Optimization
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gopt20

Two-stage stochastic equilibrium problems with equilibrium constraints: modelling and numerical schemes

Dali Zhang [a] & Huifu Xu [b]
[a] SingHealth Centre for Health Services Research, 128802, Bukit Merah Central, Singapore
[b] School of Mathematics, University of Southampton, S017 1BJ, Southampton, UK

Available online: 02 Dec 2011

Optimization
2011, 1–24, iFirst

Taylor & Francis
Taylor & Francis Group

Two-stage stochastic equilibrium problems with equilibrium constraints: modelling and numerical schemes

Dali Zhang[a]* and Huifu Xu[b]

[a]*SingHealth Centre for Health Services Research, 128802, Bukit Merah Central, Singapore;* [b]*School of Mathematics, University of Southampton, S017 1BJ, Southampton, UK*

(*Received 11 August 2010; final version received 7 October 2011*)

This article presents a two-stage stochastic equilibrium problem with equilibrium constraints (SEPEC) model. Some source problems which motivate the model are discussed. Monte Carlo sampling method is applied to solve the SEPEC. Convergence analysis on the statistical estimators of Nash equilibria and Nash stationary points are presented.

Keywords: stochastic equilibrium programs with equilibrium constraints; convergence analysis; sample average approximation; Nash stationary point

AMS Subject Classifications: 90C15; 91A15; 90C33

1. Introduction

In our earlier work [46], we discussed a one-stage stochastic Nash equilibrium model and investigated sample average approximation (SAA) of Nash equilibrium and Nash stationary points. We noted that the model may cover two-stage stochastic Nash equilibrium problem and included an application of a two-stage stochastic equilibrium program with equilibrium constraint (SEPEC) model to study competition of generators in the electricity wholesale markets with network constraints. However, the work does not explore the unique structure and characteristics of the SEPEC model. On the other hand, the SEPEC model, as a natural extension of deterministic equilibrium program with equilibrium constraints (EPEC) models, has a number of potential applications in a wide domain of engineering design, management and economics. This motivates us to write this article in an attempt to provide an independent discussion of the model and yet not overlap with our earlier work.

Let us start with some literature review. Over the past few years, deterministic EPEC and SEPEC have been developed as a new subject in optimization primarily driven by a number of practical applications particularly in deregulated electricity industry. For instances, Hobbs et al. [15] investigated an oligopolistic electricity

*Corresponding author. Email: dali.zhang.ustc@gmail.com

ISSN 0233–1934 print/ISSN 1029–4945 online

http://dx.doi.org/10.1080/02331934.2011.632418
http://www.tandfonline.com

market with several dominant generators located in an electric power network. They developed a deterministic mathematical program with equilibrium constraints (MPEC) model to study a dominant generator's optimization problem: the first-level variables consist of the generator's bids and the second-level problem is the independent system operator (ISO)'s single commodity spatial price equilibrium problem including transmission constraints. Moreover, in a game theoretic context, the problem with multiple dominant generators are described as a Nash game with equilibrium constraints which each generator solves an MPEC. Hu and Ralph [16] used EPEC to model bilevel games in a restructured electricity market where each player's decision making is a bilevel optimization problem; Yao et al. [48] seem first to use SEPEC to model generator's strategic behaviours in a spot market with two settlement system and network constraints in USA where the stochastic model is used to reflect a day-head demand uncertainty and the second-stage equilibrium constraint is used to describe the ISO's optimal decision on generators' dispatch and power flow in the network. Henrion and Römisch [13] considered a spot market competition model where generators submit their cost functions (which do not necessarily reflect true costs) to an ISO and the ISO determine the dispatch and flow of power by minimizing the total costs subject to transmission network constraints. They developed an SEPEC model where each generator's decision-making problem is formulated as a two-stage MPEC. The equilibrium constraints describe the dispatch and power flow at the second-stage when market demand is realized. A remarkable contribution of this work is that the authors presented a detailed characterization of first-order optimality conditions of the stochastic MPEC (SMPEC) in terms of Mordukhovich's coderivatives and hence M-stationarity of the problem. Surowiec [39] took it further in his PhD dissertation to explore the structural properties and explicit stationarity conditions of SEPEC models and made a number of interesting observations relating to the numerical solution of the problem and the two-stage stochastic games. Recently, Ehrenmann and Neuhoff [9] compared two market designs, the integrated market design and the coordinated transmission auction for electricity trade and transmissions. From the mathematical perspective, the authors showed that the integrated market design is an instance of an EPEC where generators know that their output decisions will influence the allocation of transmission rights by the ISO, and the model can be represented by a Stackelberg model. In a slightly different direction, Zhang et al. [49] developed a two-stage SEPEC model to study generator's competition in electricity forward and spot markets and their interactions.

Apart from applications in electricity markets, there emerges a trend of using EPEC and SEPEC to characterize two-stage games in some general oligopoly markets where a set of strategic firms or agents (called leaders) compete in a non-cooperative manner to optimize their expected objective function anticipating the reaction of the remaining nonstrategic firms or investors (called followers). Pang and Fukushima [25] recently proposed an iterative penalty method for solving a generalized Nash equilibrium where each player solves an MPEC, and introduced a class of remedial models for the multi-leader–follower games for the oligopolistic competition models in electric power markets. For a typical two-stage competition in the stochastic environment, DeMiguel and Xu [5] developed a stochastic multiple–leader Stackelberg–Nash–Cournot (SMS) model for a homogeneous product (or service) supply market. They discussed the existence and uniqueness of the SMS equilibrium and proposed a numerical procedure to solve the problem. More broadly, for the transportation systems, EPEC models have been applied to analyse the competition behaviours of strategic users in the traffic network by Yang, Xiao and Huang [47]. In a recent work, Koh [18] investigated the potential for implicit collusions between users in the traffic network, and obtained an EPEC model where players' decision problems are constrained by a variational inequality. EPEC models have also been used for a game-theoretic analysis of the implications of overlay networks traffic of internet service providers (ISP) in Wang et al. [41]. Along with the increasing interests of modelling issues on EPEC, there synchronically emerges a volume of literature on employing the mathematical programming and game-theoretic analysis to investigate the behaviours of players and the properties of equilibrium in EPEC models. One of the natural questions arising from the EPEC problems is on the existence of Nash equilibrium of these two-stage problems, where related results have been well-established for some cases with particularly structured objective functions, see [5,25,40,49] for a set of two-stage equilibrium problems. However, it is well-known that, for some general cases, EPECs may not have any global Nash equilibrium. Instead, several alternatives of *global Nash equilibrium* are introduced for describing players' strategic behaviours within a local feasible set. Hu and Ralph [16] introduced a set of new concepts as *local Nash equilibria* and *Nash stationary points* for EPEC for a bilevel noncooperative game-theoretic model of electricity markets with locational marginal prices. On the other hand, within the framework of Mordukhovich coderivatives, Mordukhovich [22,23] first investigated the necessary optimality conditions of *M-stationary points* for EPECs. Moreover, Outrata [24] addressed a set of necessary conditions on the stationary points and the local equilibria of EPEC models in the term of coderivatives.

In this article, we are concerned with the numerical methods for solving a general two-stage SEPEC. We apply the well-known Monte Carlo sampling method to the SEPEC and analyse in detail asymptotic convergence of statistical estimators of Nash equilibria and Nash stationary points obtained from solving the SAA problems. The rest of this article is organized as follows. In the following section, we present a detailed discussion on a general SMPEC model. In Section 3, we present some source problems which motivate the SEPEC model. In Section 4, we apply the SAA of our problem and carry out convergence analysis of the estimators of Nash equilibria and Nash-Clarke-stationary points. Some concluding remarks are given in Section 5.

Throughout this article, we will use the following notation. All vectors are thought as column vectors and ^{T} denotes the transpose operation. For x, $y \in \mathbb{R}^s$, $x^T y$ denotes the scalar products of two vectors x and y and $\|x\|$ denotes the Euclidean norm. When $D \subset \mathbb{R}^s$ is a nonempty compact set of vectors, we use the notation $\|\cdot\|$ to denote $\|D\| := \max_{x \in D} \|x\|$. Moreover, $d(x, D) := \inf_{x' \in D} \|x - x'\|$ denotes the distance from point x to set D. For two compact sets D_1 and D_2,

$$\mathbb{D}(D_1, D_2) := \sup_{x \in D_1} d(x, D_2)$$

denotes the deviation from set D_1 to set D_2 (in some references, e.g., [14] it is also called *excess* of D_1 over D_2), and $\mathbb{H}(D_1, D_2)$ denotes the Hausdorff distance between the two sets, that is,

$$\mathbb{H}(D_1, D_2) := \max(\mathbb{D}(D_1, D_2), \mathbb{D}(D_2, D_1)).$$

Moreover, we use $D_1 + D_2$ to denote the Minkowski addition of D_1 and D_2, that is, $D_1 + D_2 = \{x + y : x \in D_1, \ y \in D_2\}$. We also use $B(x, \delta)$ to denote the closed ball with radius δ and centre x, that is $B(x, \delta) := \{x' : \|x' - x\| \le \delta\}$. When δ is dropped, $B(x)$ represents a neighbourhood of point x. We also use $\mathcal{B}$ to denote the unit ball in a finite-dimensional space.

2. The model

Let X_i, $i = 1, 2, \ldots, M$, be a nonempty, closed and convex subset of $\mathbb{R}^{m_i}$ and $X_{-i} := X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_M$ denote the Cartesian product of the sets except X_i. Let $X = X_i \times X_{-i}$. We consider the following *SEPECs*: find $x := (x_1, x_2, \ldots, x_M) \in X$ and $y(\cdot)$ such that for $i = 1, 2, \ldots, M$, $(x_i, y(\cdot))$ solves the following problem:

$$\begin{array}{ll} \min\limits_{x_i \in X_i, y(\cdot)} & \mathbb{E}[f_i(x_i, x_{-i}, y(\omega), \xi(\omega))] \\ \text{s.t.} & 0 \in H(x, y(\omega), \xi(\omega)) + \mathcal{N}_Q(y(\omega)), \quad \text{a.e. } \omega \in \Omega, \end{array} \tag{1}$$

where $x_{-i} \in X_{-i}$, $f_i : \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_M} \times \mathbb{R}^k \times \mathbb{R}^d \to \mathbb{R}$ is a Lipschitz continuous function but not necessarily continuously differentiable, $\xi : \Omega \to \Xi \subset \mathbb{R}^d$ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$ with support set Ξ, and $\mathbb{E}[\cdot]$ denotes the mathematical expectation with respect to (w.r.t.) the distribution of ξ. The equilibrium constraint in (1) is represented by a parametric variational inequality problem (VIP), where $y(\cdot)$ is the prime variable, and x and $\xi(\omega)$ are treated as parameters, $H : \mathbb{R}^m \times \mathbb{R}^k \times \mathbb{R}^d \to \mathbb{R}^k$ is a vector-valued continuous function, Q is a nonempty, convex and closed subset of $\mathbb{R}^k$, and $\mathcal{N}_Q(y)$ is the normal cone to Q at y, which is defined in as follows:

$$\mathcal{N}_Q(y) := \begin{cases} \{\eta \in \mathbb{R}^k : \eta^T(\bar{y} - y) \le 0 \ \forall \bar{y} \in Q\}, & \text{if } y \in Q \\ \emptyset, & \text{if } y \notin Q. \end{cases} \tag{2}$$

Problem (1) is a *two-stage SEPEC*: at the first-stage, decision maker/player i, $i \in \{1, \ldots, M\}$, chooses an optimal value $x_i \in X_i$ that maximizes the expected value of f_i under Nash conjecture (for fixed $x_{-i} \in X_{-i}$). At the second-stage for a given x and a realization of the random vector ξ, decision maker i finds an optimal $y(x, \xi)$ that solves the following optimization problem,

$$\begin{array}{l} \min\limits_{y \in Q} f_i(x_i, x_{-i}, y, \xi) \\ \text{s.t. } 0 \in H(x, y, \xi) + \mathcal{N}_Q(y). \end{array} \tag{3}$$

To ease the notation, in some parts of this article we will write $\xi(\omega)$ as ξ and the context will make it clear when ξ should be interpreted as a deterministic vector. Let us use $v_i(x_i, x_{-i}, \xi)$ to denote the optimal value function of the second-stage problem (3). Then, under some moderate conditions, (1) can be written in an *implicit* form as

$$\min_{x_i \in X_i} \mathbb{E}[v_i(x_i, x_{-i}, \xi(\omega))], \tag{4}$$

where 'implicit' means that (4) does not include details of the second-stage problem.

2.1. *A discussion of the SEPEC model*

Let $Y(x, \xi)$ denote the set of solutions to the VIP

$$0 \in H(x, y, \xi) + \mathcal{N}_Q(y). \tag{5}$$

Then we can rewrite the second-stage optimization problem (3) as

$$\min_{y \in Y(x,\xi)} f_i(x_i, x_{-i}, y, \xi). \tag{6}$$

At this point, it might be helpful to give a practical interpretation of problem (6): here, an optimal (in term of '$\min_{y \in Y(x,\xi)}$') equilibrium from $Y(x, \xi)$ is not up to decision maker i to choose. The mathematical formulation (6) only represents decision maker i's *optimistic* attitude towards a possible equilibrium outcome at scenario ξ, in other words, decision maker i anticipates a best equilibrium outcome which minimizes its objective function $f_i(x_i, x_{-i}, y, \xi)$ at scenario ξ. To properly define $v_i(x_i, x_{-i}, \xi)$ mathematically, we let $v_i(x_i, x_{-i}, \xi) = +\infty$ if the corresponding equilibrium at the second-stage does not exist, i.e. the corresponding variational inequality constraint in (3) does not have any solution. Then, given that the rivals' decisions are fixed at x_{-i}, decision maker i's expected profit at the first-stage can be formulated as

$$\hat{\vartheta}_i(x_i, x_{-i}) := \mathbb{E}[v_i(x_i, x_{-i}, \xi)],$$

or equivalently

$$\hat{\vartheta}_i(x_i, x_{-i}) = \mathbb{E}\left[\min_{y(\omega) \in Y(x,\xi(\omega))} f_i(x_i, x_{-i}, y(\omega), \xi(\omega))\right]. \tag{7}$$

If all decision makers are optimistic, then the SEPEC problem can be formulated as follows:

$$\min_{x_i \in X_i} \mathbb{E}\left[\min_{y \in Y(x,\xi(\omega))} f_i(x_i, x_{-i}, y, \xi(\omega))\right]. \tag{8}$$

Under some moderate conditions, (8) coincides with (1), see Proposition 5 in [32, Chapter 1].

Let us now consider an opposite case when decision maker i is pessimistic. In such a case, his second-stage decision problem can be formulated as

$$\max_{y \in Y(x,\xi)} f_i(x_i, x_{-i}, y, \xi), \tag{9}$$

which means that, in making his decision for minimizing $\mathbb{E}[f_i(x_i, x_{-i}, y, \xi)]$ at the first-stage, decision maker i expects a worst second-stage equilibrium outcome

$y \in Y(x, \xi)$ which maximizes $f_i(x_i, x_{-i}, y, \xi)$. Denote by $\check{v}_i(x_i, x_{-i}, \xi)$ the optimal value function of decision problem (9). Then decision maker i's expected objective function can be written as

$$\check{\vartheta}_i(x_i, x_{-i}) := \mathbb{E}[\check{v}_i(x_i, x_{-i}, \xi(\omega))], \tag{10}$$

where $\check{v}_i(x_i, x_{-i}, \xi) = +\infty$ if $Y(x, \xi) = \emptyset$. If all decision makers are pessimistic and try to hedge against a worst possible equilibrium at the second-stage, then the SEPEC model becomes: find $x = (x_1, \ldots, x_M)^T$ and $y(\cdot)$ such that for $i = 1, 2, \ldots, M$, $(x_i, y(\cdot))$ solves the following problem:

$$\min_{x_i \in X_i} \mathbb{E}\left[\max_{y \in Y(x,\xi)} f_i(x_i, x_{-i}, y, \xi)\right], \tag{11}$$

where each decision maker solves a min-max problem.

There could be other cases when some decision makers are optimistic while others are pessimistic or some decision makers do not really have an extreme view about the future equilibrium. In this article, we will simplify the discussion by considering the case when the equilibrium problem has a unique solution and subsequently the minimization process in (7) can be dropped.

2.2. *Uniqueness of the second-stage equilibrium*

In the remainder of this section, we discuss sufficient conditions for the existence and uniqueness of the Nash equilibrium problem (represented by VIP (5)) at the second-stage and the Lipschitz continuity of the optimal value functions $v_i(x, \xi)$ at the second-stage. These conditions are needed in deriving the first-order equilibrium conditions of the SEPEC problem (1) and its SAA (30) in Section 4.

To this end, let us look at the solution set $Y(x, \xi)$ of VIP (5):

$$0 \in H(x, y, \xi) + \mathcal{N}_Q(y), \tag{12}$$

and decision maker i's objective function $f_i(x_i, x_{-i}, y(x, \xi), \xi)$ at every scenario $\xi \in \Xi$ for $i = 1, 2, \ldots, M$. It is well-known that the uniqueness of the solution to VIP (5) is guaranteed by the *strict monotonicity* of mapping $H(x, y, \xi)$ for any $x \in X$ and a.e. $\xi \in \Xi$, which is equivalent to the strict concavity of function $R_i(y_i, y_{-i}, x, \xi)$ w.r.t. y_i for any fixed y_{-i} and $i = 1, 2, \ldots, M$. The assumption below presents the sufficient conditions as such.

Assumption 2.1 *For $i = 1, 2, \ldots, M$ and almost every $\xi \in \Xi$,*

(a) *$H(x, y, \xi)$ is a Lipschitz continuous function of (x, ξ) on $X \times \Xi$ with a Lipschitz constant independent of y.*
(b) *$H(x, \cdot, \xi)$ is uniformly strongly monotone on set Q, that is, for any given x and $\xi \in \Xi$, there exists a constant $c > 0$ such that*

$$(H(x, y', \xi) - H(x, y, \xi))^T(y' - y) \geq c\|y' - y\|^2 \quad \forall y', y \in K. \tag{13}$$

(c) *$f_i(\cdot, x_{-i}, \cdot, \xi)$ is Lipschitz continuous on $X_i \times Q$ with modulus $\kappa_i(\xi)$, that is, for all $i = 1, 2, \ldots, M$,*

$$|f_i(x_i', x_{-i}, y', \xi) - f_i(x_i, x_{-i}, y, \xi)| \leq \kappa_i(\xi)\big(\|x_i' - x_i\| + \|y' - y\|\big),$$

where $\mathbb{E}[\kappa_i(\xi)] < \infty$.

Under Assumption 2.1(b), it follows by virtue of [10, Theorem 2.3.3] that VIP (5) has a unique solution for every given x and ξ. Moreover, under Assumption 2.1(a) and (b) we can show the Lipschitzness of $v_i(\cdot, x_{-i}, \xi)$ which will be used in the asymptotic analysis of sample average approximate Nash equilibrium in Proposition 4.1.

LEMMA 2.1 *Under Assumption* 2.1, *for every fixed* $x_{-i} \in X_{-i}$ *and a.e.* $\xi \in \Xi$, $v_i(\cdot, x_{-i}, \xi)$ *is Lipschitz continuous on* X_i *with modulus* $\kappa_i'(\xi)$, *where* $\mathbb{E}[\kappa_i'(\xi)] < \infty$.

Proof Under Assumption 2.1(b), it follows from [10, Theorem 2.3.3(c)] that the variational inequality

$$0 \in H(x, y, \xi) + \mathcal{N}_Q(y)$$

has a unique solution $y(x, \xi)$ which is Lipschitz continuous on X with a constant modulus. Moreover, under Assumption 2.1(c), $F_i(x_i, x_{-i}, \cdot, \xi)$ is Lipschitz continuous on Q with modulus $\kappa_i(\xi)$, where $\mathbb{E}[\kappa_i(\xi)] < \infty$. Consequently, for $i = 1, 2, \ldots, M$, there exists a $\kappa_i'(\xi)$ such that $v_i(\cdot, x_{-i}, \xi)$ is Lipschitz continuous on X_i with modulus $\kappa_i'(\xi)$, where $\mathbb{E}[\kappa_i'(\xi)] < \infty$. ■

3. Source problems

A number of applications of two-stage stochastic equilibrium problem with equilibrium constraints arise from a diversity of sources. In this section, we list a few examples.

3.1. *Stochastic bilevel games*

It is well-known that bilevel programming is closely related to MPEC through KKT conditions at the second-stage. It is therefore no surprise that stochastic bilevel games provide rich problem sources for the SEPEC. Consider a two-stage stochastic Nash game which consists two sets of players: a set of M players who compete at the first-stage and a set of K players who compete at the second-stage where the decisions of players at the first-stage are disclosed and exterior uncertainty (such as market demand) is realized. Mathematically, we can formulate this kind of game as follows:

$$\begin{cases} \min\limits_{x_i \in X_i, y(\cdot)} \mathbb{E}[f_i(x_i, x_{-i}, y(x, \xi), \xi)] \\ y_j(x, \xi) \text{ solves } \min\limits_{y_j \in Q_j} R_j(y_j, y_{-j}, x, \xi), \quad \text{for } j = 1, \ldots, K, \text{ a.e. } \xi \in \Xi. \end{cases} \tag{14}$$

The stochastic multiple leader–followers game investigated by DeMiguel and Xu [5] is a typical example of this kind of two-stage stochastic bilevel game.

Assuming that for $j = 1, 2, \ldots, K$, function $R_j : \mathbb{R}^{k_j} \times \mathbb{R}^{k-k_j} \times \mathbb{R}^m \times \Xi \to \mathbb{R}$ is continuously differentiable w.r.t. y_j on a nonempty convex and closed subset $Q_j \subset \mathbb{R}^{k_j}$, $Q_{-j} := Q_1 \times \cdots \times Q_{j-1} \times Q_{j+1} \times \cdots \times Q_K$, and $k = \sum_{j=1}^K k_j$, then we can characterize the optimality condition of each player at the second-stage through a generalized equation and combining them gives

$$0 \in \nabla_y R(x, y, \xi) + \mathcal{N}_Q(y), \tag{15}$$

where $\nabla_y R(x, y, \xi) := (\nabla_{y_1} R_1(y_1, y_{-1}, x, \xi), \ldots, \nabla_{y_K} R_K(y_K, y_{-K}, x, \xi))^T$. Under some convexity conditions of the objective functions, it is well-known that a solution $y^*(x, \xi)$ is a Nash equilibrium at the second-stage if and only if it is a solution to the generalized equation (15). Consequently we can reformulate the stochastic bilevel game (14) as the following two-stage SEPEC:

$$\begin{aligned} &\min_{x_i \in X_i} \ \mathbb{E}[f_i(x_i, x_{-i}, y(x, \xi), \xi)] \\ &\text{s.t. } 0 \in \nabla_y R(x, y(x, \xi), \xi) + \mathcal{N}_Q(y(x, \xi)), \quad \text{a.e. } \xi \in \Xi. \end{aligned}$$

Note that the two-stage stochastic bilevel game model (14) may cover the capacity expansion model considered by Gürkan and Pang [11] where game at the first-stage is viewed as a competition on a long-term capacity investment at present and at the second-stage as a short-term competition in future once the capacity expansion is completed and exterior uncertainty is realized. In this case players at two-stages may be identical. Note also that (14) can also be used to model competition in forward-spot electricity markets where players compete at the first-stage in the forward market for long-term contracts and then compete for dispatch at spot markets on daily basis, see [49].

3.2. *Capital tax competition*

The enforcement of an effective taxation on savings income has been a long-standing issue both in policy and in academic debates, see [7,8]. Along with the globalization of the capital market, the tax can be easily evaded if the residence country is unable to monitor the investors' foreign interest incomes, where the countries are linked through perfect capital mobility. These capital links between countries may result in a very complex investment network. Even for a two country economy the flows of real and financial capital might induce a complex system of transactions under the different tax structures of both countries. In this section, we consider a capital tax competition between the national tax authorities in two countries, denoted by i and j, respectively.

The analysis of this capital tax competition employs a two-stage stochastic equilibrium model. At the first-stage, we assume that each tax authority of country, i or j, has three different tax instruments. We give the tax instruments set by tax authority of country i (tax authority i) for example:

(a) The first tax instrument is a *wage tax rate* t_i^w at which it taxes wage income $w_i l_i$. Note that, in most of practical capital markets, wage rate w_i and labour supply l_i may be affected by the different amount of total investment level in country i, and hence we can rewrite them as $w_i(s^i)$ and $l_i(s^i)$, where s^i denotes the amount of capital invested in country i. Here, we set s^i being the sum of s_i^i, the amount of capital invested in country i by residents at home (i.e. country i), and s_j^i invested by the investors from abroad (i.e. country j). Moreover, the wage rate in a country is usually set for a long-term and hence independent of the global economic scenario ξ. On the other hand, the labour supply level fluctuates along with the change of economic environment and hence we assume that it is a function of scenario ξ, denoted by $l_i(s^i, \xi)$.

(b) The second tax instrument the authority might choose is on the capital income of residents where we denote its rate by t_i^r. By assuming the perfect information sharing between the tax authorities of the two countries, then the tax base can be formulated as

$$R(\xi)s_i = R(\xi)(s_i^i + s_i^j), \tag{16}$$

where $R(\xi)$ is the global return rate to the investment, that is, the return for every unit capital invested in either countries, and is varied by the random shock ξ in the capital market. Since in this section, we focus our investigation on the taxation problem, we generally assume that the global return rate for the investment in each country are the same. In (16), s_i is the amount of capital invested by residents at country i and is the sum of s_i^i and s_i^j where s_i^j denotes the amount of capital invested by the residents at country i into the market in country j.

(c) Third, a government may tax the capital income generated at home on a source basis, t_i^s, where the tax base can be calculated as

$$R(\xi)s^i = R(\xi)(s_i^i + s_j^i), \tag{17}$$

which includes the return of the investment by the residents at country i and from abroad, i.e. country j. Similarly as in country i, we denote the tax instruments set by the tax authority j as t_j^w, t_j^r and t_j^s, respectively.

Differing from the discussion in [7,8], in the model, we first consider the strategic behaviours of the representative investor in each country. The representative investor (or consumer) in country i maximizes a well-behaved utility function $\vartheta_i(c_{i1}, c_{i2}, l_i, \xi)$, where c_{i1} and c_{i2} are the consumption levels before and after the investment period. Denote the endowment obtained by the investor by e_i. After the realization of uncertainty ξ in the capital market, the investor in country i needs to decide the proportion of e_i to be consumed, c_{i1}, or saved, $s_i = s_i^i + s_i^j$, where $s_i = e_i - c_{i1}$. Because the decision on c_{i1} is made after knowing ξ, consumption level c_{i1} is affected by the uncertainty in the capital market and hence can be taken as a random function of ξ, which is implied by the fact that the consumption level of investors i at country i fluctuates as a response to different economic situations in the capital market. Moreover, c_{i2} denotes the consumption level after the return of the investment and can be formulated as

$$\begin{aligned} c_{i2}(s_i^i, s_i^j, s_j^i, \xi; t_i^r, t_i^s, t_j^s) = w_i(s_i)l_i\left(s_i^i + s_j^i, \xi\right) + \left[1 + R(\xi)\left(1 - t_i^r - t_i^s\right)\right]s_i^i \\ + \left[1 + R(\xi)\left(1 - t_i^r - t_j^s\right)\right]s_i^j, \end{aligned} \tag{18}$$

where the three terms on the right-hand side of (18) are the wage income, the post-tax income from the home investment and the post-tax income from the abroad investment, respectively. Note that, in (18), $t_i^r + t_i^s$ and $t_i^r + t_j^s$ are the effective tax paid by the presentative investor at country i on its capital income from country i and country j, respectively. By incorporating c_{i1} and c_{i2} into the investor's utility function, for a realized market scenario ξ, the decision problem of the representative investor in country i is to choose the amounts of $s_i^i(\xi)$ and $s_i^j(\xi)$ to maximize their utility function. Assuming that the realization of the uncertainty in the capital market is ξ and the investor in country j rationally fix their optimal investment at $(s_j^j(\xi), s_j^i(\xi))$, the representative investor in country i determines its investment policy by solving the following problem,

$$\max_{s_i^i \in S_i^i, s_i^j \in S_i^j} \vartheta_i\Big(e_i - s_i^i - s_i^j, c_{i2}(s_i^i, s_i^j, s_j^i, \xi;\, t_i^r, t_i^s, t_j^s), l_i\Big(s_i^i + s_j^i, \xi\Big), \xi\Big), \tag{19}$$

where $S_j^i := [0, \bar{s}_i^i]$ and $S_i^j := [0, \bar{s}_i^j]$, and $\bar{s}_i^i$ and $\bar{s}_i^j$ are the upper bounds of investments s_i^i and s_i^j. By assuming the convexities of functions $l_i(\cdot, \xi)$ and $\vartheta_i(c_{i1}, c_{i2}, l_i, \xi)$, we can show the existence and uniqueness of the investment equilibrium of the investors' competition in the capital market at almost every scenario ξ.

Consequently, the second-stage equilibrium problem is: for tax instruments t_i and t_j fixed at the first-stage and the realization ξ, find an equilibrium (s_i, s_j) solves the following parametric equilibrium problem,

$$\max_{s_k \in S_k} \vartheta_k\big(e_k - s_k, c_{k2}(s_k^k, s_k^{k'}, s_{k'}^k, \xi;\, t_k^r, t_k^s, t_{k'}^s), l_k\big(s^k, \xi\big), \xi\big), \tag{20}$$

where $s_k \in S_k = S_k^k \times S_k^{k'}$, $k = i, j$ and $k' \neq k \in \{i, j\}$. In (20), $t := (t_i, t_j)$ and ξ are treated as parameters. Therefore, the competition between the representative investors at country i and j can be taken as a Cournot-type game, and the equilibrium to problem (20) is a function of t and ξ which can be specified as $s_k(t, \xi)$ for $k = i$ and j. Then, we can rewrite (19) as the following general equation form,

$$0 \in H(s(t, \xi), t, \xi) + \mathcal{N}_S(s(t, \xi)), \tag{21}$$

where $s(t, \xi) = (s_i(t, \xi),\ s_j(t, \xi))$, and feasible set $S := S_i \times S_j$.

Assuming the perfect information sharing between the two countries, we have that each national authority determines its tax rates by aiming at the maximization of its expected utility function which consists of two parts: one is the expected return of its representative investor in the capital market, and the other is the production capacity of country i, denoted by g_i, which is seen as a function of the total tax revenue in the country. Consequently, at the first-stage, given that tax authority j's optimal tax policy is rationally fixed at t_j^*, the decision problem of tax authority i can be formulated as

$$\begin{aligned} &\max_{t_i}\ \mathbb{E}[\eta_i(t_i, t_j^*, s_i(t_i, t_j^*, \xi), s_j(t_i, t_j^*, \xi), \xi)], \\ &\text{s.t. } t_i^w \in [0, \bar{t}_i^w],\ t_i^r \in [0, \bar{t}_i^r],\ t_i^s \in [0, \bar{t}_i^s]. \end{aligned} \tag{22}$$

where the objective function $\eta_i(t_i, t_j^*, s_i, s_j, \xi)$ of tax authority i is defined as

$$\begin{aligned} \eta_i\big(t_i, t_j, s_i, s_j, \xi\big) := \vartheta_i\Big(&e_i - s_i, c_{i2}(s_i^i, s_i^j, s_j^i, \xi;\, t_i^r, t_i^s, t_j^s), l_i\big(s^i, \xi\big), \xi\Big) \\ &+ u\big(g_i\big(t_i, t_j, s_i, s_j, \xi\big)\big), \end{aligned} \tag{23}$$

$u(\cdot)$ is the utility function of the production capacity g_i, and the production capacity g_i is a function of the tax revenues as follows:

$$g_i\big(t_i, t_j, s_i, s_j, \xi\big) = t_i^w w_i(\xi) l_i\Big(s_i^i + s_j^i, \xi\Big) + \big(t_i^s + t_i^r\big) R(\xi) s_i^i + t_i^r R(\xi) s_i^j + t_i^s R(\xi) s_j^i.$$

Moreover, in decision problem (24), variables $s_i(t_i, t_j, \xi)$ and $s_j(t_i, t_j, \xi)$ are solved from general equation (21) for any fixed t_i, t_j and the realization ξ. By inclusively taking the investors' reactions at the second-stage into consideration, tax authority *i*'s decision problem can be written as,

$$
\begin{aligned}
&\max_{t_i}\ \mathbb{E}[\eta_i(t_i, t_j^*, s_i(t_i, t_j, \xi), s_j(t_i, t_j, \xi), \xi)],\\
&\text{s.t. } t_i^w \in [0, \bar{t}_i^w],\ t_i^r \in [0, \bar{t}_i^r],\ t_i^s \in [0, \bar{t}_i^s],\\
&\qquad 0 \in H(s(t,\xi), t, \xi) + \mathcal{N}_S(s(t,\xi)), \quad \text{a.e. } \xi \in \Xi,
\end{aligned}
\tag{24}
$$

which implies that the capital tax competition can be formulated as an SEPEC model.

3.3. *Oligopolistic transit market*

In this section, we look at a two-stage stochastic equilibrium problem for a urban transit systems. Over the past twenty years, the deregulation of *urban transit systems* has become an appealing alternative to centralized municipal transit policy. In a recent article [50], a deterministic network equilibrium model with a two-stage framework for a deregulated transit system is proposed to describe the fare competition between transit operators where every operator takes into account passengers' responses in making its decision. At the first-stage, by assuming that its rivals rationally choose their optimal decisions on their fare structures, each of the transit operators can determine its own fare structure in order to maximize its expected revenue, where in the urban transit system, the transit operators' revenues depends on the number of passengers using their lines. Then, at the second-stage, every passenger reacts to the transit operators' fares in the urban transit system.

Stochastic network equilibrium models are widely applied for predicting traffic patterns in the transportation networks at the second-stage, in which the traffic flow in the transportation network is characterized by *stochastic user equilibrium* for every possible scenario. In this two-stage model, the interaction between the transit operators and the passengers is described in the form of Stackelberg game, that is, at the first-stage, in making the decision, every operator takes the passengers' reaction to its fare plan at every traffic scenario into account. On the other hand, the competition between the transit operators can be seen as a Cournot game where each operator makes its decision regarding that its rivals' fares are fixed.

In the problem, the urban transit network is denoted by a directed traffic network as $G=(N,A)$ where N is the set of nodes (or transfer stations) and A is the set of links (or route sections). At the first-stage, the transit competition is portrayed as M player (transit operator) noncooperative game of deciding the fares for a set of transfer lines connecting origin–destination (OD) pairs w for $w \in W$ where W is the set of all OD pairs, and R_w is the set of all routes joining OD pair w. In the first-stage equilibrium, transit operator i makes its decision on the fare of the route connecting OD $w \in W$, denoted by $p_w^i = \{p_r^i\}_{r \in R_w}$, so as to maximize its expected profit, where the expectation is taken w.r.t. the distribution of the traffic uncertainty $\xi \in \Xi$. In making the decision at the first-stage, each operator predictively takes into account the passenger flow in every route of the traffic network at every possible scenario ξ, where the flows are determined by solving a stochastic user equilibrium model at the second-stage.

In the second-stage problem, with each fixed transit fare structure and a realized traffic uncertainty ξ, the stochastic user equilibrium condition can be mathematically expressed to determine the flow on every route r serving an OD pair $w \in W$ for $r \in R_w$: Denote by $y_r(\xi)$ the traffic flow on path r at traffic scenario ξ. Given that the realization of traffic uncertainty is ξ and operator i's fare for route r is p_r^i for all $r \in R_w$ and $w \in W$, the user (passenger)'s travel cost function on path r can be written as

$$u_r(y, p_w, \xi) = d_r(p_w) + C_r(y, p_w, \xi) + \theta_0(t_r(y, \xi) - \tau_w)^2, \tag{25}$$

where $y := \{y_r\}_{r \in \mathcal{R}_w}$, $p_w := (p_w^1, \ldots, p_w^M)^T$, $p_w^i = \{p_r^i\}_{r \in R_w}$. In (25), $d_r(p_w)$ represents the composite of attributes such as the travel fare for a certain distance which is independent of time/flow, $C_r(y, p_w, \xi)$ denotes the stochastic travel cost on path r which is implicitly determined by the flows on all arc r which may depends on the volume of the passenger flow between OD pair w, the fares of all operators, and changes for different traffic scenario. $t_r(y, \xi)$ denotes the stochastic travel time on path r and τ_w denote the expected travel time between OD pair w. Furthermore, in (25), θ_0 is the penalty coefficient when the actual travel time on w deviates the scheduled time τ_w. Then, by assuming the total population travelling between OD pair w is q_w being a deterministic parameter, the feasible set of the traffic flow across the whole network can be expressed as

$$Q = \left\{ y : \sum_{r \in \mathcal{R}_w} y_r = q_w \ \ \forall w, y_r \geq 0, \ \forall r \right\},$$

which is a convex set. Moreover, at the second-stage, the stochastic user equilibrium is defined as the state when no passenger believes that he can reduce his perceived travel cost by changing route unilaterally. Hence, we can write this equilibrium condition as following VIP: find y^* such that

$$\begin{cases} y^*(u(y, p, \xi) - u(y^*, p, \xi)) = 0 \ \forall y \in Q, \ \ w \in W \\ y^* \geq 0, \ \ u(y, p, \xi) - u(y^*, p, \xi) \geq 0, \end{cases} \tag{26}$$

for almost every $\xi \in \Xi$ and fixed fare $p = \{p_w^i, \ \ i = 1, 2, \ldots, M, \ \ w \in W\}$, where

$$y^*(u(y, p, \xi) - u(y^*, p, \xi)) := \sum_{r \in R_w} y_r^*(u_r(y, p, \xi) - u_r(y^*, p, \xi)),$$

and the fare p and the traffic scenario ξ are treated as parameters. VIP (26) can be equivalently rewritten in the following parametric generalized equation:

$$0 \in H(p, y, \xi) + \mathcal{N}_Q(y), \quad \text{a.e. } \xi \in \Xi, \tag{27}$$

for a vector-valued function $H(p, y(p, \xi), \xi)$ and feasible set Q of y. Usually, functions $C_r(y, p_j, \xi)$ and $t_r(y, \xi)$ are assumed to be continuously differentiable for every ξ, and hence the cost function $u_r(y, p_j, \xi)$ is a continuously differentiable function of y and $H(p, y, \xi)$ is also a continuous single-valued mapping. Hence, solution y to parametric problem (27) can be written as $y(p, \xi)$ which is a function of fare p and traffic scenario ξ.

Let us step back to the operators' problems at the first-stage, in which every operator makes its decision on the fare charged at each route to maximize its expected revenue. From the discussion on stochastic user equilibrium, we have that passenger flow $y_r(p, \xi)$ on path r can be implicitly solved by (26) or (27), and is a continuous function of p for every scenario ξ. Then, we assume that the proportion that passengers choose operator i's service on path $r \in R_w$ for travelling between OD pair w is a function of $f_w(y, p_w^i, p_w^{-i})$ where p_w^{-i} are the fare structures provided by operator i's rivals between OD pair w. In [50], these proportions are calculated according to the well-used logit-type model, and in a more general case, we might assume that $f_w(y(p,\xi), p_w^i, p_w^{-i}, \xi)$ is a continuously differentiable function of y, p_w^i and p_w^{-i} for all OD pair w, where this proportion only reflects a passenger's attitude towards each operator's fare. Consequently, with this proportion, we can estimate the number of passengers using operator i's service as $v_w^i = f_w(y(p,\xi), p_w^i, p_w^{-i}, \xi)q_w$. Then operator i's expected revenue function can be written as

$$R_i(p^i, p^{-i}) = \mathbb{E}\left[\sum_{w\in W} f_w\big(y(p_w^i, p_w^{-i}, \xi), p_w^i, p_w^{-i}, \xi\big) q_w p_w^i\right], \tag{28}$$

where $p^i = \{p_w^i\}$ and $p^{-i} = \{p_w^{-i}\}$ for all $w \in W$, and $\mathbb{E}[\cdot]$ is taken w.r.t. the distribution of the traffic uncertainty ξ. It should be noted that the operating cost does not appear in the expression due to the assumption of fixed service frequency. Thus, the two-stage equilibrium problem is: find $p = (p^1, \ldots, p^M)$ and $y(\cdot)$ such that for $i = 1, 2, \ldots, M$, $(p^i, y(\cdot))$ solves the following problem:

$$\begin{aligned} \min_{p_w^i \in [0, \bar{p}_w^i],\, y(\cdot)} \quad & \mathbb{E}\left[\sum_{w\in W} f_w\big(y(p_w^i, p_w^{-i}, \xi), p_w^i, p_w^{-i}, \xi\big) q_w p_w^i\right] \\ \text{s.t.}\ & 0 \in H(p, y(p,\xi), \xi) + \mathcal{N}_Q(y(p,\xi)), \quad \text{a.e. } \xi \in \Xi, \end{aligned} \tag{29}$$

where $\bar{p}_w^i$ is the ceiling of operator i's fare p_w^i on its route between OD w. From problem (29), we have that the competition in the oligopolistic transit market can be modelled by an SEPEC problem. In [50], a deterministic version of EPEC model is proposed for investigating the competition in a deregulated transit network market.

Apart from the applications of the two-stage SEPEC models discussed in this section, there are some potential applications in transportation and economics [13,16,22,29], internet service problems [41] and airline revenue management problems [17].

4. Sample average approximation

In this section, we discuss a numerical method for solving the SEPEC problem (1). If the random vector ξ has a finite discrete distribution and the distribution is known, then the problem can be easily formulated as a deterministic EPEC for which existing numerical methods may be readily applied to solve it [19,37]. To cover a broader spectrum of practical applications, here we assume that ξ satisfies a general distribution which could be continuous, and it is impossible to obtain a closed form of $\mathbb{E}[f_i(x_i, x_{-i}, y, \xi)]$ either because it is computationally too expensive or the distribution function is unknown. However, it might be possible to obtain samples of ξ from past data or computer simulation, and a particular numerical scheme we are looking at here is the SAA. Let $\xi^1, \ldots, \xi^N$ be an independent and identically distributed (iid) sampling of the random vector $\xi(\omega)$. We consider the following SAA of problem: find $x^N := (x_1^N, x_2^N, \ldots, x_M^N)^T \in X_1 \times X_2 \times \cdots \times X_M$ such that

$$\min_{x_i \in X_i} \frac{1}{N}\sum_{n=1}^{N} f_i(x_i, x_{-i}^N, y^n, \xi^n) \tag{30}$$

where y^n, $n=1,2,\dots,N$, is a solution of the following VIP for fixed x_i and x^n_{-i}:

$$0 \in H\big((x_i, x^N_{-i}), y^n, \xi^n\big) + \mathcal{N}_Q(y^n) \tag{31}$$

$$H(x,y,\xi) := (H_1(x,y,\xi),\dots,H_K(x,y,\xi))^T$$

and

$$\mathcal{N}_Q(y) := \mathcal{N}_{Q_1}(y_1) \times \cdots \times \mathcal{N}_{Q_K}(y_K).$$

We refer to (1) as the *true* (SEPEC) problem and (30) as the *SAA* problem. Since (31) has a unique solution, we may use the notation of the optimal value function in (4) to reformulate (30)–(31) as follows: find $(x_1^N,\dots,x_M^N)^T$ such that x_i solves

$$\min_{x_i \in X_i} \hat{\vartheta}_i^N(x_i, x_{-i}) := \frac{1}{N}\sum_{n=1}^N v_i(x_i, x^N_{-i}, \xi^n). \tag{32}$$

We call (32) the SAA of the implicit Nash SEPEC (4).

SAA is a very popular method in stochastic programming, it is known under various names such as Monte Carlo sampling, sample path optimization and stochastic counterpart, see [28,31,35] for SAA in general stochastic programming and [5,20,46] for recent applications of the method to stochastic equilibrium problems. Our focus in this section is on the convergence of SAA problems described above to their true counterparts. Specifically, if we obtain a Nash equilibrium or a Nash stationary point (to be defined shortly), denoted by x^N, from solving (30), we investigate the convergence of x^N as sample size N increases.

Proposition 4.1 (Convergence of Nash equilibrium estimators) *Let* $\{x^N\}$ *be a sequence of Nash equilibria obtained from solving* (30) *and Assumption* 2.1 *holds. Then with probability one an accumulation point of* $\{x^N\}$ *is a Nash equilibrium of the true problem* (1).

The results depend on the Lipschitz continuity of v_i (established in Lemma 2.1) rather than the details of the second-stage equilibrium. Therefore the proposition follows straightforwardly from [46, Theorem 4.2(b)]. We omit the details.

4.1. *Nash stationary points*

It is well-known that the optimal value function of a parametric mathematical program with equilibrium constraints (MPEC) is often nonconvex. In our context, this means that $v_i(x_i, x^N_{-i}, \xi^n)$ may be nonconvex in x_i for fixed x^N_{-i} and ξ^n, and consequently we may obtain a local Nash equilibrium or a Nash stationary point from solving the SAA problem (30). The concept of stationary points are important in optimization as it provides some information of optimality. This is particularly so in MPECs where obtaining a global optimal solution is often difficult and consequently various of stationary points are investigated [13,49]. The concept of Nash stationary point is relatively new: it was introduced by Hu and Ralph [16].

We start with the definition. Based on Assumption 2.1, we have that the optimal value function v_i is usually not continuously differentiable, and the concept of the generalized gradient is needed to characterize the first-order optimality conditions.

Here we use the *Clarke generalized gradient* for the analysis which is popular and mathematically easy to handle. The Clarke generalized gradient of the optimal value function $v_i(x, y, \xi)$ w.r.t. x coincides with the usual gradient at the points where $v_i(\cdot, y, \xi)$ is strictly differentiable.

Let $v : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous function. Recall that *Clarke generalized derivative* of v at point x in direction d is defined as

$$v^o(x, d) := \limsup_{y \to x, t \downarrow 0} \frac{v(y + td) - v(y)}{t}.$$

v is said to be Clarke *regular* at x if the usual one sided directional derivative, denoted by $v'(x, d)$, exists for all $d \in \mathbb{R}^n$ and $v^o(x, d) = v'(x, d)$. The *Clarke generalized gradient* (also known as Clarke subdifferential) is defined as

$$\partial v(x) := \{\zeta : \zeta^T d \leq v^o(x, d)\},$$

see [4, Chapter 2].

In Lemma 2.1, we have shown that under some appropriate conditions $v_i(x_i, x_{-i}, \xi), i = 1, \ldots, M$, is Lipschitz continuous w.r.t. x_i with integrable Lipschitz modulus. This implies that $\mathbb{E}[v_i(x_i, x_{-i}, \xi)]$ is also Lipschitz continuous w.r.t. x_i and hence $\partial_{x_i}\mathbb{E}[v_i(x_i, x_{-i}, \xi)]$ is well-defined, see [32]. We characterize the first-order equilibrium condition of (1) at a Nash equilibrium in terms of the Clarke generalized gradients as follows:

$$0 \in \partial_{x_i}\mathbb{E}[v_i(x_i, x_{-i}, \xi)] + \mathcal{N}_{X_i}(x_i), \quad i = 1, \ldots, M. \tag{33}$$

Here and later on, the addition of the sets is in the sense of Minkowski. We call a point x^* satisfying (33) a *stochastic Nash-C-stationary point*. Under some standard constraint qualifications, a Nash equilibrium is a Nash-C-stationary point. Conversely if v_i is convex, then a Nash-C-stationary point is also a Nash equilibrium.

Let us now consider the first-order necessary equilibrium condition for the SAA problem (30) in terms of Clarke generalized gradient:

$$0 \in \partial_{x_i}\vartheta_i^N(x_i, x_{-i}) + \mathcal{N}_{X_i}(x_i), \ i = 1, \ldots, M, \tag{34}$$

where

$$\partial_{x_i}\vartheta_i^N(x_i, x_{-i}) := \partial_{x_i}\left(\frac{1}{N}\sum_{n=1}^{N} v_i(x_i, x_{-i}, \xi^n)\right).$$

We call a point $\bar{x}^N$ satisfying (34) *SAA Nash-C-stationary point*. Our objective here is to investigate the convergence SAA Nash-C-stationary point to its true counterpart.

For the simplicity of notation, we denote throughout this section the following.

$$\mathcal{A}\vartheta(x) := \partial_{x_1}\mathbb{E}[v_1(x, \xi)] \times \cdots \times \partial_{x_M}\mathbb{E}[v_M(x, \xi)] \tag{35}$$

and

$$G_X(x) := \mathcal{N}_{X_1}(x_1) \times \cdots \times \mathcal{N}_{X_M}(x_M). \tag{36}$$

The first-order equilibrium condition (33) can be written as

$$0 \in \mathcal{A}\vartheta(x) + G_X(x). \tag{37}$$

Likewise, the first-order equilibrium condition (34) can be written as

$$0 \in \mathcal{A}\vartheta^N(x) + G_X(x), \tag{38}$$

where

$$\mathcal{A}\vartheta^N(x) := \partial_{x_1}\left(\frac{1}{N}\sum_{n=1}^{N} v_1(x,\xi^n)\right) \times \cdots \times \partial_{x_M}\left(\frac{1}{N}\sum_{n=1}^{N} v_M(x,\xi^n)\right). \tag{39}$$

Let $\{\bar{x}^N\}$ be a sequence of stationary points satisfying optimality condition of the SAA problem (34) with sample size N. In what follows, we investigate the convergence of the sequence as sample size N increases. First, we need the following technical results.

Lemma 4.1 *Let $F(x,\xi): \mathbb{R}^m \times \Xi \to \mathbb{R}$ be a continuous function, and $\mathcal{X}$ be a compact subset. Assume that $F(x,\xi)$ is locally Lipschitz continuous w.r.t. x for almost every ξ with modulus $L(x,\xi)$ which is bounded by a positive constant C and $\frac{1}{\tau}(F(x+\tau h,\xi) - F(x,\xi))$ is uniformly continuous in ξ for τ sufficiently small, $\|h\| \le 1$ and $x \in \mathcal{X}$. Then*

$$\lim_{N\to\infty} \sup_{x\in\mathcal{X}} \mathbb{H}\left(\partial\left(\frac{1}{N}\sum_{n=1}^{N} F(x,\xi^n)\right), \partial\mathbb{E}[F(x,\xi)]\right) = 0. \tag{40}$$

Proof The assertion is a special case of a recently established result [21, Lemma 5.1]. We include a proof for completeness. For the simplicity of the notation, let

$$P_N := \frac{1}{N}\sum_{n=1}^{N} 1_{\xi^n}(\omega),$$

where

$$1_{\xi^n}(\omega) := \begin{cases} 1, & \text{if } \xi(\omega) = \xi^{\mathrm{n}}, \\ 0, & \text{if } \xi(\omega) \neq \xi^{\mathrm{n}}. \end{cases}$$

Then $\mathbb{E}_{P_N}[F(x,\xi)] = \frac{1}{N}\sum_{n=1}^{N} F(x,\xi^n)$ and hence

$$\partial\mathbb{E}_{P_N}[F(x,\xi)] = \partial\left(\frac{1}{N}\sum_{n=1}^{N} F(x,\xi^n)\right).$$

Let $f_{P_N}(x) = \mathbb{E}_{P_N}[F(x,\xi)]$ and $f_P(x) = \mathbb{E}[F(x,\xi)]$. Under condition (a), both $f_{P_N}(x)$ and $f_P(x)$ are globally Lipschitz continuous, therefore the Clarke's generalized derivatives of $f_{P_N}(x)$ and $f_P(x)$, denoted by $f^o_{P_N}(x;h)$ and $f^o_P(x;h)$, respectively, are well-defined for every fixed nonzero vector $h \in \mathbb{R}^m$, where

$$f^o_{P_N}(x;h) = \limsup_{x'\to x,\, \tau\downarrow 0} \frac{1}{\tau}\left(f_{P_N}(x'+\tau h) - f_{P_N}(x')\right)$$

and

$$f^o_P(x;h) = \limsup_{x'\to x,\, \tau\downarrow 0} \frac{1}{\tau}(f_P(x'+\tau h) - f_P(x')).$$

Our idea is to study the Hausdorff distance $\mathbb{H}(\partial f_{P_N}(x), \partial f_P(x))$ through certain 'distance' of the Clarke generalized derivatives $f^o_{P_N}(x;h)$ and $f^o_P(x;h)$. Let D_1, D_2 be two convex and compact subsets of $\mathbb{R}^m$. Let $\sigma(D_1,u)$ and $\sigma(D_2,u)$ denote the support functions of D_1 and D_2 respectively, where $\sigma(D_i,u)$ is defined as for $i=1$ and 2

$$\sigma(D_i,u) = \sup_{x\in D_i} u^T x,$$

for every fixed $u \in \mathbb{R}^m$. Then

$$\mathbb{D}(D_1,D_2) = \max_{\|u\|\le 1}(\sigma(D_1,u) - \sigma(D_2,u))$$

and

$$\mathbb{H}(D_1,D_2) = \max_{\|u\|\le 1} |\sigma(D_1,u) - \sigma(D_2,u)|.$$

The above relationships are known as Hörmander's formulae, see [3, Theorem II-18]. Applying the second formula to our setting, we have

$$\mathbb{H}(\partial f_{P_N}(x), \partial f_P(x)) = \sup_{\|h\|\le 1} \left|\sigma(\partial f_{P_N}(x),h) - \sigma(\partial f_{P_N}(x),h)\right|.$$

Using the relationship between Clarke's subdifferential and Clarke's generalized derivative, we have that $f^o_{P_N}(x;h) = \sigma(\partial f_{P_N}(x),h)$ and $f^o_P(x;h) = \sigma(\partial f_P(x),h)$. Consequently,

$$\begin{aligned}
&\mathbb{H}\big(\partial f_{P_N}(x), \partial f_P(x)\big)\\
&\quad = \sup_{\|h\|\le 1} \left|f^o_P(x;h) - f^o_{P_N}(x;h)\right|\\
&\quad = \sup_{\|h\|\le 1} \left|\limsup_{x'\to x,\,\tau\downarrow 0} \frac{1}{\tau}(f_P(x'+\tau h) - f_P(x')) - \limsup_{x'\to x,\,\tau\downarrow 0} \frac{1}{\tau}\big(f_{P_N}(x'+\tau h) - f_{P_N}(x')\big)\right|.
\end{aligned}$$

Note that for any bounded sequence $\{a_n\}$ and $\{b_n\}$, we have

$$\left|\limsup_{n\to\infty} a_n - \limsup_{n\to\infty} b_n\right| \le \limsup_{n\to\infty} |a_n - b_n|. \tag{41}$$

To see this, let $\{a_{n_j}\}$ be a subsequence such that $\limsup_{n\to\infty} a_n = \lim_{n_j\to\infty} a_{n_j}$. Then

$$\begin{aligned}
\limsup_{n\to\infty} |a_n - b_n| &\ge \limsup_{n_j\to\infty} |a_{n_j} - b_{n_j}|\\
&\ge \limsup_{n_j\to\infty} (a_{n_j} - b_{n_j})\\
&= \limsup_{n\to\infty} a_n + \limsup_{n_j\to\infty}(-b_{n_j})\\
&\ge \limsup_{n\to\infty} a_n + \liminf_{n_j\to\infty}(-b_{n_j})\\
&\ge \limsup_{n\to\infty} a_n + \liminf_{n\to\infty}(-b_n)\\
&= \limsup_{n\to\infty} a_n - \limsup_{n\to\infty} b_n.
\end{aligned}$$

Since a_n and b_n are in a symmetric position, we have that

$$\limsup_{n\to\infty} |a_n - b_n| \geq \limsup_{n\to\infty} b_n - \limsup_{n\to\infty} a_n.$$

This verifies (41). Using (41), we have

$$\begin{aligned}\mathbb{H}(\partial f_{P_N}(x), \partial f_P(x)) &\leq \sup_{\|h\|\leq 1} \limsup_{x'\to x,\, \tau\downarrow 0} \left|\frac{1}{\tau}(f_P(x'+\tau h) - f_P(x')) - \frac{1}{\tau}\big(f_{P_N}(x'+\tau h) - f_{P_N}(x')\big)\right| \\ &= \sup_{\|h\|\leq 1} \limsup_{x'\to x,\, \tau\downarrow 0} \left|\int_\Xi \frac{1}{\tau}(F(x'+\tau h,\xi) - F(x',\xi))d(P-P_N)(\xi)\right|.\end{aligned}$$

Since P_N converges to P in distribution, and the integrand $\frac{1}{\tau}(F(x'+\tau h,\xi) - F(x',\xi))$ is uniformly continuous w.r.t ξ and it is bounded by L, by virtue of [2, Theorem 2.1]

$$\lim_{N\to\infty} \sup_{x\in\mathcal{X}} \sup_{\|h\|\leq 1} \limsup_{x'\to x,\, \tau\downarrow 0} \left|\int_\Xi \frac{1}{\tau}(F(x'+\tau h,\xi) - F(x',\xi))d(P-P_N)(\xi)\right| = 0.$$

This completes the proof. ■

Let $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real-valued function and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ a random vector defined on probability space $(\Omega, \mathcal{F}, P)$, let $\mathcal{X}$ be a subset of $\mathbb{R}^n$ and $x\in\mathcal{X}$. Recall that ϕ is said to be *almost H-clam at x from above* with modulus $\kappa(\xi)$ and order γ if for any $\epsilon > 0$, there exist a (measurable) function $\kappa : \Xi \to \mathbb{R}_+$, positive numbers γ, δ and an open set $\Xi_\epsilon \subset \Xi$ such that

$$\text{Prob}(\xi \in \Xi_\epsilon) \leq \epsilon \tag{42}$$

and

$$\phi(x',\xi) - \phi(x,\xi) \leq \kappa(\xi)||x'-x||^\gamma,$$

for all $\xi\in\Xi\backslash\Delta_\epsilon$ and all $x'\in B(x,\delta)\cap\mathcal{X}$. The notion of almost H-calm is recently introduced by Sun and Xu [38]. It is an effective extension of H-calmness which is introduced in our earlier work [46] to cover a larger class of practically interesting random functions.

Using Lemma 4.1 and the concept of almost H-calmness, we are ready to present one of the main results in this section which state the uniform and exponential convergence of the subdifferentials of underlying functions in defining the Nash equilibrium conditions.

THEOREM 4.1 *Let $\{x^N\}$ be a sequence of SAA Nash-C-stationary points. Assume*: (a) *w.p.*1 *$\{x^N\}$ is contained in a compact subset $\mathcal{X}$ of X*, (*b*) *Assumption* 2.1 *holds*, (*c*) *the Lipschitz modulus of $v_i(x_i, x_{-i}, \xi)$ w.r.t. x_i, $i=1,\ldots,M$, is bounded by a positive constant C. Then*

(i) *w.p.*1

$$\lim_{N\to\infty} \sup_{x\in\mathcal{X}} \mathbb{H}(\mathcal{A}\vartheta^N(x), \mathcal{A}\vartheta(x)) = 0. \tag{43}$$

(ii) *Assume in addition that for every $\xi\in\Xi$ and $x_{-i}\in X_{-i}$*: (d) *the Clarke generalized directional derivative $(v_i)^o_{x_i}(x,\xi^n; u)$ is almost H-calm from above w.r.t. (x,u) and $\mathbb{E}[(v_i)^o_{x_i}(x,\xi^n; u)]$ is continuous*, (e) *$v_i(x_i, x_{-i}, \xi)$ is Clarke* *regular w.r.t. x_i for a.e. $\xi \in \Xi$, (f) the support set of ξ is bounded. Then for every small positive number $\epsilon > 0$, there exist $\hat{c}(\epsilon) > 0$ and $\hat{\beta}(\epsilon) > 0$, independent of N, such that*

$$\left\{\sup_{x\in\mathcal{X}} \mathbb{D}(\mathcal{A}\vartheta^N(x), \mathcal{A}\vartheta(x)) \geq \epsilon\right\} \;\leq\; \hat{c}(\epsilon)e^{-\hat{\beta}(\epsilon)N}, \tag{44}$$

for N sufficiently large.

Proof Part (i). Observe that

$$\mathbb{H}(\mathcal{A}\vartheta^N(x), \mathcal{A}\vartheta(x)) \leq \sum_{i=1}^{M} \mathbb{H}(\mathcal{A}\vartheta_i^N(x), \mathcal{A}\vartheta_i(x)),$$

where $\mathcal{A}\vartheta_i^N(x) = \partial_{x_i}(\frac{1}{N}\sum_{n=1}^N v_i(x,\xi^n))$ and $\mathcal{A}\vartheta_i(x) = \partial_{x_i}\mathbb{E}[v_i(x,\xi)]$. Under conditions (a)–(c), it follows by Lemma 4.1 that

$$\lim_{N\to\infty} \sup_{x\in\mathcal{X}} \mathbb{H}(\mathcal{A}\vartheta_i^N(x), \mathcal{A}\vartheta_i(x)) = 0$$

w.p.1 for $i = 1, \ldots, M$, which immediately yields (43).

Part (ii). Following the proof in [26, Proposition 3.4], we have

$$\mathbb{E}[\sigma(\mathcal{A}v(x,\xi), u)] = \sigma(\mathbb{E}[\mathcal{A}v(x,\xi)], u). \tag{45}$$

On the other hand, it is easy to verify the following inequality

$$\mathbb{D}(\mathcal{A}\vartheta^N(x), \mathcal{A}\vartheta(x)) \leq \sum_{i=1}^{M} \mathbb{D}(\mathcal{A}\vartheta_i^N(x), \mathcal{A}\vartheta_i(x)).$$

By Hörmander's formulae [3, Theorem II-18],

$$\mathbb{D}(\mathcal{A}\vartheta_i^N(x), \mathcal{A}\vartheta_i(x)) = \max_{\|u\|\leq 1}[\sigma(\mathcal{A}\vartheta_i^N(x), u) - \sigma(\mathcal{A}\vartheta_i(x), u)].$$

Since

$$\mathcal{A}\vartheta_i^N(x) \subset \frac{1}{N}\sum_{n=1}^{N} \partial_{x_i} v_i(x,\xi^n),$$

and $\sigma(\partial_{x_i} v_i(x,\xi^n), u) = (v_i)^o_{x_i}(x,\xi^n; u)$, then

$$\sigma(\mathcal{A}\vartheta_i^N(x), u) \leq \frac{1}{N}\sum_{n=1}^{N} \sigma(\partial_{x_i} v_i(x,\xi^n), u) = \frac{1}{N}\sum_{n=1}^{N} (v_i)^o_{x_i}(x,\xi^n; u),$$

Moreover, under conditions (e), it follows from [4, Theorem 2.7.2]

$$\sigma(\mathcal{A}\vartheta_i(x), u) = (\vartheta_i)^o_{x_i}(x; u) = \mathbb{E}\Big[(v_i)^o_{x_i}(x,\xi; u)\Big].$$

Consequently, we have

$$\begin{aligned}\sup_{x\in\mathcal{X}} \mathbb{D}(\mathcal{A}\vartheta^N(x), \mathcal{A}\vartheta(x)) &\leq \sup_{x\in\mathcal{X}} \sum_{i=1}^{M} \mathbb{D}(\mathcal{A}\vartheta_i^N(x), \mathcal{A}\vartheta_i(x)) \\ &\leq \sup_{x\in\mathcal{X}} \max_{\|u\|\leq 1}\left(\frac{1}{N}\sum_{n=1}^{N}(v_i)^o_{x_i}(x,\xi^n; u) - \mathbb{E}\Big[(v_i)^o_{x_i}(x,\xi; u)\Big]\right).\end{aligned}$$

In what follows, we show the uniform exponential convergence of the right-hand side of the above inequality. Observe that

$$\|(v_i)^o_{x_i}(x,\xi;u_i)\| \le \|\partial_{x_i} v_i(x,\xi)\| \le \kappa_i(\xi).$$

Under condition (d), there exist positive measurable functions $a_i(\xi)$, $\kappa_i(\xi)$, positive constants γ and δ and set $\Xi_\delta \subset \Xi$ such that

$$(v_i)^o_{x_i}(x',\xi;u_i') - (v_i)^o_{x_i}(x,\xi;u_i) \le a_i(\xi)\|x'-x\|^\gamma + \kappa_i(\xi)\|u_i'-u_i\|,$$

for all $\xi \in \Xi_\delta$ and $(x',u_i') \in B((x,u_i),\delta)$, where

$$\lim_{\delta\to 0} \text{Prob}(\xi \in \Xi_\delta) = 0.$$

Let $z_i := (x,u_i)$ and $Z_i := \mathcal{X} \times \{u_i \in \mathbb{R}^{m_i} : \|u_i\| \le 1\}$. The inequalities above shows that $(v_i)^o_{x_i}(\cdot,\xi^n;\cdot)$ is almost H-calm from above on set Z_i. Moreover, under condition (f), the moment generating function $M_x(t) := \mathbb{E}\{e^{t[a_i(\xi)+\kappa_i(\xi)]}\}$ is finite valued for t close to 0. By virtue of [38, Theorem 3.1], we have that for any $\epsilon_i > 0$, there exist positive constants $\hat{c}_i(\epsilon_i)$ and $\hat{\beta}_i(\epsilon_i)$, independent of N such that

$$\text{Prob}\left\{\sup_{(x,u)\in Z_i}\left(\frac{1}{N}\sum_{n=1}^N (v_i)^o_{x_i}(x,\xi^n;u) - \mathbb{E}\Big[(v_i)^o_{x_i}(x,\xi^n;u)\Big] \ge \epsilon_i\right)\right\} \le \hat{c}_i(\epsilon_i)e^{-N\hat{\beta}_i(\epsilon_i)},$$

for $i = 1,2,\dots,M$. For any $\epsilon > 0$, let $\epsilon_i > 0$ be such that $\sum_{i=1}^M \epsilon_i < \epsilon$. Let $\hat{c}(\epsilon) = M\max_{i=1}^M \hat{c}_i(\epsilon_i)$ and $\hat{\beta}(\epsilon) = \min_{i=1}^M \hat{\beta}_i(\epsilon_i)$. Then

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}}\max_{\|u\|\le 1}\sum_{i=1}^M \frac{1}{N}\sum_{n=1}^N\Big[(v_i)^o_{x_i}(x,\xi^n;u)\Big] - \mathbb{E}\Big[(v_i)^o_{x_i}(x,\xi;u)\Big] \ge \epsilon\right\} \le \hat{c}(\epsilon)e^{-\hat{\beta}(\epsilon)N}, \tag{46}$$

for N sufficiently large. ■

In what follows, we translate the uniform convergence of the subdifferential in Theorem 4.1 into the convergence of Nash-C-stationary points. We need a perturbation theorem on generalized equation.

Consider the following generalized equation

$$0 \in G(x) + \mathcal{N}_{\mathcal{C}}(x), \tag{47}$$

where $G(x) : \mathcal{C} \to 2^{\mathbb{R}^m}$ is a closed set-valued mapping, $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^m$. Let $\tilde{G}(x)$ be a perturbation of $G(x)$ and we consider the perturbed equation

$$0 \in \tilde{G}(x) + \mathcal{N}_{\mathcal{C}}(x). \tag{48}$$

Recall that a set-valued mapping F is said to be *outer semicontinuous* (osc for brevity) at $\bar{x} \in \mathbb{R}^n$ if $\overline{\lim}_{x\to\bar{x}} F(x) \subseteq F(\bar{x})$ or equivalently $\lim_{x\to\bar{x}} \mathbb{D}(F(x),F(\bar{x})) = 0$, where

$$\overline{\lim_{x\to\bar{x}}} F(x) := \{v \in R^m : \exists \text{ sequences } x_k \to \bar{x}, v_k \to v \text{ with } v_k \in F(x_k)\}.$$

The following lemma states that when $\mathbb{D}(\tilde{G}(x)), G(x))$ is sufficiently small uniformly w.r.t x, then the solution set of (48) is close to the solution set of (47).

LEMMA 4.2 [43] *Let $\mathcal{W}$ be a compact subset of $\mathcal{C}$. Let X^* denote the set of solutions to* (47) *in $\mathcal{W}$ and Y^* denote the set of solutions to* (48) *in $\mathcal{W}$. Assume that X^* and Y^* are nonempty. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that if $\sup_{x \in \mathcal{C}} \mathbb{D}(\tilde{G}(x), G(x)) < \delta$ and G is osc in $\mathcal{W}$, then $\mathbb{D}(Y^*, X^*) < \epsilon$.*

THEOREM 4.2 *Let $\{x^N\}$ be a sequence of Nash-C-stationary points satisfying* (34). *If conditions* (a)–(c) *of Theorem* 4.1 *hold, then w.p.*1, *an accumulation point of $\{x^N\}$ is a Nash-C-stationary point of the true problem which satisfies the first-order necessary equilibrium condition* (33). *If, in addition, conditions* (d)–(f) *of Theorem* 4.1 *are satisfied, then for every small positive number $\epsilon > 0$, there exist $\hat{c}(\epsilon) > 0$ and $\hat{\beta}(\epsilon) > 0$, independent of N, such that*

$$\text{Prob}\left\{d(x^N, X^*) \geq \epsilon\right\} \leq \hat{c}(\epsilon) e^{-\hat{\beta}(\epsilon) N}, \tag{49}$$

for N sufficiently large, where X^* denotes the set of Nash-C-stationary point of the true problem.

Proof The conclusion follows straightforwardly from Theorem 4.1 and Lemma 4.2. We omit the details. ■

Remark 4.1 The convergence results established here are stronger than those in our previous work [46]. To see this, recall that in [46] we considered the so-called weak Nash equilibrium conditions for a one-stage stochastic Nash equilibrium problem:

$$0 \in \mathbb{E}[\partial_{x_i} v_i(x_i, x_{-i}, \xi)] + \mathcal{N}_{X_i}(x_i), \; i = 1, \ldots, M, \tag{50}$$

where v_i is player i's objective function and $\mathbb{E}[\partial_{x_i} v_i(x_i, x_{-i}, \xi)]$ denotes Aumann's integral of Clarke subdifferential $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ [1]. Condition (50) is weaker than (33) in that

$$\partial_{x_i} \mathbb{E}[v_i(x_i, x_{-i}, \xi)] \subset \mathbb{E}[\partial_{x_i} v_i(x_i, x_{-i}, \xi)],$$

see [4, Theorem 2.7.2]. The corresponding first-order optimality conditions for the SAA problem considered there are:

$$0 \in \left(\frac{1}{N} \sum_{n=1}^{N} \partial_{x_i} v_i(x_i, x_{-i}, \xi^n) \right) + \mathcal{N}_{X_i}(x_i), \; i = 1, \ldots, M. \tag{51}$$

Since

$$\partial_{x_i} \left(\frac{1}{N} \sum_{n=1}^{N} v_i(x_i, x_{-i}, \xi^n) \right) \subset \left(\frac{1}{N} \sum_{n=1}^{N} \partial_{x_i} v_i(x_i, x_{-i}, \xi^n) \right),$$

condition (51) is also weaker than (34). Roughly speaking, the convergence results established in [46] are about weak Nash-C-stationary point defined through (51) to its true counterpart defined through (50), whereas the convergence results in Theorem 4.2 are for normal SAA Nash-C-stationary point defined by (51) to its true counterpart which satisfies (33).

5. Concluding remarks

In this article, we discuss a two-stage stochastic equilibrium problem with equilibrium constraints model and present a few source problems to motivate the model. The model may be extended by including some terms either in the objective or in the constraints which reflect risks such as variance, conditional value at risk [30], chance constraints [27] or certain dominance constraints [6]. To solve the two-stage stochastic equilibrium model, we propose to apply the well-known SAA method. The exponential rate of convergence means that the sample size will not be very large to obtain a reasonably reliable solution. In the case when the distribution of ξ is finite and known, SAA is not needed. The true problem may be solved through decomposition method or stochastic approximation method, see [33,42] for stochastic bilevel programs. More recently, by applying the decomposition-based splitting algorithm to a mixed-linear complementarity problem, Shanbhag et al. [34] provided computational evidence to show the convergence of the algorithm for stochastic bilevel programs with a discrete distribution.

Note also that the first-order optimality conditions in this article are characterized in terms of Clarke generalized gradient. It is possible to derive these conditions in terms of Mordukhovich limiting subdifferentials and through sensitivity analysis of two-stage SMPECs [44], the conditions in terms of the underlying functions in the equilibrium constraints. Indeed this has been done in [12,13,39] for some SEPEC models where the random variables have finite distribution. In our case, the limiting subdifferential approach can be applied to our SAA problem which can be viewed as SMPEC with finite distribution. The convergence results however will not be improved in that Aumann's integral convexifies the integrand, which means that the expected value of the limiting subdifferential of a Lipschitz function coincides with that of Clarke's. Finally, we note that the EPEC model is often nonconvex and therefore it would be practically interesting but challenging to identify a Nash equilibrium from a set of obtained Nash stationary points. The M-stationarity approach developed in [12,13,39] might provide a promising avenue towards this.

Acknowledgements

The authors thank the anonymous referee and the associate editor for their valuable comments.

References

[1] R.J. Aumann, *Integrals of set-valued functions*, J. Math. Anal. Appl. 12 (1965), pp. 1–12.
[2] P. Billingsley, *Convergence of Probability Measures*, Wiley & Sons, New York, 1968.
[3] C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, Lecture Notes in Mathematics, Vol. 580, Springer, Berlin, 1977.
[4] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
[5] V. DeMiguel and H. Xu, *A stochastic multiple leader stackelberg model: Analysis, computation, and application*, Oper. Res. 57 (2009), pp. 1220–1235.
[6] D. Dentcheva and A. Ruszczyński, *Optimality and duality theory for stochastic optimization with non-linear dominance constraints*, Math. Program. 99 (2004), pp. 329–350.
[7] W. Eggert and A. Haufler, *Capital taxation and production efficiency in an open economy*, Econom. Lett. 62 (1999), pp. 85–90.
[8] W. Eggert and M. Kolmar, *The taxation of financial capital under asymmetric information and the tax–competition paradox*, Scand. J. Econom. 106 (2004), pp. 83–104.

[9] A. Ehrenmann and K. Neuhoff, *A comparison of electricity market designs in networks*, Oper. Res. 57 (2009), pp. 274–286.

[10] F. Facchinei and J.-S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.

[11] G. Gürkan and J.-S. Pang, *Approximations of Nash equilibria*, Math. Program. 117 (2008), pp. 223–253.

[12] R. Henerion, J. Outrata and T. Surowiec, *Analysis of M-stationary points to an EPEC modelling oligopolistic competition in an electricity spot market*, preprint (2009), Weierstrass Institute for Applied Analysis and Stochastics, Berlin, p. 10117.

[13] R. Henerion and W. Römisch, *On M-stationary point for a stochastic equilibrium problem under equilibrium constraints in electricity spot market modelling*, Appl. Math. 52 (2007), pp. 473–494.

[14] C. Hess, *Set-valued Integration and Set-valued Probability Theory: An Overview*, *Handbook of Measure Theory*, Vols. I and II, North-Holland, Amsterdam, 2002, pp. 617–673.

[15] B.F. Hobbs, C.B. Metzler, and J.-S. Pang, *Strategic gaming analysis for electric power systems: An MPEC approach*, IEEE Trans. Power Syst. 15 (2000), pp. 637–645.

[16] X. Hu and D. Ralph, *Using EPECs to model bilevel games in restructured electricity markets with locational prices*, Oper. Res. 55 (2007), pp. 809–827.

[17] H. Jiang, D. Ralph and S. Scholtes, *Equilibrium problems with equilibrium constraints: A new modelling paradigm for revenue management*, The 4th Annual INFORMS Revenue Management and Pricing Section Conference, 2004.

[18] A. Koh, Nash Equilibrium, Collusion in games and the coevolutionary particle swarm algorithm, White Rose Research Papers, 2008. Available at http://eprints.whiterose.ac.uk/7784/

[19] S. Leyffer and T. Munson, *Solving multi-leader-follower games*, Optim. Methods Softw. 25 (2010), pp. 601–623.

[20] G. Lin, H. Xu, and M. Fukushima, *Monte Carlo and quasi-Monte Carlo sampling methods for a class of stochastic mathematical programs with equilibrium constraints*, Math. Methods Oper. Res. 67 (2008), pp. 423–441.

[21] Y. Liu, H. Xu, and G.H Lin, *Stability analysis of two-stage stochastic mathematical programs with complementarity constraints via NLP-regularization*, SIAM J. Optim. 21 (2011), pp. 609–705.

[22] B.S. Mordukhovich, *Equilibrium problems with equilibrium constraints via multiobjective optimization*, Optim. Methods Softw. 19 (2004), pp. 479–492.

[23] B.S. Mordukhovich, *Optimization and equilibrium problems with equilibrium constraints*, Omega 33 (2005), pp. 379–384.

[24] J. Outrata, *A note on a class of equilibrium problems with equilibrium constraints*, Kybernetika 40 (2004), pp. 585–594.

[25] J.S. Pang and M. Fukushima, *Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games*, Comput. Manag. Sci. 2 (2005), pp. 21–56.

[26] N. Papageorgiou, *On the theory of Banach space valued multifunctions 1. Integration and conditional expectation*, J. Multivariate Anal. 17 (1985), pp. 185–206.

[27] G.C. Pflug and N. Broussev, *Electricity swing options: Behavioural models and pricing*, Eur. J. Oper. Res. 197 (2009), pp. 1041–1050.

[28] E.L. Plambeck, B.R. Fu, S.M. Robinson, and R. Suri, *Sample-path optimization of convex stochastic performances functions*, Math. Program. 75 (1996), pp. 137–176.

[29] D. Ralph, *Mathematical Programs with Complementarity Constraints in Traffic and Telecommunications Networks*, Philosophical Transactions of Royal Society of London: Mathematical, Physical and Engineering Sciences, Vol. 366, The Royal Society, London, 2008, pp. 1973–1987.

[30] U. Ravat and U.V. Shanbhag, *On the characterization of solution sets of smooth and nonsmooth stochastic Nash games*, Proceedings of the American Control Conference (ACC), Baltimore, 2010.

[31] S.M. Robinson, *Analysis of sample-path optimization*, Math. Oper. Res. 21 (1996), pp. 513–528.

[32] A. Ruszczyński and A. Shapiro, *Stochastic programming models*, in *Stochastic Programming*, A. Ruszczyński and A. Shapiro, eds., Handbooks in OR & MS, Vol. 10, North-Holland, Amsterdam, 2003.

[33] U.V. Shanbhag, *Decomposition and sampling methods for stochastic equilibrium problems*, Ph.D. thesis, Management Science and Engineering (Operations Research), Stanford, CA, 2006.

[34] U.V. Shanbhag, G. Infanger, and P.W. Glynn, *A Complementarity Framework for Forward Contracting under Uncertainty*, Oper. Res. (2011), in press.

[35] A. Shapiro, *Monte Carlo sampling methods*, in *Stochastic Programming*, A. Ruszczyński and A. Shapiro, eds., Handbooks in OR & MS, Vol. 10, North-Holland, Amsterdam, 2003.

[36] A. Shapiro, *Stochastic programming with equilibrium constraints*, J. Optim. Theory Appl. 128 (2006), pp. 223–243.

[37] A. Shapiro and H. Xu, *Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation*, Optimization 57 (2008), pp. 395–418.

[38] H. Sun and H. Xu, *A note on uniform exponential convergence of sample average approximation of random functions*, J. Math. Anal. Appl. 385 (2011), pp. 698–708.

[39] H.T.M. Surowiec, *Explicit stationarity conditions and solution characterization for equilibrium problems with equilibrium constraints*, PhD thesis, Humboldt University of Berlin, Berlin, 2010.

[40] C.-L. Su, *Analysis on the forward market equilibrium model*, Oper. Res. Lett. 35 (2007), pp. 74–82.

[41] J. Wang, D. Chiu, and C. Lui, *A game theoretic analysis of the implications of overlay network traffic on ISP peering*, Comput. Networks 52 (2008), pp. 2961–2974.

[42] A.S. Werner, *Bilevel stochastic programming problems: Analysis and application to telecommunications*, PhD diss., Norwegian University of Science and Technology, 2004.

[43] H. Xu, *Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming*, J. Math. Anal. Appl. 368 (2010), pp. 692–710.

[44] H. Xu and J.J. Ye, *Necessary optimality conditions for two-stage stochastic programs with equilibrium constraints*, SIAM J. Optim. 20 (2010), pp. 1685–1715.

[45] H. Xu and D. Zhang, *Smooth sample average approximation of nonsmooth stochastic problems and applications*, Math. Program. Ser. A 119 (2009), pp. 371–401.

[46] H. Xu and D. Zhang, Stochastic Nash equilibrium problems: Sample average approximation and applications, 2011. Available at www.optimization-online.org

[47] H. Yang, F. Xiao, and H. Huang, *Private road competition and equilibrium with traffic equilibrium constraints*, J. Adv. Transport. 43 (2009), pp. 21–45.

[48] J. Yao, S.S. Oren, and I. Adler, *Two-settlement electricity markets with price caps and Cournot generation firms*, Eur. J. Oper. Res. 181 (2007), pp. 1279–1296.

[49] D. Zhang, H. Xu, and Y. Wu, *A stochastic two-stage equilibrium model for electricity markets with two way contracts*, Math. Methods Oper. Res. 71 (2010), pp. 1–45.

[50] J. Zhou, W.H.K. Lam, and B.G. Heydecker, *The generalized Nash equilibrium model for oligopolistic transit market with elastic demand*, Transport. Res. Part B 39 (2005), pp. 519–544.