensive bidding strategies with genetic programming/finite automata. *IEEE Transactions on Power Systems*, 14(4):1207–1212, 1999.

[59] P. Attaviriyanupap, H. Kita, E. Tanaka, and J. Hasegawa. New bidding strategy formulation for day ahead energy and reserve markets based on evolutionary programming. *Electrical Power and Energy Systems*, 27:157–167, 2005.

[60] F. Wen and A. K. David. Optimal bidding strategies and modeling of imperfect information among competitive generators. *IEEE Transactions on Power Systems*, 16(1):15–21, 2001.

[61] R. Rajaraman and F. Alvarado. Optimal bidding strategies in electricity markets under uncertain energy and reserve prices. Technical report, Power Systems Engineering Research Center, Paper 03-05, 2003.

[62] C. Li, A. J. Svoboda, X. Guan, and H. Singh. Revenue adequate bidding strategies in competitive electricity markets. *IEEE Transactions on Power Systems*, 14(2):492–497, 1999.

[63] A. Seeger. *Recent Advances in optimization.* Lecture Notes in Economics and Mathematical Systems. Springer, 2004.

[64] K. G. Murthy. *Linear Complementarity, Linear, and Nonlinear Programming*, chapter The LCP and its Geometry, pages 1–61. Internet Edition, 1999.

[65] K. H. Lee and R. Baldick. Tuning of discretization in bimatrix game approach to power system market analysis. *IEEE Transactions on Power Systems*, 18(2):830–836, MAY 2003.

[66] J. Yao. *Cournot equilibrium in two settlement electricity markets:formulation and computation.* PhD thesis, University of California, Berkeley, 2006.

[67] A. Gosavi. A convergent reinforcement learning algorithm for solving semi-Markov decision problems under average cost for an infinite time horizon. *In press, European Journal of Operational Research*, 2003.

[68] M.L. Puterman. *Markov Decision Processes.* John Wiley and Sons, New York Chichester Brisbane Toronto Singapore, 1994.

[69] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statist.*, 22:400–407, 1951.

[70] A. Gosavi. *Simulation-Based Optimization : Parametric Optimization Techniques and Reinforcement Learning.* Operations Research/Computer Science Interfaces Series. Springer, 2003.

[71] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An introduction.* MIT Press, Cambridge, MA, 1998.

[72] T. R. Rockafellar and S. Uryasev. Optimization of conditional value at risk. *Journal of Risk*, 2(3):21–41, 2000.

[73] T. R. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):21–41, 2002.

[74] J. Su and F. Wu. Evaluation of generation expansion investment under competitive market environment. *IEEE Power and Energy Society GM Proceedings*, pages 2136–2140, 2005.

[75] C. Acerbi and D. Tasche. Expected shortfall: A natural coherent alternative to value-at-risk. *Economic Notes*, 31(2):379–388, 2002.

[76] C. A. Berry, B. F. Hobbs, W. A. Meroney, R. P. ONeill, and W. R. Stewart Jr. Understanding how market power can arise in network competition: A game theoretic approach. *Utilities Policy*, 8:139–158, 1999.

[77] L.S. Shapley. Stochastic game, 1953. In H.W Kuhn, editor, *Classics in Game Theory*. Princeton University Press, 1997.

[78] M.L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *In Proceedings of the Eleventh International Conference on Machine Learning*, pages 151–163, San Francisco, CA, 1994.

[79] J. Hu and M. P. Wellman. Nash Q-learning for general-sum stochastic games. *Journal of Machine Learning Research*, 4:1039–1069, 2003.

[80] L.P. Kaelbling, M.L. Littman, and A.W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

[81] M. Bowling and M.M. Veloso. An analysis of stochastic game theory for multiagent reinforcement learning. Technical Report CMU-CS-00-165, Computer Science Department, Carnegie Mellon University, 2000.

[82] C. Darken, J. Chang, and J. Moody. Learning rate schedules for faster stochastic gradient search. In D.A. White and D.A. Sofge, editors, *Neural Networks for Signal Processing 2 - Proceedings of the 1992 IEEE Workshop*. IEEE Press, Piscataway, NJ, 1992.

[83] PowerWorld Simulator. *Version 8.0 Glover/ Sarma Build 11/ 02/ 01, PowerWorld Corporation, 2004 South Wright Street, Urbana, IL-61801, http://www. powerworld. com.*

[84] W. B. Arthur. Out-of-equilibrium economics and agent-based modeling. In K. Judd and L. Tesfatsion, editors, *Handbook of Computational Economics*, volume 2. Elsevier/North-Holland, 2005.

[85] G. L. Doorman and A. Botterud. Analysis of generation investment under different market designs. *IEEE Transactions on Power Systems*, 23(3):859–867, August 2008.

[86] R. Baldick, R. Grant, and E. Kahn. Theory and application of linear supply function equilibrium in electricity markets. *Journal of Regulatory Economics*, 25(2):143–167, 2004.

[87] V. Nanduri and T. K. Das. A reinforcement learning model to assess market power under auction-based energy pricing. *IEEE Transactions on Power Systems*, 22(1):85–95, 2007.

[88] A. D. Ellerman. Are cap-and-trade programs more enviromentally effective than conventional regulation. In J. Freeman and C. D. Kolstad, editors, *Moving to markets in enviromental regulation: Lessons from Twenty Years of Experience*, chapter 3, pages 48–62. Oxford University Press US, 2006.

[89] R. N. Stavins. A U.S. cap-and-trade system to address global climate change. *The Hamilton Project, Discussion Paper 2007-13. Washington, D.C.: The Brookings Institution,*, 2007.

[90] Market Advisory Committee. Recommendations for designing a greenhouse gas cap-and-trade system for california. *Recommendations of the Market Advisory Committee to the California Air Resources Board, http:// www. climatechange. ca. gov/ market_ advisory_ committee/ index.html*, 2007.

[91] European Union Emission Trading Scheme (EU ETS). *http: //ec. europa. eu /environment /climat /emission*, 2009.

[92] Renewable Portfolio Standards. *http: //apps1. eere. energy. gov/ states/ maps/ renewable_portfolio_states.cfm*, 2009.

ABOUT THE AUTHOR

Vishnuteja (Vishnu) Nanduri, received his Ph.D. in Industrial Engineering in 2009 and M.S. in Industrial Engineering in 2005, both from University of South Florida. He received a B.E. in Industrial Engineering & Management from Visweswariah Technological University, India, in 2002. Currently, he is an Assistant Professor in the Department of Industrial & Manufacturing Engineering at the University of Wisconsin-Milwaukee.

His Master's Thesis won the First place in the Institute of Industrial Engineer's (IIE) Graduate Research Award national competition. He won the Best paper award in the Energy sponsored sessions of Institute for Operations Research and Management Sciences (INFORMS) at Washington D.C. Between 2005-2009 he served as the Project Manager for the NSF funded GK-12 program at USF (Students, Teachers, and Resources in the Sciences, STARS).

His methodological areas of interest are reinforcement learning, game-theoretic modeling, stochastic optimization, and applied probability. Major areas of application of his research are energy and the environment.