

## An approximation scheme for a class of risk-averse stochastic equilibrium problems

Juan Pablo Luna · Claudia Sagastizábal ·  
Mikhail Solodov

Received: June 2013 / Accepted: January 2016

**Abstract** We consider two models for stochastic equilibrium: one based on the variational equilibrium of a generalized Nash game, and the other on the mixed complementarity formulation. Each agent in the market solves a single-stage risk-averse optimization problem with both here-and-now (investment) variables and (production) wait-and-see variables. A shared constraint couples almost surely the wait-and-see decisions of all the agents. An important characteristic of our approach is that the agents hedge risk in the objective functions (on costs or profits) of their optimization problems, which has a clear economic interpretation. This feature is obviously desirable, but in the risk-averse case it leads to variational inequalities with set-valued operators – a class of problems for which no established software is currently available. To overcome this difficulty, we define a sequence of approximating differentiable variational inequalities based on smoothing the nonsmooth risk measure in the agents’ problems, such as average or conditional value-at-risk. The smoothed variational inequalities can be tackled by the PATH solver, for example. The approximation scheme is shown to converge, including the case when smoothed problems are solved approximately. An interesting by-product of our proposal is that the smoothing approach allows us to show existence of an equilibrium for the original problem. To assess the proposed technique, numerical results are presented. The first set of experiments is on randomly generated equilibrium problems, for which we compare the proposed methodology with the standard smooth reformulation of average value-at-risk minimization (using additional variables to rewrite the associated max-function). The second set of experiments deals with a part of the real-

---

The first author was supported by CNPq post-doctoral scholarship 503441/2012-0. The second author is partially supported by Grants CNPq 303840/2011-0, AFOSR FA9950-11-1-0139, as well as by PRONEX-Optimization and FAPERJ. The third author is supported in part by CNPq Grants 302637/2011-7 and PVE 401119/2014-9, by PRONEX-Optimization and by FAPERJ.

J. P. Luna  
IMPA – Instituto de Matemática Pura e Aplicada,  
Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.  
(Currently with: COPPE-UFRJ, Engenharia de Produção, Cidade Universitária, Rio de Janeiro - RJ, 21941-972, Brazil.) E-mail: jluna@impa.br

C. Sagastizábal  
Visiting researcher, IMPA. E-mail: sagastiz@impa.br

M. Solodov  
IMPA. E-mail: solodov@impa.br

life European gas network, for which Dantzig-Wolfe decomposition can be combined with the smoothing approach.

**Keywords** stochastic equilibrium · risk aversion · generalized Nash game · variational inequality · complementarity · Dantzig–Wolfe decomposition.

**Mathematics Subject Classification (2010)** 65K10 · 90C30 · 49J53.

## 1 Introduction

Over the last decades and in a worldwide trend, many industries that were considered as “natural monopolies” in the 1970s (electricity, telecommunications, natural gas, water supply) were restructured to introduce competition into one or more of their horizontal segments. Nevertheless, even in the most liberalizing countries, some strategical sectors continue to be subject to regulations in quality, price and entry. Such is the case for networks transmitting and distributing electricity or transporting natural gas. For these oligopolistic industries, the regulatory mechanisms have important implications not only for supporting wholesale and retail competition but also to maintain the network reliability. To have a full understanding of the market behaviour, it is necessary to understand the impact of distortions introduced by the regulator when capping prices, or when applying rewards and incentives for efficient production, like the emission allowance allocation system in [49].

We consider equilibrium problems for a market with competing *risk-averse* agents that try to maximize profit subject to coupling constraints resulting from regulatory interventions or market clearing conditions. Among related literature, however mostly casting the problem in deterministic or risk-neutral settings, we refer to [11, 25, 1, 35, 36, 38, 26, 16, 10, 48]. We also mention [9], which emphasizes the tradeoff between maximal profit and minimal risk for thermal power producers in a pool-based electricity market. In this work, we adopt the modeling of [32]. Our approach is comprehensive enough to include the electricity generation-capacity expansion model [17] and the European natural gas market model in [23], set in a stochastic environment.

This type of problems requires the interplay of various areas, like optimization, game theory, mixed complementarity problems (MCP), and variational inequalities (VI). The latter, in particular, can be employed for capturing certain solutions of convex Generalized Nash Equilibrium Problems (GNEP). Specifically, the VI formulation finds a *Variational Equilibrium*, a special Nash equilibrium that privileges points that lead to equal marginal values for all the players in the game; see [19] and also [18, 28, 29].

In the presence of uncertainty, VIs give rise to numerous challenges already in the risk-neutral case. To start with and as discussed in Section 2 below, where and how to introduce uncertainty and risk in a tractable/convenient way is not straightforward, and a number of different approaches have been considered.

After having decided how to handle uncertainty, a further challenge is how to endow a VI with risk aversion in an economically meaningful and computationally tractable manner. We focus on *computable* stochastic equilibrium prices for a market in which agents exhibit risk aversion. Specifically, in our model each agent in the market solves a single-stage risk-averse optimization problem with:

- here-and-now and wait-and-see variables (investment and production, respectively), and
- a shared constraint that couples almost surely the wait-and-see decisions of all the agents (related to market clearing); see (1) below.

Regarding the modeling paradigms of *competition*, *dynamics* and *hierarchy* described in [34], since we put all agents in the same decision level and uncertainty is not revealed stage-wise but once for all, our proposal emphasizes competition over the two other paradigms.

The introduction of risk functionals in the agents' problems gives rise to some numerical issues that have to be dealt with. For nonsmooth risk measures like the widely used *conditional or average value-at-risk (AVaR [42])*, we obtain VIs with set-valued operators. Currently, there appears to be no established efficient software available for VIs of this class, unless the VI is *integrable*, i.e., it is actually an optimization problem (in our equilibrium model, this amounts to eliminating the shared constraint, a nonrealistic setting, since clearing the market is of course a necessity). We therefore build a family of approximating VIs with single-valued operators, comprised by gradients of functions obtained from smoothing the *AVaR*. The smoothed functions are also risk measures but, unlike the *AVaR*, they are not coherent. When the smoothing parameter goes to zero, we show convergence of the approximation scheme for stochastic equilibrium models based on both the GNEP and the MCP approaches. By this token, we can also show existence of an equilibrium for the original risk-averse problem, under some natural conditions.

The sequence of smoothed approximating VI problems is solved by PATH [13,21]. We note that nonsmoothness resulting from minimization and from the positive-part function in the definition of *AVaR* can in principle be handled using the standard reformulation which introduces extra variables and shifts certain terms from the objective functions to the constraints in the agents' problems. However, as discussed in Section 4, such a strategy does not scale well, as it makes the problem size grow quickly with the number of scenarios, hence leading to unsolvable instances much faster.

The rest of the paper is organized as follows. In Section 2 we describe our setting: the approach to handling uncertainty, risk aversion, and the GNEP and MCP models of the market. An approximation scheme for the set-valued VIs obtained by smoothing the non-differentiable risk measures, and the corresponding convergence theory, are presented in Section 3. Section 4 describes our solution method based on random samples for a sample space with a finite number of scenarios, and the variational formulations of the two models. This section also contains structural and computational comparisons with the reformulation resulting from what would appear to be the standard (smooth) reformulation of the average-value-at-risk, which requires extra variables and constraints. In particular, we provide numerical validation of the advantage of our approach as the number of scenarios grows. Section 5 contains numerical results investigating the effects of risk aversion, volatility, and other issues, for part of the real-life European gas network. In addition, it is shown how the Dantzig-Wolfe decomposition [22,31] of VIs can improve solution times in this context.

Our notation is mostly standard. The inner product (in an arbitrary space) is denoted by  $\langle \cdot, \cdot \rangle$ , and the notation  $x \perp v$  means that  $\langle x, v \rangle = 0$ . By  $[\cdot]^+$  we denote the positive-part function, i.e.,  $[x]^+ = \max\{0, x\}$ . The nonnegative orthant in  $\mathbb{R}^l$  is denoted by  $\mathbb{R}_+^l$ . For a convex function  $\psi$ , its subdifferential at a point  $x$  is the set  $\partial\psi(x) = \{v : \psi(x') \geq \psi(x) + \langle v, x' - x \rangle \text{ for all } x'\}$ ; for a convex function of two variables like  $\psi(x, x')$ ,  $\partial_x\psi$  stands for the subdifferential mapping of  $\psi$  with respect to the variable  $x$ . For a vector  $x$ , the notation  $(x^i, \bar{x}^{-i})$  means that the second block of components of  $x$  is fixed in the given context.

Let  $C$  be a closed convex set. Its normal cone at the point  $x$  is the set  $N_C(x) = \{w : \langle w, x' - x \rangle \leq 0 \text{ for all } x' \in C\}$  if  $x \in C$  and the empty set otherwise. By  $P_C(\cdot)$  we denote the orthogonal projection mapping onto the closed convex set  $C$ .

For a single-valued mapping  $F$ , the usual variational inequality [20] is stated as

$$\text{VI}(F, C) : \text{find } \bar{x} \in C \text{ such that } \langle F(\bar{x}), x - \bar{x} \rangle \geq 0 \quad \text{for all } x \in C.$$

Equivalently,  $\text{VI}(F, C)$  means finding an  $\bar{x}$  that satisfies the inclusion  $F(\bar{x}) + N_C(\bar{x}) \ni 0$ , or the projection equation  $\bar{x} - P_C(\bar{x} - F(\bar{x})) = 0$ .

If  $\Psi$  is a set-valued mapping  $\Psi$ , the associated VI is the problem

$$\text{VI}(\Psi, C) : \text{find } \bar{x} \in C \text{ such that for some } \bar{w} \in \Psi(\bar{x}), \quad \langle \bar{w}, x - \bar{x} \rangle \geq 0 \quad \text{for all } x \in C.$$

## 2 Modelling Equilibrium Problems in the Presence of Risk Aversion

We consider two different models for stochastic equilibrium, based on game theory and on complementarity. It is worth mentioning several related works, starting with the study on endogenous stochastic equilibria in [39, Sec. 5]. Stochastic linear complementarity models are explored in [5, 7] from an expected-residual and robust perspectives, respectively. Stochastic mathematical programs with equilibrium constraints are analyzed in [37, 44, 30]. For the natural gas market, the deterministic MCP model from [15] and [23] was revisited in a stochastic risk-neutral setting in [14]. For electricity markets, game-theoretical and MCP formulations have been considered in [27] and [17], respectively.

Regarding the merits of the two modelling approaches, it is sometimes argued that the MCP formulation provides an adequate framework for imperfect markets. Nevertheless, at least in a deterministic environment, it is shown in [31] that equivalent equilibria can be obtained by GNEP and by MCP. Since the two models are eventually cast in this paper as related stochastic VIs, we shall carry on our discussion for both of them in parallel.

### 2.1 The Setting

The VI under consideration comes from writing an equilibrium problem for a market with  $N$  agents trying to optimize their activities. In a deterministic environment the VI operator has components of the form  $F^i(x) = \nabla_{x^i} f^i(x)$ ,  $i = 1, \dots, N$ , for certain functions  $f^i$  and subvectors  $x^i$ , representing the objective functions and the actions of each agent.

In a stochastic environment, the operator and the feasible set depend on uncertain parameters, say  $\xi$ . In this context, deciding what is a “good” VI formulation is not straightforward. In particular, the mathematical model needs to define how to deal with the fact that parameters are random. In addition, if agents are risk-averse,  $F^i$  is a (set-valued) *subdifferential* instead of a (single-valued) gradient.

As a first step to handle uncertainty, decisions are separated into here-and-now and wait-and-see types, [45]. For a market with  $N$  agents trying to optimize their activities of production and investment in capacity expansion, the  $i$ th agent distinguishes between:

- “investment” variables  $z^i \in \mathbb{R}^{n^i}$ , of the here-and-now type: they are decided prior to any realization of uncertainty (the capacity expansion is decided before knowing the amount of future product demand).
- There are also “generation” variables  $q_{\omega}^i \in \mathbb{R}^{m^i}$ , of the wait-and-see type, depending on some uncertainty that becomes known when the decision must be taken. These variables are random functions in the probability space  $L_p(\Omega, \mathcal{F}, \mathbb{P})$  for  $p \in [1, +\infty)$ , a measure  $\mathbb{P}$ , and a sample space  $\Omega$  equipped with a sigma-algebra  $\mathcal{F}$ . For example, when the random

vector  $\xi_\omega$  represents the product demand, the decision of how much to produce can be taken after observing a realization of  $\omega$ .

The second step when handling uncertainty is to determine how random variables unveil along time steps and if the decision maker has access to this information. There is a variety of situations, ranging from deterministic *clairvoyant* models to multistage models with recourse. In the former, decisions are taken considering only one realization of uncertainty (wait-and-see variables are meaningless in this setting). In the latter, here-and-now variables are fixed before all realizations, but wait-and-see variables are determined stagewise, as the random variables unveil.

For our problem we choose a single-stage model with recourse, with both here-and-now and wait-and-see variables, as some decisions need to be taken before any realization, while others may wait until randomness is realized. Differently from the multistage model, uncertainty is unveiled at once: realizations in  $\Omega$  are known to the decision maker *a priori*.

The single-stage model with recourse is the simplest one involving wait-and-see variables. Handling uncertainty by a two-stage model with recourse would be desirable (the impact of stochastic data would be better dealt with) but, contrary to what is customary in optimization, there are some equilibrium problems for which multistage formulations are not possible. We shall come back to this important issue in Section 2.4.

Wait-and-see variables are functions of the uncertainty. Thus, formally, the most suitable notation for such variables would be  $q^i(\omega)$ . Nevertheless in what follows we write  $q_\omega^i$  instead, for the sake of better readability in various long expressions involving  $\omega$  that will appear in our development; for example, but not only, (4) and (5).

## 2.2 The $i$ th Agent Problem

Continuing with the investment-production illustration, for each realization  $\omega \in \Omega$  the  $i$ th agent decisions to invest and produce depend on available resources and technology, represented by feasible sets  $X_\omega^i$ , endogenous to each agent.

The wait-and-see actions of the other agents, represented by the quantities  $\bar{q}_\omega^{-i}$  below, have an impact on the  $i$ th agent's costs and remuneration. The total cost of the agent is the sum of an investment cost  $I^i$  and (stochastic) production costs  $c_\omega^i$ :

$$I^i(z^i) + c_\omega^i(q_\omega^i, \bar{q}_\omega^{-i}).$$

Here again, the above notation for the cost shortens the functional dependence on the random variable:  $c_\omega^i(q_\omega^i, \bar{q}_\omega^{-i})$  stands for  $c^i(q^i(\omega), q^{-i}(\omega); \omega)$ .

After transformation of the product and eventual transportation losses, the produced volume  $q_\omega^i$  differs from the actual amount reaching its final destination. This phenomenon is represented here by the random affine functions  $h_\omega^i$ , depending only on wait-and-see variables: out of the production  $q_\omega^i$ , the amount  $h_\omega^i(q_\omega^i)$  is eventually delivered to the client.

If the product is remunerated at a unit price  $\pi_\omega$ , the random revenue of the  $i$ th agent is

$$\langle \pi_\omega, h_\omega^i(q_\omega^i) \rangle - \left( I^i(z^i) + c_\omega^i(q_\omega^i) \right).$$

The remuneration price  $\pi_\omega$  is considered exogenous for most of the agents, except for those agents large enough to exert *market power*, see Remark 1 below. The price  $\pi_\omega$  is a Lagrange

multiplier of some constraint coupling the agents' actions in the market, for example a market clearing condition requiring that supply meets demand, written in the abstract form:

$$S := \left\{ \left( (q_\omega^i)_{\omega \in \Omega} \right)_{i=1}^N : \sum_{j=1}^N h_\omega^j(q_\omega^j) \leq 0 \in \mathbb{R}^{m^0} \quad \text{a.e. } \omega \in \Omega \right\}. \quad (1)$$

We associate to the inequality in  $S$  a nonnegative Lagrange multiplier  $\pi_\omega \in L_{p^*}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{m^0})$  with  $1/p + 1/p^* = 1$  ( $\pi_\omega$  denotes a vector function of  $\omega$ ).

To ensure feasibility of the agents' problems, in what follows we assume that the property of *relatively complete recourse* is satisfied:

$$\text{for all } (z^1, \dots, z^N) \text{ there exists } (q_\omega^1, \dots, q_\omega^N) \text{ such that } \begin{cases} (z^i, q_\omega^i) \in X_\omega^i \\ \sum_{j=1}^N h_\omega^j(q_\omega^j) \leq 0 \end{cases} \quad \text{a.e. } \omega \in \Omega.$$

### 2.3 Two Equilibrium Models with Risk Aversion

The market equilibrium will be given by a primal-dual point of the form

$$(\bar{z}, \bar{q}, \bar{\pi}) = \left( (\bar{z}^i, (\bar{q}_\omega^i)_{\omega \in \Omega})_{i=1}^N, (\bar{\pi}_\omega)_{\omega \in \Omega} \right),$$

representing decisions of the agents and equilibrium prices, respectively.

The concept of equilibrium depends on two important issues: how each agent hedges risk, and how the agents interpret their own actions on the market. Regarding the first item, we model the  $i$ th agent aversion to volatility by a monotone convex risk measure  $\rho^i(\cdot)$ , assumed to be a proper function, see [12, Chapter 6].

For a random outcome  $Y \in L_1(\Omega, \mathcal{F}, \mathbb{P})$  representing a loss and a given confidence level  $1 - \varepsilon \in (0, 1)$ , an important example of a coherent risk measure is the well-known *average value-at-risk* introduced in [42] as *conditional value-at-risk*:

$$AVaR_\varepsilon(Y) := \min_u \left\{ u + \frac{1}{1 - \varepsilon} \mathbb{E}[Y_\omega - u]^+ \right\}. \quad (2)$$

An observation relevant for the subsequent computational developments is that the *AVaR* measure is nondifferentiable. In fact, as explained in [24, Sec.VI.4.5], potential nonsmoothness of a function like *AVaR* is induced not only by the positive-part function involved in its definition, but also by the fact that it is a value-function (i.e., optimal value of an optimization problem). In particular, differentiability at a point depends on the minimum being attained; we refer to Remark 4.5.4 in [24, Chapter VI] for more details.

For a given risk-aversion parameter  $\kappa^i \in [0, 1]$ , and letting  $\mathbb{E}$  denote the expected value function, the agent's different degrees of risk aversion are expressed by the function

$$\rho^i(Y) := (1 - \kappa^i)\mathbb{E}(Y) + \kappa^i AVaR_{\varepsilon^i}(Y). \quad (3)$$

Determining how the agents interpret their own actions on the market depends on the modeling choice. We shall focus on equilibrium models based on either the GNEP or the MCP approaches. In the first case each agent in the market minimizes cost, *guessing* the decisions  $\bar{q}_\omega^{-i}$  of the other agents, paying attention to satisfy the shared constraint in the set  $S$  defined in (1). A *variational equilibrium* [18] corresponds to the additional requirement that the optimal Lagrange multipliers associated to the shared constraint are the same for all

the players, say  $\bar{\pi}_\omega$ . Such an equilibrium is desirable, as it has the economic interpretation of being *fair* – at an equilibrium all the agents are remunerated at the same price.

Alternatively, when modeling the equilibrium as a MCP, each agent maximizes the remuneration. The agents' actions are coupled by the constraint set  $S$  and a complementarity condition between each constraint in  $S$  and  $\pi_\omega$ , a.e.  $\omega \in \Omega$ . Coupling constraints *do not* enter the agents' optimization problems. In the MCP model, each agent takes decisions *independently* of both the other agents' decisions  $\bar{q}_\omega^i$  and the shared constraint.

For the risk-averse GNEP model, we let  $R^i$  be a risk measure (in our computational developments, it will be a smoothed version of (3), aiming to approximate a solution of the problem associated to (3)). We thus have

$$\text{Stochastic GNEP:} \quad \begin{cases} \text{solve} \\ \text{for } i = 1, \dots, N \end{cases} \left\{ \begin{array}{l} \min I^i(z^i) + R^i\left(c_\omega^i(q_\omega^i, \bar{q}_\omega^{-i})\right) \\ \text{s.t. } z^i \in \mathbb{R}^{m^i}, q^i \in L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{m^i}), \\ \text{and, a.e. } \omega \in \Omega : \\ (z^i, q_\omega^i) \in X_\omega^i, \\ h_\omega^i(q_\omega^i) + \sum_{i \neq j=1}^N h_\omega^j(\bar{q}_\omega^j) \leq 0. \end{array} \right. \quad (4)$$

A variational equilibrium of the game is the triple  $(\bar{z}, \bar{q}, \bar{\pi})$ , where  $\bar{\pi} \in L_{p^*}(\Omega, \mathcal{F}, \mathbb{P})$  is the multiplier associated to the last constraint above (the same for all the players),

For the risk-averse complementarity equilibrium, we consider the model

$$\text{Stochastic MCP:} \quad \begin{cases} \text{solve} \\ \text{for } i = 1, \dots, N \end{cases} \left\{ \begin{array}{l} \min I^i(z^i) + R^i\left(c_\omega^i(q_\omega^i, \bar{q}_\omega^{-i}) - \langle \pi_\omega, h_\omega^i(q_\omega^i) \rangle\right) \\ \text{s.t. } z^i \in \mathbb{R}^{m^i}, q^i \in L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{m^i}), \\ (z^i, q_\omega^i) \in X_\omega^i \quad \text{a.e. } \omega \in \Omega, \end{array} \right. \quad (5)$$

together with  $0 \leq \pi_\omega \perp \sum_{j=1}^N h_\omega^j(q_\omega^j) \leq 0 \quad \text{a.e. } \omega \in \Omega$ .

As shown in Section 4 below, both models (4) and (5) of equilibrium with risk aversion can be cast as stochastic VIs; see also [32]. Theorem 3 shows existence of an equilibrium for these models, using an argument which in fact makes use of the smoothing approach itself.

*Remark 1 (Price cap, regulatory interventions, and market power)*

In energy markets there is often a higher entity; for example, a regulating agency representing the consumers, or an independent system operator. This entity acts on the market by capping prices, or by encouraging the demand satisfaction and the respect of environmental constraints. As shown in [32], for both the GNEP and MCP models it is possible to incorporate an additional player, say the 0th agent, to represent such a regulator. For example, in the presence of a price cap  $PC \in L_{p^*}(\Omega, \mathcal{F}, \mathbb{P})$ , the variable  $q_\omega^0 \in \mathbb{R}^{m^0}$  represents a deficit in demand while  $c_\omega^0(q_\omega^0, \bar{q}_\omega^0) = \langle PC_\omega, q_\omega^0 \rangle$  measures the impact that capping prices has on the market when the realization of uncertainty is given by  $\omega$ .

Both our models can also incorporate market power. Following [23], when a price-sensitive demand curve is available, there are given intercept  $d_0$  and matrix  $P$  defining the inverse demand function as  $P \cdot + d_0$ . In this case, for the MCP model the remuneration changes from  $\langle \pi_\omega, h_\omega^i(q_\omega^i) \rangle$  to

$$\left\langle (1 - \varphi^i) \pi_\omega + \varphi^i \left( P h_\omega^i(q_\omega) + d_0 \right), h_\omega^i(q_\omega^i) \right\rangle,$$

for a factor  $\varphi^i \in [0, 1]$  representing the strength of the agent in the market. For the GNEP, [32] shows that modelling market power is equivalent to having an additional player who tries to maximize the revenue resulting from the portion of the total production that is not affected by the market power of the agents, at the price given by the inverse demand function.  $\square$

## 2.4 On Two-stage Models and Shared Constraints

In both (4) and (5), each agent in the market solves a single-stage risk-averse optimization problem. To justify this modeling choice we now explain why two-stage formulations, such as the ones proposed in [6] and [47], are not applicable in our case.

In the stochastic model in [6] the wait-and-see variables account for feasibility deviation. The modification of the D-gap function therein gives a new residual function whose recourse term “adjusts” the here-and-now decision. Being based on a VI’s gap function, such an approach would be suitable for risk-neutral agents, which is not our setting.

A two-stage stochastic Nash game with risk aversion is considered in [47]. Regarding this work, the slight yet fundamental difference with our formulation is the following. The shared constraint in (1) makes the  $i$ th decision  $q_\omega^i$  depend on the *whole* wait-and-see vector:  $q_\omega^i = q_\omega^i(\bar{q}_\omega^{-i})$ . This is in contrast to relation (1.3) in [47], where the wait-and-see variable  $q_\omega^i$ , denoted by  $y_i$ , depends *only* on the here-and-now variables  $z^i, z^{-i}$  denoted by  $x_i, x_{-i}$  in that work. The fact that  $y_i$  is independent of  $y_{-i}$  in [47] becomes evident when inspecting the sets (1.4) and (1.5) therein, and comparing their expression with our feasible set (1).

It is precisely because of the feasible set of the format (1) that two-stage formulations for (4) or (5) are not possible. As there can be multiple second-stage solutions, it is not clear how to define the recourse function. This difficulty is *independent of risk aversion*; it arises already when trying to give a meaning to two-stage Nash games for problems whose shared constraints couple second-stage variables (corresponding to our wait-and-see variables). To illustrate this phenomenon, take a singleton set  $\Omega$ , so that there are no subindices  $\omega$  in (4). To further simplify the writing, let the  $i$ th-cost function depend only on the  $i$ th player and suppose the endogenous feasible sets are polyhedral,  $X^i = \{(z^i, q^i) : T^i z^i + W^i q^i = b^i\}$ . The corresponding GNEP has deterministic problems, for  $i = 1, \dots, N$ :

$$\left\{ \begin{array}{ll} \min_{(z^i, q^i) \in \mathbb{R}^{n^i} \times \mathbb{R}^{m^i}} & I^i(z^i) + R^i(c^i(q^i)) \\ \text{s.t.} & T^i z^i + W^i q^i = b^i \\ & h^i(q^i) + \sum_{j=1, j \neq i}^N h^j(q^j) \leq 0. \end{array} \right.$$

To reformulate the  $i$ th player problem in two levels (corresponding to a two-stage formulation in the stochastic setting), we would need a first level problem only on variables  $z^i$ :

$$\min_{z^i \in \mathbb{R}^{n^i}} I^i(z^i) + R^i(Q^i(z^i)).$$

The second-level function, akin to the recourse function in a two-stage problem, is given by

$$Q^i(z^i) := \left\{ \begin{array}{ll} \min_{q^i \in \mathbb{R}^{m^i}} & c^i(q^i) \\ \text{s.t.} & W^i q^i = b^i - T^i z^i \\ & h^i(q^i) + \sum_{j=1, j \neq i}^N h^j(q^j) \leq 0. \end{array} \right.$$



Observing the second-level optimization problem above we realize that, for the function  $Q^i$  to be well defined, some rule regarding the variables  $q^{-i} = (q^j : j \neq i)$  needs to be established (a “selection”). This difficulty is a direct consequence of the combination of two features: having here-and-now variables and having a shared constraint. Without a shared constraint, the second-level problems can be solved. Without the here-and-now variables, the VI can be derived, and a variational equilibrium can be found. With both the shared constraints and the here-and-now variables, the VI involves the gradient of the second-level function, which is not explicit, but known only implicitly (because it is a value function).

Likely for this reason, the work [27] considers a single-stage game formulation with here-and-now and wait-and-see variables, similar to (4). There are two important differences with our approach, however:

- First, the risk measure  $R^i$  in the cited work is set on a cost function that depends only on here-and-now variables and on an uncertain parameter, say  $\varphi_\omega$ . In our notation, this would correspond to replacing  $c_\omega^i(q_\omega^i, \bar{q}_\omega^{-i})$  in (4) by  $\varphi_\omega z^i$ .
- Second, in [27] the feasible sets  $X_\omega^i$  do not involve here-and-now variables. Instead of  $(z^i, q_\omega^i) \in X_\omega^i$  in (4), the constraint is  $q_\omega^i \in X_\omega^i$ .

### 3 A Convergent Approximation Scheme

From this section on we now focus on equilibrium models (4) and (5) with finite sampling space: given a number  $K > 0$ ,

$$\Omega = \Omega_K := \{\omega_1, \dots, \omega_K\},$$

for scenarios  $\omega_k$  with probability  $p_k$  for  $k = 1, \dots, K$ . In (4) and (5),  $q_\omega$  becomes  $q_k$  for some  $k \in \{1, \dots, K\}$  and similarly for all the random elements. Hence, instead of having  $L_p$ -functions we deal with concatenation of objects with  $K$ -components, for instance

$$q_K := \left( q_k = q_{\omega_k} \right)_{k=1}^K \text{ is a vector of dimension } K \sum_{i=1}^N m^i.$$

When the agents hedge risk in their individual problems with nonsmooth measures, such as  $\rho^i = AVaR$ , the mapping in the corresponding variational inequality has *set-valued* components. As there currently appears to be no established software to handle such problems, the corresponding formulation is basically intractable, at least unless one resorts to some reformulations, discussed in detail in Section 4.2 below. Roughly speaking, nonsmoothness associated to minimization which involves the positive-part function (like in the definition (2) of  $AVaR$ ) can be removed by introducing additional variables and moving certain terms from the objective functions to the constraints in the agents’ problems. This approach, and its disadvantages in our setting, are discussed in Section 4 below. In this section, we consider how to deal with the nonsmoothness issues without going through such reformulations.

To overcome the difficulty presented by the set-valued VI operators, while preserving the decomposable structure, we define a sequence of approximating problems that employ *smooth* risk measures  $R^i = \rho_\tau^i$  where  $\tau > 0$  is a parameter. The corresponding VIs are smooth single-valued, and can be tackled using the popular PATH solver [13, 21]. As the smoothing parameter  $\tau$  tends to zero, the corresponding solutions of the smoothed VIs asymptotically approach solutions of the original problems, i.e., those with the nonsmooth risk measures  $R^i = \rho^i$ . It is interesting to mention that thanks to this convergence result, we can also show the existence of an equilibrium for the original problem; see Theorem 3 below.

Smoothing techniques are common in optimization and complementarity; see, e.g., [3, 40, 4, 2] and also [20, Chapter 11.8]. In the stochastic setting those ideas have been used, for example, by [33] to smooth risk constraints, by [30] to handle stochastic optimization problems with equilibrium constraints, and by [6] in a VI setting.

An important difference between [33] and our work is that in the former smoothing is employed for an optimization problem with only here-and-now variables. As explained in Section 2 and Remark 2, dealing with VIs and wait-and-see variables coupled by the shared constraint notably complicates the issues, and therefore makes our development substantially different.

Let us start with a general setting, without considering for now any special structure of the nonsmooth risk measures  $\rho^i$ ; the *AVaR* structure, however, will be important eventually. Let  $\text{VI}(F, C)$  be the variational inequality associated to (4) or to (5), with the risk measure  $R^i$  therein being this  $\rho^i$  for a finite sample space  $\Omega_K$ . For brevity, we shall not define here the operator  $F$  and the set  $C$  formally, as their versions for a finite number of scenarios will be stated in Section 4 (see Proposition 3, which describes the structure). For now, the important point is that  $F$  contains some single-valued continuous components and set-valued components given by subdifferentials  $\partial\rho^i$ , and  $C$  is a closed convex set. Let  $\text{VI}(F_\tau, C)$  be the variational inequality associated to (4) or to (5), with the risk measure  $R^i$  therein being some smooth function  $\rho_\tau^i$  approximating  $\rho^i$ . As a practical computational matter, for some applications (in particular, when many smoothed problems need to be solved to arrive at satisfactory solutions) it can make good sense to solve the problems  $\text{VI}(F_\tau, C)$  approximately, tightening the stopping criterion along the way. It is natural to measure approximation to a solution of  $\text{VI}(F_\tau, C)$  in terms of its *natural residual* [20]. Thus, for  $\tau > 0$  given, we say that a point  $x_\tau \in C$  is an  $\delta(\tau)$ -approximate solution of  $\text{VI}(F_\tau, C)$ , if it holds that

$$\|x_\tau - P_C(x_\tau - F_\tau(x_\tau))\| \leq \delta(\tau). \quad (6)$$

Naturally,  $\delta(\cdot) \geq 0$ , and if  $\delta(\tau) = 0$  then  $x_\tau$  is an exact solution of  $\text{VI}(F_\tau, C)$ .

The first convergence statement below (Theorem 1) can clearly be made more general if we do not refer specifically to VIs associated to (4) or (5). Moreover, certainly various conceptually related results can be found in the literature, as the proof relies on the property known as *gradient consistency* [4, 2]. We note, however, that our setting is that of variational inequalities (and with inexact solutions) rather than optimization, where gradient consistency had been mostly explored. Thus a proof of Theorem 1 is in order. We also mention that, according to Theorem 1, we would have to show continuous convergence and differentiability of our smoothed risk measures, which is not automatic from more general smoothing schemes. Recall that potential sources for nonsmoothness in *AVaR* come not only from the positive-part function, but also from it being a value-function of an optimization problem. Value-functions are not *composite functions*, for which smoothing schemes are commonly designed. Also, in our case continuous convergence is not enough to ensure that the smoothed risk measures we introduce fit the assumptions in general smoothing schemes such as [40, 46, 33]. Remark 2 gives further details related to this discussion.

### Theorem 1 (Convergence of solutions of approximating VIs)

Let  $\{\rho_\tau^i(\cdot)\}$  be a sequence of differentiable convex functions, which converges continuously to the (possibly nonsmooth) risk measure  $\rho^i(\cdot)$ , as  $\tau \rightarrow 0$ . Let  $\text{VI}(F_\tau, C)$  be the VI associated to (4) or to (5), with the risk measure  $R^i(\cdot)$  therein being  $\rho_\tau^i(\cdot)$  for a finite sample space  $\Omega_K$ . Let  $\delta(\cdot)$  be any function such that  $\delta(\tau) \rightarrow 0$  as  $\tau \rightarrow 0$ .

As  $\tau \rightarrow 0$ , every accumulation point of any sequence of any  $\delta(\tau)$ -approximate (in the sense of (6)) solutions of VIs  $(F_\tau, C)$  is a solution of the VI  $(F, C)$ .

*Proof* Since  $\rho_\tau^i(\cdot)$  converges continuously to  $\rho^i(\cdot)$ , the following “gradient consistency property” takes place: for any sequences  $\tau_j \rightarrow 0$ ,  $y_{\tau_j} \rightarrow \bar{y}$  and  $\nabla \rho_{\tau_j}^i(y_{\tau_j}) \rightarrow \bar{w}$ , it holds that  $\bar{w} \in \partial \rho^i(\bar{y})$ . This property then implies that if  $\{x_j\} \rightarrow \bar{x}$  and  $\tau_j \rightarrow 0$ , then for every accumulation point  $\bar{v}$  of the sequence  $\{F_{\tau_j}(x_j)\}$  it holds that  $\bar{v} \in F(\bar{x})$ .

Let now  $\bar{x}$  be an accumulation point of some sequence  $\{x_j\}$  of  $\delta(\tau_j)$ -approximate solutions of VI  $(F_{\tau_j}, C)$  as  $\tau_j \rightarrow 0$ . Without loss of generality (passing onto a further subsequence if necessary), we can assume that  $\{x_j\} \rightarrow \bar{x}$  as  $j \rightarrow \infty$ . Under our assumptions, it is immediate that  $\bar{x} \in C$  and that the sequence  $\{F_{\tau_j}(x_j)\}$  is bounded. Passing yet onto a further subsequence if necessary, we can assume that  $\{F_{\tau_j}(x_j)\}$  converges to  $\bar{v}$ . Then we have that  $\bar{v} \in F(\bar{x})$ .

Since  $x_j$  is an  $\delta(\tau_j)$ -approximate solution of VI  $(F_{\tau_j}, C)$ , the condition (6) means that there exists some element  $e_{\tau_j}$  in the unit ball such that

$$x_j - P_C(x_j - F_{\tau_j}(x_j)) = \delta(\tau_j)e_{\tau_j}.$$

By the classical characterization of the orthogonal projection, this means that

$$x_j - F_{\tau_j}(x_j) - (x_j - \delta(\tau_j)e_{\tau_j}) \in N_C(x_j - \delta(\tau_j)e_{\tau_j}),$$

so that

$$-F_{\tau_j}(x_j) + \delta(\tau_j)e_{\tau_j} \in N_C(x_j - \delta(\tau_j)e_{\tau_j}).$$

As it holds that  $-F_{\tau_j}(x_j) + \delta(\tau_j)e_{\tau_j} \rightarrow \bar{v}$  and  $x_j - \delta(\tau_j)e_{\tau_j} \rightarrow \bar{x}$  as  $j \rightarrow \infty$ , using the fact that the normal cone mapping is outer semicontinuous, we conclude that  $-\bar{v} \in N_C(\bar{x})$ , and so  $0 \in F(\bar{x}) + N_C(\bar{x})$ . This means that  $\bar{x}$  solves the VI  $(F, C)$ , as claimed.  $\square$

The method for defining smoothing approximations of nonsmooth functions depends strongly on the structure of the function at hand. In the rest of this section, we study the case of the nonsmooth risk measure *AVaR*.

By the definition of *AVaR* in (2), we see that there are two sources of nonsmoothness: the positive-part function  $[\cdot]^+$  and the definition of the risk measure through a minimization problem (as a value-function).

**Table 1** Examples of smoothing functions.

| $d(t)$  | $\sigma_\tau(x)$   | $D_1$         | $D_2$         |
|---|--|---------------|---------------|
| $\frac{e^{-t}}{(1+e^{-t})^2}$   | $x + \tau \log(1 + e^{-\frac{x}{\tau}})$   | $\log 2$      | 0             |
| $\frac{2}{(t^2+4)^{\frac{3}{2}}}$   | $\frac{x + \sqrt{x^2 + 4\tau^2}}{2}$   | 1             | 0             |
| $\begin{cases} 1 & \text{if } 0 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases}$                      | $\begin{cases} 0 & \text{if } x < 0 \\ \frac{x^2}{2\tau} & \text{if } 0 \leq x \leq \tau \\ x - \frac{\tau}{2} & \text{if } x > \tau \end{cases}$                                  | 0             | $\frac{1}{2}$ |
| $\begin{cases} 1 & \text{if } -\frac{1}{2} \leq t \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$ | $\begin{cases} 0 & \text{if } x < -\frac{\tau}{2} \\ \frac{1}{2\tau} (x + \frac{\tau}{2})^2 & \text{if }  x  \leq \frac{\tau}{2} \\ x & \text{if } x > \frac{\tau}{2} \end{cases}$ | $\frac{1}{8}$ | 0             |

The nonsmoothness of the positive-part function and its smooth approximations is a well-studied issue. We start with recalling the main elements of the general smoothing framework of [3]; see also [20, Chapter 11.8]. To construct smooth approximations of the

$[\cdot]^+$  function, one starts with a nonnegative piecewise continuous function  $d$  with finite number of pieces, such that

$$\int_{-\infty}^{+\infty} d(t)dt = 1 \quad \text{and} \quad \int_{-\infty}^{+\infty} |t|d(t)dt < \infty.$$

Then, for  $\tau > 0$ , the approximating function  $\sigma_\tau : \mathbb{R} \rightarrow \mathbb{R}$  is given by

$$\sigma_\tau(x) := \int_{-\infty}^x \int_{-\infty}^y \frac{1}{\tau} d\left(\frac{t}{\tau}\right) dt dy. \quad (7)$$

This function is well defined and it is convex, because  $d$  is nonnegative ( $\sigma_\tau$  becomes strictly convex when  $d$  is positive).

Table 1 gives some examples of smoothing functions, their generators  $d$ , and the constants

$$D_1 := \int_{-\infty}^0 |t|d(t)dt \quad \text{and} \quad D_2 := \left[ \int_{-\infty}^{+\infty} td(t)dt \right]^+ \quad (8)$$

useful to bound the closeness of  $\sigma_\tau$  to the positive-part function.

In particular, we recall the following facts from [3, Proposition 2.2]:

- the function  $\sigma_\tau$  is nondecreasing, convex, and continuously differentiable (if  $d$  is  $k$ -times continuously differentiable, then  $\sigma_\tau$  is  $(k+2)$ -times continuously differentiable);
- for every  $x \in \mathbb{R}$  it holds that

$$-D_2 \tau \leq \sigma_\tau(x) - [x]^+ \leq D_1 \tau. \quad (9)$$

Given a function  $\sigma_\tau(\cdot)$  as in (7), consider

$${}^\tau\text{AVaR}_\varepsilon(Y) := \min_{u \in \mathbb{R}} \left\{ u + \frac{1}{1-\varepsilon} \mathbb{E} \left( \sigma_\tau(Y_\omega - u) \right) \right\}, \quad (10)$$

defined for  $Y \in L_1 := L_1(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R})$ . The functions  $\rho_\tau$  and  $\rho$  in Section 3 correspond, respectively, to  ${}^\tau\text{AVaR}_\varepsilon$  and  $\text{AVaR}_\varepsilon$ .

Note again that smoothness of  ${}^\tau\text{AVaR}_\varepsilon$  is *not* automatic from smoothness of  $\sigma_\tau$ , as we have a value-function (rather than a composite function, say). That  ${}^\tau\text{AVaR}_\varepsilon$  is indeed (twice) differentiable will be shown in Proposition 2. Before, we establish some basic properties including, in particular, the continuous convergence of smoothed risk measures to  $\text{AVaR}$ . The smoothed functions being risk measures on their own is also clearly meaningful.

**Proposition 1** ( *${}^\tau\text{AVaR}_\varepsilon$  is a well-defined risk measure continuously converging to  $\text{AVaR}_\varepsilon$* )

*Given a function  $\sigma_\tau(\cdot)$  as in (7) and  ${}^\tau\text{AVaR}_\varepsilon$  in (10), the following hold:*

- (i) *The function  ${}^\tau\text{AVaR}_\varepsilon$  is well defined. In particular, for any  $\tau > 0$  and all  $Y \in L_1$  there exists  $u^\tau(Y)$ , the “smoothed value-at-risk”, realizing the minimum in (10).*
- (ii) *The function  ${}^\tau\text{AVaR}_\varepsilon$  is a risk measure.*
- (iii) *For  $D_1$  and  $D_2$  from (8) and all  $Y \in L_1$ , it holds that*

$$\left| \text{AVaR}_\varepsilon(Y) - {}^\tau\text{AVaR}_\varepsilon(Y) \right| \leq \frac{\max(D_1, D_2)\tau}{1-\varepsilon}.$$

- (iv) *The family of risk measures  ${}^\tau\text{AVaR}_\varepsilon$  converges continuously to the risk measure  $\text{AVaR}_\varepsilon$  as  $\tau$  tends to zero, i.e., whenever a sequence  $\{y_\tau\}$  converges to  $y$  as  $\tau \rightarrow 0$ , the sequence  $\{{}^\tau\text{AVaR}_\varepsilon(y_\tau)\}$  converges to  $\text{AVaR}_\varepsilon(y)$ .*

*Proof* Given  $Y \in L_1$ , we exhibit that the  $AVaR_\varepsilon$ -minimand  $T := u + \frac{1}{1-\varepsilon} \mathbb{E} \left( [Y_\omega - u]^+ \right)$  is coercive in the  $u$ -variable, because

$$T \geq \begin{cases} u & \text{if } u \geq 0, \\ \left(1 - \frac{1}{1-\varepsilon}\right)u + \frac{1}{1-\varepsilon} \mathbb{E} \left( Y_\omega \right) & \text{if } u < 0. \end{cases}$$

As a result,

$$-\frac{D_2 \tau}{1-\varepsilon} + T \leq u + \frac{1}{1-\varepsilon} \mathbb{E} \left( \sigma_\tau(Y_\omega - u) \right) \leq \frac{D_1 \tau}{1-\varepsilon} + T. \quad (11)$$

So the  $\tau AVaR_\varepsilon$ -minimand is coercive too, and there exists  $u^\tau(Y)$  realizing the minimum in (10), as claimed.

The second item now follows from the risk measure definition in [12, Chapter 6.3]. In particular, in addition to being convex, the smoothed function satisfies the properties of translation-equivariance:

$$\tau AVaR_\varepsilon(Y + a) = a + \tau AVaR_\varepsilon(Y) \quad \text{and} \quad \tau AVaR_\varepsilon(Y) \geq \tau AVaR_\varepsilon(Y') \quad \text{for all } a \in \mathbb{R},$$

and monotonicity:  $Y_\omega \geq Y'_\omega$  a.e.  $\omega \in \Omega$ .

Regarding the third assertion, we take the infimum values over  $u$  in the chain of inequalities (11) to obtain

$$-\frac{D_2 \tau}{1-\varepsilon} + AVaR_\varepsilon(Y) \leq \tau AVaR_\varepsilon(Y) \leq \frac{D_1 \tau}{1-\varepsilon} + AVaR_\varepsilon(Y).$$

Finally, to show the last statement, note that the previous item implies that the sequence  $\{\tau AVaR_\varepsilon\}$  converges uniformly to  $AVaR_\varepsilon$ . Since the limit function  $AVaR_\varepsilon$  is itself a continuous function, [43, Thm. 5.43] implies that the convergence is in fact continuous.  $\square$

*Remark 2 (On some other smoothing schemes of nonsmooth functions)*

A very general smoothing scheme is given in [40], which covers also nonconvex functions. It was subsequently used in a number of works related to smoothing techniques; e.g., [46, 33].

It can be seen that the generality of [40] (related to being able to handle nonconvexity) is not suitable for our case, because assumptions therein allow functions  $\sigma_\tau$  for which the objective function of the optimization problem in (10) would be unbounded below. In this case, the definition (10) returns  $\tau AVaR_\varepsilon = -\infty$ , i.e., the smoothed risk measure is not even well-defined.  $\square$

*Remark 3 ( $\tau AVaR_\varepsilon$  is not coherent)*

As shown in Proposition 1(ii), our smoothed functions are risk measures. For the risk measure to be coherent, it must also be positively homogeneous, i.e, it should hold that

$$\rho^\tau(tY) = t\rho^\tau(Y) \quad \text{for any } t > 0 \text{ and } Y \in L_1.$$

Since the smoothing  $\sigma_\tau$  does not satisfy this property in general (c.f. Table 1),  $\tau AVaR_\varepsilon$  is not coherent.

Nevertheless, all the smoothing functions in Table 1 satisfy for  $t > 0$  and  $x \in \mathbb{R}$  the relation  $\sigma_\tau(tx) = t\sigma_{\tau/t}(x)$ . As a result, the  $\tau AVaR_\varepsilon$  risk measure satisfies the property

$$\rho^\tau(tY) = t\rho^{\tau/t}(Y) \quad \text{for any } t > 0 \text{ and } Y \in L_1,$$

which gives coherence in the limit, i.e., for  $\tau = 0$  and, hence, for the  $AVaR_\varepsilon$  (a well-known result). We mention that the *threshold risk measures* from [8] propose (in a different context) a relation similar to the above, as a substitute for coherence.  $\square$

By Proposition 1, we see that to deduce from Theorem 1 convergence of our approximating approach, we still need to show that the smoothed risk measures in (10) are differentiable. Note that differentiability of the smoothed positive-part function is immediate, of course, but there is still the issue of the risk measures being value-functions of an optimization problem. We now state the differentiability properties of smoothed risk measures for a finite sample space  $\Omega_K$ .

**Proposition 2 (Differentiability properties of  ${}^\tau\text{AVaR}_\varepsilon$ )**

Let a function  $\sigma_\tau(\cdot)$  be given as in (7),  ${}^\tau\text{AVaR}_\varepsilon$  be defined by (10), and let  $u^\tau(Y)$  be a minimizer in (10). Suppose  $\Omega = \Omega_K = \{\omega_1, \dots, \omega_K\}$  is finite and let  $p_k$  be the probability of an event  $\omega_k \in \Omega$  for  $k = 1, \dots, K$ . The following statements hold:

(i) The function  ${}^\tau\text{AVaR}_\varepsilon$  is differentiable, with its gradient given by

$$\nabla {}^\tau\text{AVaR}_\varepsilon(Y) = \frac{1}{1-\varepsilon} \left( p_k \sigma'_\tau(Y_{\omega_k} - u^\tau(Y)) \right)_{k=1}^K.$$

(ii) If  $\sigma'_\tau(\cdot) > 0$  then the function  ${}^\tau\text{AVaR}_\varepsilon$  is twice differentiable, with the second derivative given by

$$\frac{\partial^2 [{}^\tau\text{AVaR}_\varepsilon(Y)]}{\partial Y_{\omega_i} \partial Y_{\omega_j}} = \frac{p_{\omega_i} \sigma''_\tau(Y_{\omega_i} - u^\tau(Y))}{1-\varepsilon} \left[ \delta_{ij} - \frac{p_j \sigma''_\tau(Y_{\omega_j} - u^\tau(Y))}{\mathbb{E}[\sigma''_\tau(Y - u^\tau(Y))]} \right],$$

where  $\delta_{ij}$  denotes the Kronecker's delta function and  $i, j \in \{1, \dots, K\}$ .

*Proof* In the finite-dimensional setting,  ${}^\tau\text{AVaR}_\varepsilon$  is a value-function

$${}^\tau\text{AVaR}_\varepsilon(Y) = \min_u g(Y, u) = g(Y, u^\tau(Y))$$

for  $g(Y, u) := u + \frac{1}{1-\varepsilon} \mathbb{E}(\sigma_\tau(Y_{\omega_k} - u))$ . Since by Proposition 1(i) the minimum in (10) is attained “at finite distance” and  $g$  is differentiable, the value-function in question is differentiable by Corollary 4.5.3 in [24, Ch. VI], and

$$\frac{\partial {}^\tau\text{AVaR}_\varepsilon(Y_{\omega_k})}{\partial Y_{\omega_k}} = \frac{\partial g(Y_{\omega_k}, u^\tau(Y))}{\partial Y_{\omega_k}} = \frac{p_k}{1-\varepsilon} \sigma'_\tau(Y_{\omega_k} - u^\tau(Y)).$$

We next analyze the existence and calculus of second derivatives. Given a minimizer  $u^\tau(\bar{Y})$  corresponding to  ${}^\tau\text{AVaR}_\varepsilon(\bar{Y}) = g(\bar{Y}, u^\tau(\bar{Y}))$ , it holds that

$$\frac{\partial g}{\partial u}(\bar{Y}, u^\tau(\bar{Y})) = 0.$$

By the convexity of  $\sigma_\tau$ ,

$$\frac{\partial^2 g}{\partial u^2}(Y, u) = \frac{1}{1-\varepsilon} \mathbb{E}(\sigma''_\tau(Y_k - u)) \geq 0.$$

As  $\frac{\partial^2 g}{\partial u^2}(\bar{Y}, u^\tau(\bar{Y})) > 0$  if  $\sigma_\tau(\cdot) > 0$ , by the Implicit Function Theorem there exist neighborhoods  $V$  of  $\bar{Y}$  and  $U$  of  $u^\tau(\bar{Y})$  and a differentiable function  $u : V \rightarrow U$  such that, among other properties, for all  $Y \in V$  it holds that

$$\frac{\partial g}{\partial u}(Y, u(Y)) = 0. \tag{12}$$

By the convexity of  $g(Y, \cdot)$ , we have that  ${}^\tau\text{AVaR}(Y) = g(Y, u(Y))$ ;  ${}^\tau\text{AVaR}$  is differentiable and

$$\nabla {}^\tau\text{AVaR}(Y) = \frac{\partial g}{\partial Y}(Y, u(Y)) + \frac{\partial g}{\partial u}(Y, u(Y))\nabla u(Y) = \frac{\partial g}{\partial Y}(Y, u(Y)).$$

This shows that  ${}^\tau\text{AVaR}$  is twice differentiable. Differentiating (12) we obtain

$$\frac{\partial^2 g}{\partial Y \partial u}(Y, u(Y)) + \frac{\partial^2 g}{\partial u^2}(Y, u(Y))\nabla u(Y) = 0,$$

implying the explicit formula for  $\nabla u(Y)$  as  $\frac{\partial^2 g}{\partial u^2}(Y, u(Y)) > 0$  on the relevant neighborhood. Substituting the expression for  $\nabla u(Y)$  into

$$\nabla^2[{}^\tau\text{AVaR}](Y) = \frac{\partial^2 g}{\partial Y^2}(Y, u(Y)) + \frac{\partial^2 g}{\partial u \partial Y}(Y, u(Y))\nabla u(Y),$$

we obtain the needed formula in item (ii).  $\square$

Our final result of this section states that, for both the GNEP model (4) and the MCP model (5), accumulation points of (approximate) solutions of the smoothed VIs solve the original VI formulations with nonsmooth measures.

**Theorem 2 (Convergence of solutions of VIs with smoothed risk measures)**

Let  $VI(F_\tau, C)$  be the (single-valued operator) variational inequality associated to (4) or to (5), with the risk measure  $R^i(\cdot)$  therein defined by

$$\rho_\tau^i(\cdot) = (1 - \kappa^i)\mathbb{E}(\cdot) + \kappa^i {}^\tau\text{AVaR}_{\varepsilon^i}(\cdot), \quad \kappa^i \in [0, 1],$$

for a finite sample space  $\Omega_K$ . Let  $\delta(\cdot)$  be any function such that  $\delta(\tau) \rightarrow 0$  as  $\tau \rightarrow 0$ .

Then, as  $\tau \rightarrow 0$ , every accumulation point of any sequence of any  $\delta(\tau)$ -approximate solutions of VIs  $(F_\tau, C)$  solves the (set-valued operator) VI  $(F, C)$ , associated to (4) or (5), with the risk measure  $R^i(\cdot)$  therein given by

$$\rho^i(\cdot) = (1 - \kappa^i)\mathbb{E}(\cdot) + \kappa^i \text{AVaR}_{\varepsilon^i}(\cdot).$$

*Proof* The assertion follows combining Theorem 1 with Propositions 1 and 2.  $\square$

#### 4 Assessing the practical value of the smoothing approach

We assume that all the investment and production functions are convex. Recall that the coupling constraint  $h_k(\cdot)$  in the set  $S_k$  is affine. In both models and for each player  $i = 1, \dots, N$ , we take endogenous sets of polyhedral form  $X_k^i$  for  $k = 1, \dots, K$ . We make the natural assumptions that the sets in question are bounded and there is relatively complete recourse.

In Section 3 we outlined our proposal on how to deal with the set-valued VI operators associated to (4) or (5) by the smoothing approximation technique. Another possibility is to reformulate the positive-part function in (3) in the following (standard) way, introducing *additional* variables and constraints:

$$\text{AVaR}_\varepsilon(Y_K) = \begin{cases} \min u + \frac{1}{1 - \varepsilon} \mathbb{E}[T_K] \\ \text{s.t. } u \in \mathbb{R}, \text{ and, for } k = 1, \dots, K : \\ \quad T_k \geq Y_k - u \text{ and } T_k \geq 0. \end{cases} \quad (13)$$

In this section we compare, theoretically and computationally:

- VI obtained from the GNEP (4) with smoothed risk measure  $R = {}^\tau AVaR_\varepsilon$  defined in (10).
- VI obtained from GNEP (4) reformulating the original risk measure  $R = AVaR_\varepsilon$  using extra variables and constraints, as outlined in (13).

#### 4.1 VI derivation and existence results

Some computational drawbacks resulting from the reformulation (13) are commented in Section 4.2 below. We mention here that the constraint  $T_k^i \geq c_k^i - u^i$  couples *all* the agents' variables, while *only* appearing in the  $i$ th agent's problem. Note also that this changes the nature of the game, which is no longer of Generalized Nash type (the constraint in  $S_k$  is individual to each player, no longer shared). For this reason, to obtain a VI formulation for computing a variational equilibrium using (13), in item (ii) of Proposition 3 below, and in each player's problem, the constraints in question have to be "lifted" into the VI operator introducing Lagrange multipliers  $\eta_k^i$ . And there is actually yet another reason that requires lifting those constraints into the VI operator: grouping those constraints for all the agents in a VI formulation may give nonconvex feasible sets. This is because while the cost functions  $c_k^i$  are convex in the variables  $q_k^i$ , it is not uncommon for them to be nonconvex in the entire variable  $q_k$ . In that case, combining those constraints would give a problem with a nonconvex set, with all the resulting difficulties.

We now state the VIs given by our approach and by the reformulation (13). Without loss of generality, we shall assume that  $\kappa^i = 1$  in (3), so that the objective functions in the agents' problems have simpler expressions.

#### Proposition 3 (Two single-valued VIs for GNEP)

(i) For the smoothed stochastic GNEP based on  ${}^\tau AVaR_\varepsilon$ , with the agents' problems

$$\left\{ \begin{array}{l} \min I^i(z^i) + {}^\tau AVaR_{\varepsilon^i}(c_K^i(q_K^i, \bar{q}_K^{-i})) \\ \text{s.t. } z^i \in \mathbb{R}^{n^i}, q_K^i \in \mathbb{R}^{m^K}, \\ \text{and, for } k = 1, \dots, K : \\ (z^i, q_k^i) \in X_k^i, \\ h_k^i(q_k^i) + \sum_{i \neq j=1}^N h_k^j(\bar{q}_k^j) \leq 0, \end{array} \right. \quad (14)$$

for  $i = 1, \dots, N$ , the associated VI  $(F_\tau, C)$  has the following structure:

(a) The VI operator is  $F_\tau(z, q_K) = \prod_{i=1}^N F_\tau^i(z^i, q_K)$ , with components

$$F_\tau^i(z^i, q_K) := \left( \begin{array}{c} \nabla I^i(z^i) \\ \left( \nabla_{q_k^i} [{}^\tau AVaR_{\varepsilon^i}(c_k^i(q_k^i, \bar{q}_k^{-i}))] \right)_{k=1}^K \end{array} \right).$$

(b) The VI feasible set is

$$C := \prod_{k=1}^K \left( \prod_{i=1}^N X_k^i \cap S_k \right),$$

$$\text{where } S_k := \left\{ (z^i, q_k^i)_{i=1}^N : \sum_{i=1}^N h_k^i(q_k^i) \leq 0 \in \mathbb{R}^{m^0} \right\}.$$



(ii) For the stochastic GNEP reformulating  $AVaR_\varepsilon$  via additional variables and constraints, with agents' problems given by

$$\left\{ \begin{array}{l} \min I^i(z^i) + u^i + \frac{1}{1-\varepsilon^i} \mathbb{E}[T_K^i] \\ \text{s.t. } z^i \in \mathbb{R}^{n^i}, q_k^i \in \mathbb{R}^{m^i K}, u^i \in \mathbb{R}, T_K^i \in \mathbb{R}^K, \\ \text{and, for } k=1, \dots, K: \\ (z^i, q_k^i) \in X_k^i, \\ h_k^i(q_k^i) + \sum_{\substack{j=1 \\ i \neq j}}^N h_k^j(\bar{q}_k^j) \leq 0, \\ T_k^i \geq c_k^i(q_k^i, \bar{q}_k^{-i}) - u^i, T_k^i \geq 0, \end{array} \right. \quad (15)$$

for  $i=1, \dots, N$ , the associated VI  $(\tilde{F}, \tilde{C})$  has the following structure:

(a) The VI operator is

$$\tilde{F}(z, q_K, u, T_K, \eta_K) = \prod_{k=1}^K \prod_{i=1}^N \tilde{F}_k^i(z^i, q_k, u^i, T_k^i, \eta_k^i),$$

with components

$$\tilde{F}_k^i(z^i, q_k, u^i, T_k^i, \eta_k^i) := \begin{pmatrix} \nabla I^i(z^i) \\ \eta_k^i \nabla_{q_k^i} c_k^i(q_k^i, \bar{q}_k^{-i}) \\ 1 - \eta_k^i \\ \frac{p_k}{1-\varepsilon^i} - \eta_k^i \\ -c_k^i(q_k^i, \bar{q}_k^{-i}) + u^i + T_k^i \end{pmatrix}.$$

(b) The VI feasible set is

$$\tilde{C} := \begin{pmatrix} C \\ \prod_{i=1}^N \mathbb{R} \\ \prod_{k=1}^K \prod_{i=1}^N \mathbb{R}_+ \\ \prod_{k=1}^K \prod_{i=1}^N \mathbb{R}_+ \end{pmatrix}.$$

*Proof* Direct, by writing the optimality conditions for the  $N$  problems of the agents.  $\square$

Existence of an equilibrium for risk-averse GNEPs is a difficult subject. Existence was shown for a risk-averse Nash game with a market for risk (but no shared constraints) in [39, Sec. 5]. When there are only finitely many scenarios for the random variable ( $\Omega$  is finite), both [41] and [27] provide sufficient conditions for the existence of a variational equilibrium based on coercivity of the VI operator, following [20]. The next theorem states an important consequence of our approach, that ensures existence of an equilibrium for our models, with an argument which in fact makes use of our smoothing approach itself.

### Theorem 3 (Existence of solutions in stochastic GNEP with risk aversion)

Let VI  $(F, C)$  be associated to the stochastic GNEP (4) with the risk measure  $R^i(\cdot)$  therein given by  $\rho^i(\cdot) = AVaR_{\varepsilon^i}(\cdot)$  for a finite sample space  $\Omega_K$ , i.e., let the set  $C$  be defined in

Proposition 3 (i(b)) and let the VI operator be  $F(z, q_K) = \prod_{i=1}^N F^i(z^i, q_K)$  with components

$$F^i(z^i, q_K) := \begin{pmatrix} \nabla I^i(z^i) \\ \left( \partial_{q_k^i} [AVaR_{\varepsilon^i}(c_k^i(q_k^i, \bar{q}_k^{-i}))] \right)_{k=1}^K \end{pmatrix}.$$

If the sets  $X_k^i$ ,  $i = 1, \dots, N$ ,  $k = 1, \dots, K$ , are convex and compact, then  $VI(F, C)$  has a solution.

*Proof* By compactness of the sets  $X_k^i$ , the set  $C$  is compact. Further,  $F_\tau$  defined in Proposition 3 (i(a)) is a single-valued continuous mapping. Hence, the  $VI(F_\tau, C)$  has a solution [20, Corollary 2.2.5], for each  $\tau > 0$ . Belonging to  $C$ , any sequence of such solutions (indexed by  $\tau$ ) is bounded, and thus has accumulation points. By Theorem 2, any of those accumulation points (which exist) solves  $VI(F, C)$ . Hence, the latter VI has a solution.  $\square$

Existence for the MCP formulation (5) can be obtained similarly, recalling the difference in the argument of the risk measure and in the treatment of the shared constraint in  $S_k$ .

#### 4.2 Computational comparison with the smooth reformulation

We first discuss some structural properties of the two variational formulations in Proposition 3. At first sight, because the original risk measure is employed in (15) (though via its reformulation!) and the resulting problem is single-valued, it might seem that the  $VI(\tilde{F}, \tilde{C})$  in item (ii) might be preferable. This first impression neglects several challenges that complicate the solution of the reformulation  $VI(\tilde{F}, \tilde{C})$ , especially when  $K$  is large (to better represent uncertainty, having more events in the sampling space is preferable). In particular:

- (i) As the number of scenarios grows, the additional number of variables and constraints induced by the reformulation poses a clear numerical drawback: it adds  $(2K + 1)N$  variables to the VI, and  $2KN$  nonnegativity constraints. This is already quite significant in itself, as  $K$  grows.
- (ii) Note that the shared constraint in (1) couples all agents  $i = 1, \dots, N$  for *each realization*  $k$ . In contrast, the additional relations involving the terms

$$T_k^i - c_k^i(q_k^i, q_k^{-i}) + u^i, \quad i = 1, \dots, N, \quad k = 1, \dots, K,$$

couple also *all the realizations*  $k = 1, \dots, K$  via the value-at-risk variable  $u^i$ .

The latter also corresponds to requiring that  $(z, q, u, T) \in \prod_{i=1}^N \prod_{k=1}^K \tilde{S}_k^i$ , where

$$\tilde{S}_k^i := \left\{ (z, q_k, u^i, T_k^i) : \tilde{h}_k^i(q_k, u^i, T_k^i) := c_k^i(q_k^i, q_k^{-i}) - u^i - T_k^i \leq 0 \right\}.$$

The set  $\tilde{S}_k^i$  can be nonconvex if the function  $c_k^i$  is nonconvex on the entire variable  $q_k$  (this is the case of agent “0” in the game employed for numerical experiments below). Thus, the corresponding formulation may have this underlying (undesirable) nonconvex feature of the feasible set, even if it is made somewhat implicit by “lifting” the constraint in question into the VI operator (using a Lagrange multiplier).

- (iii) Another possibly negative impact on the numerical performance is that the second block of components in  $\tilde{F}_k^i$  becomes nonlinear on variables  $q_k$  and  $\eta_k^i$ . Even in the simple setting with quadratic costs, with linear gradients  $\nabla c_k^i$ , the corresponding bilinear relation presents a computationally undesirable feature.
- (iv) Finally, we now explain how the fact of *having wait-and-see variable complicates solving the AVaR reformulation*. As mentioned in Remark 2.4, the work [27] also deals with risk-averse GNEPs. To handle AVaR’s nonsmoothness, the authors employ the usual reformulation with more variables and constraints, like in Proposition 3(ii). A modeling

issue that makes the VI in that work easier to handle than ours is the following. In [27], the *AVaR* only involves here-and-now variables that are *endogenous*, there are no other agents' variables in the *AVaR* argument (see equation (4) in [27]). This corresponds to having the costs  $c_i$  depending only on  $q_i$ . The difference is crucial, as the constraints resulting from the *AVaR* reformulation will *not couple* along scenarios the agents' decisions (see page 1104 in [27]). The reformulated Nash game remains a GNEP (with an increased number of variables and constraints, of course) that the authors solve by means of a distributed gradient projection scheme with good scalability properties for the considered problems, thanks to the cutting-plane method employed for the projection step. In our setting, this approach is not directly applicable, for the reasons above.

#### 4.2.1 Benchmark specifications

To compare computational issues induced by the two VI formulations in Proposition 3, we consider the following simplification of the gas network model discussed in Section 5 below:

- The game has three players: producers 1 and 2 and the agent 0, the latter representing the consumers. Their wait-and-see decisions are  $q_k^1$ ,  $q_k^2$  and  $q_k^0$  for  $k = 1, \dots, K$ , the number of scenarios. The here-and-now variables  $z^1$  and  $z^2$  limit the production capacity; for example, as  $0 \leq q_k^1 \leq z^1$ . There is no variable  $z^0$ , and  $q_k^0$  corresponds to demand left unsatisfied.
- Agent 0 aims at minimizing the deficit in demand by means of an inverse demand function, as in Remark 1 (deficit is discouraged by setting a very low price at low production levels). Producers 1 and 2 minimize investment and production costs.
- The product demand follows a normal process with mean 90, and 5 as standard deviation.

The model was coded in Python, using the Matlab interface of PATH to solve the VIs with tolerance  $10^{-6}$ . The runs were performed on a PC operating under Ubuntu 12.04 LTS with Intel Core i7-2600 3.40GHz  $\times$  8 processors and 15GB of memory.

We ran 10 random instances of the game for each of the following sizes of  $\Omega_K$ :

$$\Omega_K \text{ has } K \text{ scenarios, for } K \in \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}.$$

#### 4.2.2 Performance of the smoothing approach

Our solution method solves a sequence of smoothed VIs, indexed by  $j = 0, 1, 2, \dots$ , using for the positive-part function the approximation  $\sigma_\tau(x) = \frac{x + \sqrt{x^2 + 4\tau^2}}{2}$  (the second option in Table 1). Starting with  $\tau_0 = 10^{-3}$ , we set  $\tau_{j+1} = 0.5\tau_j$ , and increase  $j$  until certain stability is observed. Specifically, the process stops when consecutive values of *all* the components of the decision variables are sufficiently close. I.e., for a decision variable called, say,  $\bar{x}$ , it holds that

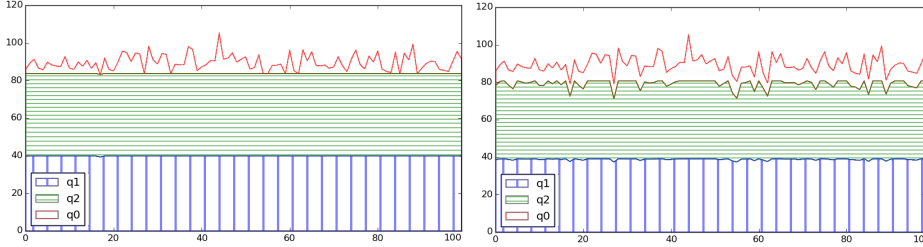
$$\frac{|\bar{x}_{j+1} - \bar{x}_j|}{\max(1, |\bar{x}_{j+1}|)} \leq 0.01, \quad (16)$$

and the same is valid for all the components of all the decision variables.

We ran a first set of experiments with quadratic generation costs and risk aversion parameter  $\varepsilon = 0.25$  for all players. The average solving time was 45 seconds; it took less than five VI solves for our smoothing method to trigger the stopping test ( $j \leq 5$  in (16)) on average, i.e., taking the mean over all the instances and all the considered scenarios.

The excellent quality of the primal and dual solutions obtained with the  $\mathcal{AVaR}$ -smoothed game (14) was confirmed by the fact that in all the runs the final values were (numerically) identical to the “exact” solutions obtained with the  $AVaR$ -reformulation (15).

For the  $\mathcal{AVaR}$  solution and the three players ( $i = 0, 1, 2$ ), Figure 1 shows the value of variables  $(q_i^k)_{k=1}^{K=100}$  taking  $\varepsilon = 0.75$  and  $\varepsilon = 0.25$  in (14) (left and right, respectively).

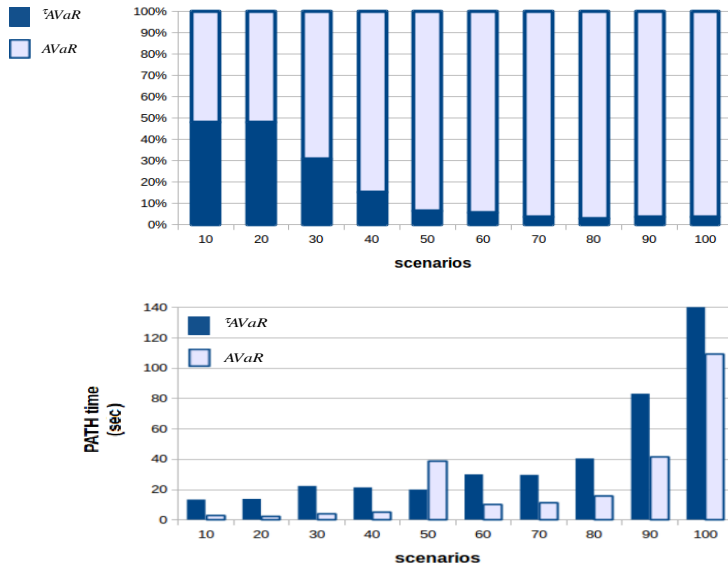


**Fig. 1** Players behaviour with low (left) and high (right) levels of risk aversion

We observe a clear impact of the risk aversion factor: on the left graph, agents are relatively risk-neutral, and their production is practically the same for all scenarios, because the variation of demand is entirely absorbed by the consumers, in the form of a deficit,  $q_0^k$ . By contrast, on the right graph, agents are more risk averse and their production follows closely the demand profile. The higher sensitivity of player 2 (horizontal lines in the graph) is explained by this player having higher generation costs. For these experiments, both the  $\mathcal{AVaR}$ -methodology and solving (15) ran in less than 50 seconds, averaging over all instances and scenarios.

To better assess the impact of the smoothing approach on computing times, we created a second set of instances with nonquadratic convex generation costs. This yields highly nonlinear VI operators in Proposition 3. We first note that the stopping test for the smoothing method was triggered after six VI solves on average:  $j \leq 6$  in (16). The second observation is a noticeable difference in the total solution time: with our approach the average time spent in each scenario was about half a second, while the  $AVaR$  reformulation needed more than eight seconds per scenario. We note that the above refers to total time, i.e., time required to build each VI from the input problem data, and the PATH solution time. In addition, we computed the PATH time separately.

Figure 2 shows these times for both approaches, as a function of the number of scenarios. The top graph in Figure 2 illustrates the increasing time that the  $AVaR$ -reformulation needs, as (15) uses more and more scenarios. For  $K = 10$  and 20, both approaches take approximately the same total time (the length of the darker and lighter color bars are similar). By contrast, for  $K \geq 50$ , the reformulation takes at least 9 times longer than our smoothing approach. The bottom graph makes clear that the reason for this difference in performance is not the PATH time, which is actually better for the reformulation (though becomes quite comparable as  $K$  grows), but the time required to mount the VI models. Except for the case of  $K = 50$ , we see that PATH solves (15) faster than the sequence of (6 in average) smoothed VIs (14) (the case  $K = 50$ , which is out of the trend, should be thought of as an example of the importance of running a good number of realizations when dealing with stochastic



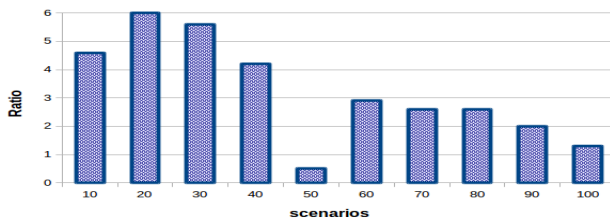
**Fig. 2** Total solution time and PATH time, for a sequence of smoothed  $\varepsilon AVaR$  problems (dark color) and for reformulated  $AVaR$  problems (lighter color).

data)). But the time taken to build larger reformulation models (15) becomes eventually a problem, as  $K$  grows.

As a complement of information, for each  $K$  we computed the ratios

$$\mathcal{R}_K := \frac{\text{PATH time to solve a sequence of smoothed VIs derived from (14) using } \varepsilon AVaR}{\text{PATH time to solve one VI derived from } AVaR \text{ reformulation (15)}}$$

and obtained the values  $\mathcal{R}_K = \{4.60, 6.00, 5.60, 4.20, 0.50, 2.90, 2.60, 2.60, 2.00, 1.30\}$ , displayed graphically in Figure 3. Taking for example the case with  $K = 90$  scenarios, which



**Fig. 3** PATH time ratios between solving VI with  $AVaR$  reformulation and a sequence of smoothed VIs using  $\varepsilon AVaR$

required seven VI solves in the smoothing approach, we see that solving seven smoothed VIs (14) demanded only twice the time required for the single reformulated VI (15). Solving faster one smoothed VI is partly explained by its smaller size. But we also observed that PATH was always much faster solving consecutive smoothed VIs, after having solved the first one, thanks to a good use of warm starts.

In terms of scalability, the trend is clear. At least for our test problems, the single VI for the *AVaR* reformulation will become unsolvable much earlier than VIs in our approach. Note also that our approach allows for decomposition, and even parallelization (see [31]), thus pushing the boundary even further. In fact, one of the subjects of future research would be precisely combining Benders decomposition [14] along scenarios with Dantzig-Wolfe decomposition for the players [31]. Using the latter decomposition technique is discussed in the next section.

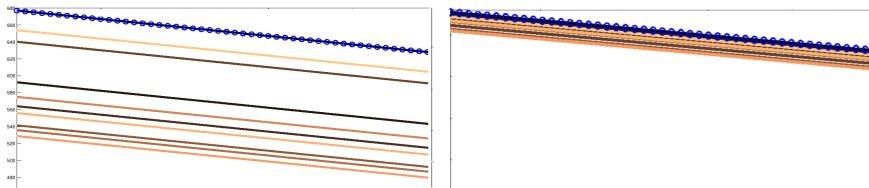
## 5 Assessment on the European Natural Gas Network

In this section we use data from a real-world problem, the gas network specified in [23], with two particular goals. First, we analyze the agents' strategies regarding aversion to risk under different conditions of market volatility. Second, as a proof of concept, we show that further advantage is to be gained with our approach applying decomposition. At this time this is just a preliminary assessment, as after decomposing as in [31], one could also parallelize, an option not yet implemented. Moreover, our formulation allows for Benders decomposition [31] along scenarios as well. This is another subject for future implementations of our proposals, currently underway.

The gas network [23] considers agents of several types: producers, traders, liquefiers, regasifiers, storage and pipeline operators. This is an energy-only market (there is no capacity market; so the investment variables and the corresponding objects disappear from the general model described above). Consumers are represented by the inverse demand function, a modelling referred to as "implicit" in [32]; see also Remark 1. As there are no here-and-now variables, the risk-neutral and risk-averse MCP models are equivalent in this case: our numerical results concern the GNEP model only.

### 5.1 Information about Solvers, Problem and Data

We use data of a "subnetwork" in [23], with 3 producers (Russia, the Netherlands and Norway), 1 trader, 1 regasifier, 1 liquefier, and 1 storage operator (the Netherlands, Belgium, Nigeria, and France, respectively), and the pipelines linking the agents. The deterministic version of the model has 44 variables and 22 constraints.



**Fig. 4** Inverse demand function scenarios for France, high and low volatility (left and right)

To generate the stochastic demand, we created random reductions in the deterministic values of the intercept of the inverse demand function, sampling from a uniform distribution and multiplying by a factor representing volatility. The left and right graphs in Figure 4 show scenarios of the inverse demand function for France, for higher and lower volatility, respectively. The deterministic function is plotted with circles. We used the smoothing parameters in Section 4.2.2,  $\varepsilon = 0.1$ , and  $\kappa^i = 0.1$  for all the agents, except the producers, for which we let  $\kappa^i = 0.75$  (producers are more risk averse).

## 5.2 Impact of Volatility and Risk Aversion

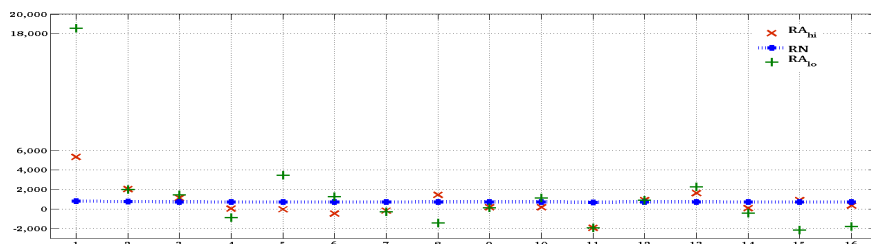
We solved the GNEP model for an increasing number of scenarios, and made 5 runs for every cardinality. Table 2 reports the average solution time in seconds (failures are denoted with \*\*\*\*\*) for the risk neutral (RN), and the risk-averse game (RA) models. A suffix *lo* or *hi* refers to the data with low or high volatility represented by Figure 4.

**Table 2** Average CPU time in seconds - Low and High Volatility

| $K$ | RN <sub>lo</sub> | RN <sub>hi</sub> | RA <sub>lo</sub> | RA <sub>hi</sub> |
|-----|------------------|------------------|------------------|------------------|
| 1   | 0.0              | 0.0              | 2.3              | 2.3              |
| 2   | 0.0              | 0.0              | 11.8             | 5.8              |
| 4   | 0.0              | 0.1              | 77.7             | 25.0             |
| 8   | 0.0              | 0.1              | 185.8            | 94.4             |
| 16  | 0.3              | 0.3              | 541.3            | 326.1            |
| 32  | 0.8              | 0.9              | *****            | *****            |
| 64  | 1.4              | 1.9              | *****            | *****            |

We observe in Table 2 that (naturally) for all the instances the risk-neutral model ( $\kappa^i = 0$  in (3)) is solved in less than one second. The interest of the case with just one scenario ( $K = 1$ ) is to check if all the models give the same solution. This was the case for all of our runs. The increased CPU time for the risk-averse model when  $K = 1$  can be seen as the price to pay for the nonlinearity introduced by the risk measure  $\tau\text{AVaR}$ . Regarding the two instances with less or more volatility, it appears as if the low volatility data made the VI more difficult to solve: with 8 and 16 scenarios it took PATH twice the time for RN<sub>lo</sub> compared to RN<sub>hi</sub>. We observed this phenomenon in all of our runs. We conjecture that when uncertainty is more “alike” (as in Figure 4 for the right graph, with low volatility), there are more “similar” feasible points, which can be interpreted as kind of degeneracy that might give Newtonian solvers like PATH some problems. We observe that for  $K \leq 16$  all of the runs were successful (no failures) while for  $K = 32, 64$  no risk averse instance could be solved. For this reason, the runs below consider  $K = 16$ .

To compare the quality of the output, we compute the profit of the producers, averaged over 5 runs. Figure 5 shows the profit for each scenario, with and without risk aversion, and for the two instances of volatility for a problem with  $K = 16$ . Since the risk-neutral profit was (practically) identical in both volatility conditions, Figure 5 displays only one RN curve. The risk-averse solutions are very sensitive to different scenarios, especially if volatility is high. This is shown by the RA<sub>hi/lo</sub> values for scenarios 1 and 11 which represent, respectively, a very favourable and unfavourable situations. This behaviour is more extreme for the instance with high volatility: the producers accept to incur into losses for scenarios 6, 7, and 11, in view of the very high gain in scenario 1. Table 3 reports the mean and 90%-quantile for each



**Fig. 5** Producers Profit per Scenario, risk neutral and risk averse profit under low and high volatility

situation, exhibiting a pattern similar to the one observed in Figure 5 (note in particular the spread between the mean and the quantile).

**Table 3** Producers' Profit Statistics under varying volatility ( $K = 16$ )

| $RN_{lo}$ mean | quantile <sub>0.9</sub> | $RA_{lo}$ mean | quantile <sub>0.9</sub> | $RN_{hi}$ mean | quantile <sub>0.9</sub> | $RA_{hi}$ mean | quantile <sub>0.9</sub> |
|----------------|-------------------------|----------------|-------------------------|----------------|-------------------------|----------------|-------------------------|
| 732.4          | 756.8                   | 727.8          | 1989.3                  | 285.9          | 406.8                   | 1402.3         | 3347.8                  |

It took longer for PATH to solve problems with higher risk-aversion parameters. We conjecture this is because risk-averse producers make it more difficult for the market to reach an equilibrium. The profit statistics, as producers become more risk averse, are reported in Table 4. By comparing the values with  $\kappa \leq 0.75$  in Table 4, it appears that increasing the

**Table 4** Producers' Profit Statistics for increasing risk aversion

| $\kappa$ | $RA_{lo}$ mean | quantile <sub>0.9</sub> | $RA_{hi}$ mean | quantile <sub>0.9</sub> |
|----------|----------------|-------------------------|----------------|-------------------------|
| 0.25     | 760.8          | 2804.4                  | 911.0          | 6673.9                  |
| 0.50     | 758.2          | 2944.8                  | 1291.5         | 8548.0                  |
| 0.75     | 740.1          | 3827.4                  | 1315.6         | 12016.8                 |
| 1.00     | 748.3          | 3468.2                  | 1810.7         | 10712.6                 |

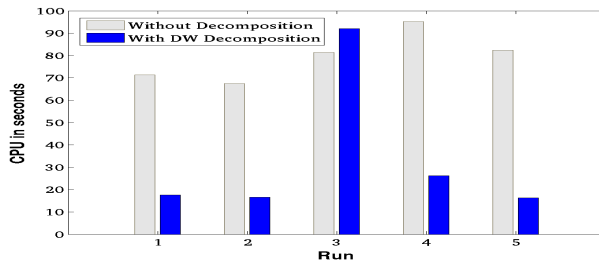
aversion to risk is beneficial for the producer (even more so in the high volatility setting). However, becoming completely risk averse may not pay off: the values with  $\kappa \leq 0.75$  are preferable for the producer compared to those obtained setting  $\kappa = 1$ . A closer inspection of the output shows that when  $\kappa = 1$ , the high volatility problem results in an equilibrium such that all of the profit quantiles give losses to the producers, until the 8th one. But since scenario number 1 is highly favourable, the corresponding gain pulls up the 9th quantile.

### 5.3 The Benefits of Decomposing

When using Matlab, the (compiled) mex-file of PATH limits the size of solvable VIs. For larger instances it would inevitably become necessary to apply some type of decomposition. For instance, the VI Dantzig-Wolfe approach [22] and [31]; the Benders' decomposition [14]; or the distributed method in [27]. In particular, [31] presents a rather broad and flexible framework, allowing for various kinds of data approximations, inexact solution of subproblems, and a potential for parallelization; all useful for the model at hand.



In the current case, the problem has no (investment) here-and-now variables, and the variables  $z^i$  disappear from  $C^i$  in Proposition 3(i)(b). The feasible set is separable by scenarios ( $k = 1, \dots, K$ ), with a structure amenable to Dantzig-Wolfe decomposition using the techniques in [31]. To induce separability along scenarios, we use the VI operator approximation called in [31] *constant approximation*. Specifically, at each iteration for each fixed scenario, we replace the last term in the VI operator by the vector with all the terms fixed to the last available value. With such an approximation and for our data, the subproblems in the Dantzig-Wolfe decomposition scheme of [31] become quadratic programming problems; they are solved by Mosek solver [www.mosek.org](http://www.mosek.org). The master problems of [31] have a simplicial feasible set and a differentiable VI operator that involves the derivatives of the smoothed risk measures; the master problems are solved by PATH. For more details on this class of decomposition methods and its convergence properties, see [31]. The interest of the approach as  $K$  increases is investigated in [31]; for an alternative decomposition having also good scalability properties, we refer to the distributed gradient projection scheme in [27].



**Fig. 6** The advantage of decomposing

Figure 6 shows the CPU times in seconds required by the direct solution of the problem by PATH and by using the Dantzig-Wolfe decomposition (stopped when its “gap measure” becomes smaller than  $0.05\sqrt{\dim}$ , with  $\dim$  the dimension of the VI variables). We considered an instance of high volatility with  $K = 4$  scenarios, and run the solvers 5 times. We observe that in most cases the decomposition approach found an equilibrium (the same as PATH) in about 1/3 of the time required by applying PATH directly to the VI. The exception is the third run. After a closer examination, our understanding is that this instance was a difficult one in terms of feasibility, due to a high stochastic demand. In such a case, solving a sequence of VIs sort of repeats the difficulty on every iteration. This third run is also an example of the importance of running several realizations (instances) of the same problem, before attempting any conclusions in a stochastic setting.

Since various specific schemes within the Dantzig-Wolfe class of [31] can be parallelized (in particular, the one used in our implementation here), additional speedup is to be gained with a professional implementation.

**Acknowledgements** The authors acknowledge insightful and significant comments and remarks from the reviewers and the editors, which led to a substantially improved version of this work.

The authors also thank Alexey Izmailov on the issue of computing the second derivatives described in Proposition 2.

## References

1. Baldick, R., Helman, U., Hobbs, B., O'Neill, R.: Design of efficient generation markets. *Proceedings of the IEEE* **93**(11), 1998–2012 (2005). DOI 10.1109/JPROC.2005.857484
2. Burke, J., Hoheisel, T.: Epi-convergent smoothing with applications to convex composite functions. *SIAM J. optimization* **23**, 1457–1479 (2013)
3. Chen, C., Mangasarian, O.: A class of smoothing functions for nonlinear and mixed complementarity problems. *Computational Optimization and Applications* **5**, 97–138 (1996)
4. Chen, X.: Smoothing methods for nonsmooth, nonconvex minimization. *Mathematical Programming* **134**, 71–99 (2012)
5. Chen, X., Fukushima, M.: Expected residual minimization method for stochastic linear complementarity problems. *Math. Oper. Res.* **30**(4), 1022–1038 (2005). DOI 10.1287/moor.1050.0160. URL <http://dx.doi.org/10.1287/moor.1050.0160>
6. Chen, X., Wets, R.J.B., Zhang, Y.: Stochastic variational inequalities: Residual minimization smoothing sample average approximations. *SIAM Journal on Optimization* **22**(2), 649–673 (2012)
7. Chen, X., Zhang, C., Fukushima, M.: Robust solution of monotone stochastic linear complementarity problems. *Math. Program.* **117**(1-2), 51–80 (2009). DOI 10.1007/s10107-007-0163-z. URL <http://dx.doi.org/10.1007/s10107-007-0163-z>
8. Collado, R.A., Powell, W.B.: Threshold risk measures part 1: Finite horizon (2013)
9. Conejo, A., Nogales, F., Arroyo, J., Garcia-Bertrand, R.: Risk-constrained self-scheduling of a thermal power producer. *Power Systems, IEEE Transactions on* **19**(3), 1569–1574 (2004). DOI 10.1109/TPWRS.2004.831652
10. Conejo, A.J., Carrión, M., Morales, J.M.: *Decision Making Under Uncertainty in Electricity Markets*. International Series in Operations Research & Management Science. Springer, New York (2010)
11. David, A., Wen, F.: Strategic bidding in competitive electricity markets: a literature survey. In: *Power Engineering Society Summer Meeting, 2000. IEEE*, vol. 4, pp. 2168–2173 vol. 4 (2000). DOI 10.1109/PSS.2000.866982
12. Dentcheva, D., Ruszczyński, A., Shapiro, A.: *Lectures on Stochastic Programming*. SIAM, Philadelphia (2009)
13. Dirkse, S., Ferris, M.: The PATH solver : A nonmonotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software* **5**, 123–156 (1995)
14. Egging, R.: Benders decomposition for multi-stage stochastic mixed complementarity problems – applied to a global natural gas market model. *European Journal of Operational Research* **226**(2), 341–353 (2013). DOI <http://dx.doi.org/10.1016/j.ejor.2012.11.024>
15. Egging, R.G., Gabriel, S.A.: Examining market power in the european natural gas market. *Energy Policy* **34**(17), 2762–2778 (2006). DOI <http://dx.doi.org/10.1016/j.enpol.2005.04.018>
16. Ehrenmann, A., Neuhoff, K.: A comparison of electricity market designs in networks. *Operations Research* **57**(2), 274–286 (2009)
17. Ehrenmann, A., Smeers, Y.: Generation capacity expansion in risky environment: A stochastic equilibrium analysis. *Operations Research* **59**(6), 1332–1346 (2011)
18. Facchinei, F., Fischer, A., Piccialli, V.: On generalized Nash games and variational inequalities. *Operations Research Letters* **35**, 159–164 (2007)
19. Facchinei, F., Kanzow, C.: Generalized Nash Equilibrium Problems. *Annals OR* **175**(1), 177–211 (2010)
20. Facchinei, F., Pang, J.S.: *Finite-dimensional variational inequalities and complementarity problems*. Vol. II. Springer Series in Operations Research. Springer-Verlag, New York (2003)
21. Ferris, M.C., Munson, T.S.: Interfaces to PATH 3.0: Design, Implementation and Usage. *Computational Optimization and Applications* **12**, 207–227 (1999)
22. Fuller, J.D., Chung, W.: Dantzig-Wolfe Decomposition of Variational Inequalities. *Comput. Econ.* **25**, 303–326 (2005). DOI 10.1007/s10614-005-2519-x. URL <http://dl.acm.org/citation.cfm?id=1075373.1075374>
23. Gabriel, S.A., Zhuang, J., Egging, R.: Solving stochastic complementarity problems in energy market modeling using scenario reduction. *European Journal of Operational Research* **197**(3), 1028–1040 (2009). DOI 10.1016/j.ejor.2007.12.046. URL <http://www.sciencedirect.com/science/article/pii/S0377221708002749>
24. Hiriart-Urruty, J.B., Lemaréchal, C.: *Convex Analysis and Minimization Algorithms*. No. 305-306 in *Grund. der math. Wiss.* Springer-Verlag (1993). (two volumes)
25. Hobbs, B., Metzler, C., Pang, J.S.: Strategic gaming analysis for electric power systems: an MPEC approach. *Power Systems, IEEE Transactions on* **15**(2), 638–645 (2000). DOI 10.1109/59.867153
26. Hu, X., Ralph, D.: Using EPECs to model bilevel games in restructured electricity markets with locational prices. *Operations Research* **55**(5), 809–827 (2007)

27. Kannan, A., Shanbhag, U.V., Kim, H.M.: Addressing supply-side risk in uncertain power markets: stochastic nash models, scalable algorithms and error analysis. *Optimization Methods and Software* **28**(5), 1095–1138 (2013). DOI 10.1080/10556788.2012.676756. URL <http://dx.doi.org/10.1080/10556788.2012.676756>
28. Kulkarni, A., Shanbhag, U.: On the variational equilibrium as a refinement of the generalized Nash equilibrium. *Automatica* **48**, 45–55 (2012)
29. Kulkarni, A.A., Shanbhag, U.V.: Revisiting generalized Nash games and variational inequalities. *Journal of Optimization Theory and Applications* **154**, 175–186 (2012)
30. Lin, G.H., Chen, X., Fukushima, M.: Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization. *Math. Program.* **116**(1-2), 343–368 (2009)
31. Luna, J.P., Sagastizábal, C., Solodov, M.: A Class of Dantzig-Wolfe Type Decomposition Methods for Variational Inequality Problems. *Mathematical Programming* **143**, 177–209 (2014). DOI 10.1007/s10107-012-0599-7. URL <http://dx.doi.org/10.1007/s10107-012-0599-7>
32. Luna, J.P., Sagastizábal, C., Solodov, M.: Complementarity and game-theoretical models for equilibria in energy markets: deterministic and risk-averse formulations. In: R. Kovacevic, G. Pflug, M. Vespucci (eds.) *International Series in Operations Research and Management Science*. Springer (2014). Vol. 199, Chapter 10, pp. 237–264
33. Meng, F., Sun, J., Goh, M.: A smoothing sample average approximation method for stochastic optimization problems with cvar risk measure. *Computational Optimization and Applications* **50**(2), 379–401 (2011)
34. Pang, J.S.: Three modeling paradigms in mathematical programming. *Mathematical Programming* **125**(2), 297–323 (2010). DOI 10.1007/s10107-010-0395-1. URL <http://dx.doi.org/10.1007/s10107-010-0395-1>
35. Pang, J.S., Fukushima, M.: Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. *Computational Management Science* **2**(1), 21–56 (2005). DOI 10.1007/s10287-004-0010-0. URL <http://dx.doi.org/10.1007/s10287-004-0010-0>
36. Pang, J.S., Fukushima, M.: Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. *Computational Management Science* **6**(3), 373–375 (2009). DOI 10.1007/s10287-009-0093-8. URL <http://dx.doi.org/10.1007/s10287-009-0093-8>
37. Patriksson, M., Wynter, L.: Stochastic mathematical programs with equilibrium constraints. *Operations Research Letters* **25**(4), 159–167 (1999). DOI 10.1016/S0167-6377(99)00052-8. URL <http://www.sciencedirect.com/science/article/pii/S0167637799000528>
38. Ralph, D., Smeers, Y.: EPECs as models for electricity markets. In: *Power Systems Conference and Exposition, 2006. PSCE '06. 2006 IEEE PES*, pp. 74–80 (2006). DOI 10.1109/PSCE.2006.296252
39. Ralph, D., Smeers, Y.: Pricing risk under risk measures: an introduction to stochastic-endogenous equilibria (2011). URL [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1903897](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1903897). Social Science Research Networks
40. Ralph, D., Xu, H.: Implicit smoothing and its application to optimization with piecewise smooth equality constraints I. *Journal of optimization theory and applications* **124**(3), 673–699 (2005)
41. Ravat, U., Shanbhag, U.V.: On the characterization of solution sets of smooth and nonsmooth convex stochastic Nash games. *SIAM Journal on Optimization* **21**(3), 1168–1199 (2011)
42. Rockafellar, R., Uryasev, S.: Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance* **26**(7), 1443–1471 (2002)
43. Rockafellar, R., Wets, J.B.: *Variational Analysis*. Springer-Verlag, New York (1997)
44. Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications* **128**(1), 223–243 (2006). DOI 10.1007/s10957-005-7566-x
45. Wets, R.J.B.: Stochastic programming models: Wait-and-see versus here-and-now. In: C. Greengard, A. Ruszczyński (eds.) *Decision Making Under Uncertainty, The IMA Volumes in Mathematics and its Applications*, vol. 128, pp. 1–15. Springer New York (2002). DOI 10.1007/978-1-4684-9256-9\_1. URL [http://dx.doi.org/10.1007/978-1-4684-9256-9\\_1](http://dx.doi.org/10.1007/978-1-4684-9256-9_1)
46. Xu, H., Zhang, D.: Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications. *Mathematical Programming* **119**(2), 371–401 (2009)
47. Xu, H., Zhang, D.: Stochastic Nash equilibrium problems: sample average approximation and applications. *Computational Optimization and Applications* **55**(3), 597–645 (2013)
48. Yin, H., Shanbhag, U., Mehta, P.: Nash equilibrium problems with scaled congestion costs and shared constraints. *Automatic Control, IEEE Transactions on* **56**(7), 1702–1708 (2011). DOI 10.1109/TAC.2011.2137590
49. Zhao, J., Hobbs, B.F., Pang, J.S.: Long-run equilibrium modeling of emissions allowance allocation systems in electric power markets. *Operations Research* **58**(3), 529–548 (2010)