

# STOCHASTIC SWITCHING GAMES AND DUOPOLISTIC COMPETITION IN EMISSIONS MARKETS

MICHAEL LUDKOVSKI

ABSTRACT. We study optimal behavior of energy producers under a  $CO_2$  emission abatement program. We focus on a two-player discrete-time model where each producer is sequentially optimizing her emission and production schedules. The game-theoretic aspect is captured through a reduced-form price-impact model for the  $CO_2$  allowance price. Such duopolistic competition results in a new type of a non-zero-sum stochastic switching game on finite horizon. Existence of game Nash equilibria is established through generalization to randomized switching strategies. No uniqueness is possible and we therefore consider a variety of correlated equilibrium mechanisms. We prove existence of correlated equilibrium points in switching games and give a recursive description of equilibrium game values. A simulation-based algorithm to solve for the game values is constructed and a numerical example is presented.

## 1. INTRODUCTION

In this paper we study a new class of non-zero-sum stochastic *switching games* with continuous state-space. Such games have natural applications in economics and finance, in particular for describing oligopolistic competition between large commodity producers. Our analysis is motivated by the  $CO_2$  cap-and-trade markets and provides new quantitative insight into the game-theoretic aspects of these schemes.

Switching game are a special class of dynamic non-zero-sum state-space games. They are characterized by a finite number of system states  $\vec{u}$ , jointly selected by the players. The players dynamically react to actions of other players and the evolution of state variables, represented as controlled stochastic processes, by strategically modifying the system state. Our contribution is a first rigorous probabilistic analysis of switching games. Because multiple game Nash equilibria are possible in our model, we propose to apply the wider concept of correlated equilibria. Correlated equilibria give a clear financial mechanism for stepwise equilibrium selection. Our key result is the construction of correlated equilibria in switching games in Section 3.3. The resulting representation in Theorem 3.4 of switching games in terms of a recursive sequence of stopping games leads to a constructive characterization of equilibrium strategies. Namely, we prove the analogue of the dynamic programming equation for the game values which enables numerical solution through backward recursion. Thus, the complexity of switching games is only slightly higher than of regular optimal switching problems.

---

2000 *Mathematics Subject Classification.* 91A15, 60G40, 93E20, 91B76.

*Key words and phrases.* stopping games, optimal switching, correlated equilibrium, carbon trading.

In terms of existing literature this paper extends two separate strands of research. Work on stochastic zero-sum *stopping games* dates back to Dynkin [14]. Such Dynkin games were progressively generalized in [1, 11, 16, 19, 40]. Later extensions also treated special cases of non-zero-sum stopping games, especially the so-called monotone type [23, 32, 33]. The key tool of correlated equilibria in stochastic dynamic games was studied by [31, 35, 36, 37, 38, 39]. We augment these results by explicitly characterizing correlated equilibria in repeated stopping games using the methods of Ramsey and Szajowski [35]. Contemporaneously, the theory of *optimal switching* for a single agent was developed and extensively studied in the past decade, see [8, 12, 20, 34]. In Section 3.3 we extend these results to a game setting by showing that at switching game equilibrium each player faces a optimal switching problem with *randomized* controls.

Another contribution of this work is a construction of numerical schemes to compute game-values and equilibrium strategies of switching games. This is achieved by combining backward recursion together with sequential solution of local  $2 \times 2$  games. We suggest two approaches, one based on the Markov chain approximation method and a second algorithm that relies on least squares Monte Carlo. The latter is a novel extension of our previous work in [8, 28] and borrows ideas from standard optimal stopping theory to implement the analogue of the dynamic programming recursion on a set of Monte Carlo simulations.

A significant portion of the paper is dedicated to the application of our model to emissions trading. With imminent ramping up of  $CO_2$ -emissions markets around the world (see e.g. the Western Climate Initiative in the US and the EU ETS Phase III in Europe both set to start in 2012), it is crucial to understand energy producer behavior under the new frameworks. By design, the carbon allowances will be scarce and market participants will be competing for finite permit resources. Our analysis is a first pass at oligopolistic competition in  $CO_2$  markets using game-theoretic methods. We hope it can serve as a stepping-stone to more sophisticated modeling that addresses market design and comparative statics of our framework.

The rest of the paper is organized as follows. In Section 2 we define the precise stochastic model representing oligopolistic competition in  $CO_2$  markets. Section 3 constructs the representation of switching games in terms of repeated stopping games and culminates with Theorem 3.4 that establishes the dynamic programming equations. Section 4 describes our numerical solution algorithms and presents a computational example. Finally, Section 5 discusses extensions of the model and points directions for future work.

## 2. COMPETITIVE EQUILIBRIUM AMONG OLIGOPOLISTIC PRODUCERS

In this section we formally define a model for the competitive dynamic equilibrium between producers with market power.

**2.1. Price Dynamics under Cap-and-Trade Emission Schemes.** The carbon allowance markets are intrinsically linked to other energy markets, notably electricity whose production accounts for the bulk of regulated emissions. Furthermore, due to their size, major electricity producers often

have the ability to dramatically move carbon prices based on their emission schedules. Hence, to understand equilibria in  $CO_2$ -markets it is useful to consider them from the point of view of *large traders*. The model proposed below captures this phenomenon for a joint electricity-carbon market.

To fix ideas, start with a filtered probability space  $(\Omega, \mathcal{H}, (\mathcal{F}_t), \mathbb{P})$ ,  $t \in \mathbb{T} \triangleq \{0, 1, \dots, T\}$ . The terminal time  $T$  is the expiration date of the current vintage of permits. We consider two heterogeneous producers (henceforth termed players) who each produce commodity  $P$  (electricity) and consume commodity  $X$  (carbon allowances). These two producers generate “dirty” electricity from e.g. coal or gas and can be viewed as representative agents of a park of power plants with identical engineering characteristics. All the other participants in the electricity and  $CO_2$  markets are not modeled explicitly; rather we postulate that their collective actions induce *stochastic* fluctuations in the prices of  $P$  and  $X$ . We assume that the two players are large traders in the carbon market, but small players on the electricity market. This reflects the fact that the other “green” producers (who use nuclear, hydroelectric, renewable, etc. sources) create a competitive electricity market while remaining passive in the  $CO_2$ -permits arena.

The electricity price is given by the exogenous discrete-time stochastic process  $(P_t)$ , for simplicity taken to be one-dimensional,

$$P_{t+1} = G(P_t, \epsilon_t^P),$$

where the innovations  $(\epsilon_t^P)$  are i.i.d. standard Gaussian. Our canonical example is the logarithmic Ornstein-Uhlenbeck process which is a log-normal stationary Gaussian process with

$$(1) \quad P_{t+1} = P_t \cdot \exp(\kappa_P(\bar{P} - \log P_t) + \sigma_P \epsilon_t^P), \quad \epsilon_t^P \sim \mathcal{N}(0, 1),$$

for some positive constants  $\kappa_P, \bar{P}, \sigma_P$ .

The objective of the producers is to maximize their expected net profits over the planning horizon  $T$ . The producers’ profit is given by their *clean dark spread* [17] which is defined as the difference between electricity price and the carbon-adjusted production cost. We assume that input fuel costs are fixed, as is often the case for power generators with long-term supply contracts. The strategy of each player is described by a repeated start-up/shut-down option. Namely, if the market conditions are unfavorable, a player can stop production, eliminate  $CO_2$  emissions and avoid losses; she can then restart production when the profit spread improves. As a first approximation we assume that these choices of production regimes are binary and denoted as “off” (0) and “on” (1). Formally, the production schedule of each producer is described by a stochastic process  $u_i$ ,  $u_i(t) \in \{0, 1\}$ ,  $t \in \mathbb{T}$ . In a single-player model, such timing optionality is known as a real option and has been thoroughly investigated since the seminal work of [4, 13]. Repeated real options have attracted considerable attention recently, see [8, 20, 28], and others.

*Remark 2.1.* An alternative formulation is to take  $u_i$  to be continuous, so that the producers can choose emissions levels smoothly. This would lead to a non-zero-sum stochastic dynamic game. Such models have been extensively studied both in discrete and continuous time, see [22, 30] and references therein. While presenting other formidable technical challenges, the problem of equilibrium selection is less severe with continuous controls thanks to the convexity of value

functions. In this work we focus on the timing flexibility and therefore maintain the discrete control space.

Let  $X_t$  be the permit price at date  $t$ . Based on above discussion, the actions of each player influence the dynamics of  $X_t$ . Namely, conditional on player actions  $u_1(t), u_2(t)$ , we model  $X_t$  as another mean-reverting process with a policy-dependent mean and log-Gaussian increments,

$$(2) \quad \begin{cases} X_{t+1} = X_t \cdot \exp(\kappa_X(f(u_1(t), u_2(t)) - \log X_t) + \sigma_X \epsilon_t^X) & \text{with} \\ f(u_1, u_2) = \log(\bar{X} + g_1 u_1 + g_2 u_2). \end{cases}$$

The sequence  $(\epsilon_t^X)$  is again Gaussian, with correlation parameter  $\rho$  to  $(\epsilon_t^P)$ , i.e.  $\epsilon_t^X = \rho \epsilon_t^P + \sqrt{1 - \rho^2} \epsilon_t^\perp$  with  $\epsilon_t^\perp \sim \mathcal{N}(0, 1)$  independent of  $\epsilon_t^P$ . Rising electricity prices are likely to increase the overall  $CO_2$  emission rates and therefore we expect that  $P$  and  $X$  are positively correlated,  $\rho > 0$ .

*Remark 2.2.* To motivate (2), we recall from [18] that in a carbon market “ $X_t = \bar{x} \bar{\mathbb{P}}\{C_T > \bar{c} | \mathcal{F}_t\}$ ,” where  $C_T$  is the cumulative total  $CO_2$  emissions on  $[0, T]$ ,  $\bar{x}$  is the penalty for going over the allowance limit,  $\bar{c}$  is the total amount of allowances allocated and  $\bar{\mathbb{P}}$  is the equilibrium pricing measure. We postulate that  $C_T = \sum_{s=0}^{T-1} \{b_1 u_1(s) + b_2 u_2(s) + \tilde{u}(s)\}$ , where  $b_i u_i(s)$  are the emissions of producer  $i$  in period  $s$  and  $\tilde{u}(s)$  are the emissions by all other market participants. Assuming independent increments (due to external shocks such as weather effects, etc.) in  $\tilde{u}(s)$ , the dynamics (2) follow, with some complicated and time-dependent functions  $f(t, \cdot)$  and volatility  $\sigma_X(t)$ . In (2) we give a simplified or reduced-form version of this description to capture the temporal feedback between  $u_i$ 's and  $X_t$ . If the supply curve for the  $CO_2$  allowances is convex, then the price impact  $f(u_1, u_2)$  in (2) would be nonlinear and further magnify the competitive effects.

**2.2. Optimization Objective.** We assume that the producers have zero allowance allocations and cannot bank allowances; therefore they must purchase the requisite allowances at each stage of the game. The total P&L of the players then consists of (i) revenue from selling electricity, minus the (ii) cost of buying emission allowances, as well as (iii) operational costs due to adopted strategy  $u_i$ . An important case of operational costs are *fixed switching costs*  $K_{\{i, j_1, j_2\}}$  that are paid each time the production regime of agent  $i$  is changed from  $j_1$  to  $j_2$  and corresponding to the ramping-up/winding-down costs associated with the electricity turbines [8, 17]. We postulate  $K_{\{i, j_1, j_1\}} = 0 \forall j_1$  and the triangle inequality  $K_{\{i, j, \ell\}} \leq K_{\{i, j, k\}} + K_{\{i, k, \ell\}}$  for all  $j, k, \ell$ .

Let  $\mathbb{P}^{\vec{u}}$  be the law of  $(P_t, X_t)$  given a strategy pair  $\vec{u} \equiv (u_1(t), u_2(t))_{t=0}^T$ . The expected cumulative net profit of producer  $i$  starting with  $P_s = p$ ,  $X_s = x$  and initial production regime  $\vec{\zeta} \in \{0, 1\}^2$  is

$$(3) \quad V_i(s, p, x, \vec{\zeta}; \vec{u}) \triangleq \mathbb{E}^{\vec{u}} \left[ \sum_{t=s}^{T-1} \{(a_i P_t - b_i X_t - c_i) u_i(t) - K_{\{i, u_i(t-), u_i(t)\}}\} \middle| P_s = p, X_s = x, \vec{u}(s-) = \vec{\zeta} \right].$$

The constants  $a_i, b_i, c_i, K_{\{i, j_1, j_2\}}$ ,  $i = 1, 2$ , represent the maximum quantity of electricity produced by the facility in one period, the amount of corresponding  $CO_2$  allowances needed, fixed production costs and switching costs, respectively. Due to switching costs, current production regime is also a

state variable. Below, the theorems on existence of equilibria in stochastic games require bounded payoffs; therefore we assume that profits are truncated from above at some large positive level.

In the duopoly setting, while each player aims to maximize her own profits, the competitor actions will also affect her decisions. Indeed, emissions today shrink remaining permit supplies and tend to increase *future*  $CO_2$  prices. Therefore, if player 1 is emitting, player's 2 expected future profits are reduced. Overall, the producers are facing a stochastic game where actions correspond to the latest choice of production regime by each player and payoffs are a function of the exogenous  $P_t$  and the partly controlled  $X_t$ . Since the game is stochastic and multi-period with Markov state variables, we restrict our attention to Markovian (feedback) equilibria. Our main task for the remainder of the paper is to characterize such game equilibria and then compute the corresponding game value functions (i.e. expected profits)  $V_i$  and equilibrium emission schedules  $(u_1^*, u_2^*)$ .

**2.3. Randomized Emission Schedules.** The strategies  $u_i$  may be mixed or randomized, i.e.  $u_i(t)$  is not necessarily adapted to the market filtration  $\mathcal{F}$ . However, we also assume a full-information setting, whereby the emission schedules of each agent are publicly known after the fact. Accordingly, the market observables are  $\mathcal{F}_t = \sigma(X_0, P_0, u_1(0), u_2(0), \dots, u_1(t-1), u_2(t-1), X_t, P_t)$ , the filtration generated by the price histories and past actions. An  $\mathcal{F}$ -randomized emission strategy is a pair  $(u_i(t), \mathcal{G}^i(t))$  where  $\mathcal{G}^i$  is an independent enlargement of the filtration  $\mathcal{F}$  (i.e.  $\mathbb{P}(A|\mathcal{F}_t) = \mathbb{P}(A|\tilde{\mathcal{G}}_t^i)$  for all  $A \in \mathcal{F}_t$ ) and  $u_i$  is  $\mathcal{G}^i$ -adapted. Let

$$(4) \quad p_i(t) \triangleq \mathbb{P}(u_i(t) = 1 | \mathcal{F}_t)$$

denote the stage- $t$  probability that the control will be 'on', given observable information so far. If  $p_i(t) \in \{0, 1\}$  then the strategy is *pure* at stage  $t$ , otherwise it is *mixed* and can be represented via a randomization parameter  $\eta_i(t)$  as

$$(5) \quad u_i(t) = 1_{\{\eta_i(t) \leq p_i(t)\}}, \quad \eta_i(t) \sim Unif(0, 1), \eta_i(t) \perp \mathcal{F}_t.$$

The full mixed strategy is the vector  $\vec{\pi}_i(t) \equiv (1 - p_i(t), p_i(t))$  belonging to the 2-simplex  $\vec{\pi}_i(t) \in \Delta_2 \triangleq \{(\pi^0, \pi^1) : \pi^j \geq 0, \pi^0 + \pi^1 = 1\}$ . The joint action is given by the *strategy profile*  $\vec{\pi}(t)$ , with  $\pi_i^j(t)$  denoting the probability that player  $i$  emits at level  $j$ .

The set  $\mathcal{U}_i$  of admissible production schedules for player  $i$  consists of  $\mathcal{F}$ -randomized  $\{0, 1\}$ -valued processes and can be canonically identified with an  $\mathcal{F}$ -adapted process  $(p_i(t))$ ,  $0 \leq p_i(t) \leq 1$  and independent sequence  $\eta_i(t)$  as in (5). Let  $\mathfrak{D}^i(t)$  denote the set of  $\mathcal{G}^i$ -stopping times bigger than  $t$ . Because  $u_i(t) \in \{0, 1\}$ , we have a one-to-one correspondence between admissible  $u_i$ 's and sequences  $(\tau_k^u)_{k=1}^\infty$  satisfying  $\tau_{k+1}^u \in \mathfrak{D}^i(\tau_k^u)$ ,

$$(6) \quad u_i(t) = \sum_{k=0}^T u_i(0) 1_{[\tau_{2k}^u, \tau_{2k+1}^u)} + (1 - u_i(0)) 1_{[\tau_{2k+1}^u, \tau_{2k+2}^u)}, \quad \tau_0^u = 0.$$

The switching times  $\tau_k^u$  encode the times of production regime shifts defined by  $u_i$ . The representation (6) holds because at most one regime switch can be made by each player at any given stage. Indeed, multiple simultaneous regime switches by the same producer are strongly sub-optimal if  $K_{\{i, j_1, j_2\}} > 0$  and weakly suboptimal otherwise. A  $\mathcal{G}^i$ -adapted stopping time  $\tau$  can also be viewed as

a randomized  $\mathcal{F}$ -stopping time, via its conditional stopping probabilities  $p_t = \mathbb{P}(\tau = t | \tau > t-1, \mathcal{F}_t)$ , namely

$$(7) \quad \tau(p) \triangleq \inf\{t : \eta_t \leq p_t\}, \quad \eta_t \sim Unif(0, 1) \perp \mathcal{F}_t.$$

When  $p_t \in \{0, 1\}$  for all  $t$ , we are back in the case of regular  $\mathcal{F}$ -stopping times.

If  $\mathcal{G}_t^1 \cap \mathcal{G}_t^2 = \mathcal{F}_t$  then the randomizations of the two players are independent. Alternatively, correlated decision making can be introduced by making the  $\eta_i$ 's in (5) dependent. Let  $\gamma$  be an  $\mathcal{F}$ -adapted stochastic process taking values in  $\Delta_4$ . Following [35] we interpret  $\gamma(t)$  as a *weak (stepwise) communication device*, with  $\gamma_{ij}(t)$  specifying the probability that player 1 takes action  $i \in \{0, 1\}$  and player 2 applies action  $j \in \{0, 1\}$ ,

$$\gamma_{ij}(t) \triangleq \mathbb{P}(u_1(t) = i, u_2(t) = j | \mathcal{F}_t).$$

The correlation is implemented via a third party that directs the players to implement a particular action pair through private signals. Namely, the players receive signals

$$(8) \quad \mu_1(t, \gamma) = 1_{\{\gamma_{10}(t) + \gamma_{11}(t) < \eta(t)\}} \quad \text{and} \quad \mu_2(t, \gamma) = 1_{\{\gamma_{01}(t) + \gamma_{11}(t) < \eta(t)\}},$$

where  $\eta \sim Unif(0, 1)$  is only observed by the third party. Setting  $u_i(t) = \mu_i(t, \gamma)$ , the resulting strategy profile is denoted  $\vec{u}(t, \gamma) = (\vec{\pi}_1(t, \gamma), \vec{\pi}_2(t, \gamma))$  and has dependent marginals and joint distribution  $\gamma$ . Conditional on the signal at date  $t$ , an agent can impute the strategy of the other player by e.g.  $\vec{\pi}_2(t, \gamma) \big|_{\mu_1(t, \gamma)=0} = \left( \frac{\gamma_{00}(t)}{\gamma_{00}(t) + \gamma_{01}(t)}, \frac{\gamma_{01}(t)}{\gamma_{00}(t) + \gamma_{01}(t)} \right)$ . With  $\gamma$  in place, the space of randomized strategies is now adjusted to  $\mathcal{U}_i(\gamma) = \{(u_i, \mathcal{G}^i) \in \mathcal{U}_i \text{ such that } \mathcal{G}_t^i \supseteq \mathcal{F}_t \vee \sigma(\mu_i(t))\}$ .

**2.4. Correlated Equilibria in  $CO_2$  markets.** The introduced correlation mechanism  $\gamma$  can be used to define correlated equilibrium points (CEP) in the  $CO_2$  emissions duopoly game. To motivate the need for such mechanisms, we observe that intuitively the dynamic switching game is a sequence of one-period bimatrix games. At each stage  $t$ , the control  $u_i(t) \in \{0, 1\}$  of each player  $i \in \{1, 2\}$  is simply “on/off”, leading to the classic  $2 \times 2$  game. From a dynamic point of view, the relevant payoff to the players at stage  $t$  is then the sum of the current clean spread and the continuation value that corresponds to the game value that can be realized in the future by the respective player contingent on current state of the world. In our repeated game setting, the players must *a priori* agree on how to implement future equilibria, otherwise the computation of continuation values would not be possible. Hence, to have a well-defined switching game value, we need existence-uniqueness of equilibria in the one-period sub-games.

Accordingly, we briefly recall the structure of 2-by-2 one-shot game. Consider the  $2 \times 2$  game  $H$  with normal form

$$(9) \quad H = \begin{pmatrix} (z_1^{00}, z_2^{00}) & (z_1^{01}, z_2^{01}) \\ (z_1^{10}, z_2^{10}) & (z_1^{11}, z_2^{11}) \end{pmatrix},$$

where the rows of  $H$  are chosen by player 1, and the columns by player 2. A strategy profile  $(\bar{\pi}_1^*, \bar{\pi}_2^*)$  is a *Nash equilibrium point* (NEP) of  $H$  if we have

$$\sum_{j,k} \pi_1^{*,j} \pi_2^{*,k} z_1^{jk} = \sup_{\bar{\pi}_1 \in \Delta_2} \sum_{j,k} \pi_1^j \pi_2^{*,k} z_1^{jk}, \quad \text{and} \quad \sum_{j,k} \pi_1^{*,j} \pi_2^{*,k} z_2^{jk} = \sup_{\bar{\pi}_2 \in \Delta_2} \sum_{j,k} \pi_1^{*,j} \pi_2^k z_2^{jk}.$$

Hence,  $\bar{\pi}_i^*$  is a best-response for player  $i$ , given that the other player uses  $\bar{\pi}_{-i}^*$ . While the classical theorem of Nash shows that a mixed NEP is always available,  $H$  may have zero (strictly competitive case), one (standard case) or two (coordination case) pure NEP's [5].

Thus, to establish existence of NEP in a multi-period game, one must consider mixed strategies. Furthermore, since multiple equilibria are possible, an *equilibrium selection* rule is needed. In the context of the  $CO_2$  emission game, because the  $(P, X)$ -prices are stochastic, it is impossible to *a priori* rule out some of the above scenarios for all possible state variable realizations. In particular, the case of the anti-coordination “battle-of-the-sexes” or “chicken” game is likely to appear when the electricity-carbon spread is slightly positive. In such a situation, each of the players will have an incentive to emit; however, if the price impact is strong enough, it is not profitable for *both* of them to consume permits. As a result, two pure Nash equilibria are possible whereby one producer yields the market to the other.

The communication device  $\gamma$ , introduced in one-shot games by [2, 29], provides a general method for describing such coordination while maintaining the non-cooperative game setting.

**Definition 2.1.** *A Markovian correlated equilibrium point for the switching game is a Markov communication device  $\gamma : (s, p, x, \vec{\zeta}) \rightarrow \Delta_4$  inducing admissible stage strategy profiles  $\bar{u}^*(t; \gamma) = (\mu_1(t, \gamma), \mu_2(t, \gamma))$  such that  $\forall (s, p_0, x_0, \vec{\zeta})$  (recall definition of  $V_i$  in (3))*

$$(10) \quad \begin{cases} V_1(s, p_0, x_0, \vec{\zeta}; u_1^*, u_2^*) \geq V_1(s, p_0, x_0, \vec{\zeta}; u_1, u_2^*) & \forall u_1 \in \mathcal{U}_1, \\ V_2(s, p_0, x_0, \vec{\zeta}; u_1^*, u_2^*) \geq V_2(s, p_0, x_0, \vec{\zeta}; u_1^*, u_2) & \forall u_2 \in \mathcal{U}_2. \end{cases}$$

*The resulting game values are denoted as  $V_i(s, p_0, x_0, \vec{\zeta}; \gamma)$ .*

The meaning of the correlated equilibrium in (10) is that conditional on the private signal sequence, neither player has an incentive to deviate from the prescribed action. Therefore, given  $\mu_i(t, \gamma)$  and market information  $\mathcal{F}_t$ , it is optimal to take  $u_i(t) = \mu_i(t, \gamma)$ . Note that in (10), even if a player chooses to deviate from the recommendation  $\mu_i(t, \gamma)$  she continues to receive future signals  $\mu_i(s, \gamma)$ ,  $s > t$  and therefore information about the implied strategy of the other player. Existence of CEP of switching games will be established in Theorem 3.4. We will also provide a recursive construction of  $V_i(t, \cdot)$  in terms of conditional expectations of  $V_i(t+1, \cdot)$  and one-shot  $2 \times 2$  games. This allows for a solution method, detailed in Section 4, analogous to the dynamic programming paradigm for ordinary stochastic control problems. Finally, we will show that CEP of switching game induces rational behavior at each stage, i.e. matches with a CEP of a one-stage sub-game.

*Remark 2.3.* A related concept of competitive equilibrium in the industrial organization literature is that of a stage Stackelberg game [3]. In a Stackelberg game, at each stage one player is the leader

and has priority in making decisions; the second player then follows. This description corresponds closely to the preferential mechanism of equilibrium selection which always favors the leader.

Economically, the third party implementing the correlation could be a government regulator, market watchdog, or just a proxy for market frictions that make one equilibrium most preferable. Thus, no inherent collusion is required and the game is still non-cooperative. If a regulator is involved, a socially beneficial correlation can be selected. For instance, a “utilitarian” communication device maximizes the (weighted) sum of the firms continuation values so that the producers as a whole have best economic health. Alternatively, a “green” device minimizes  $CO_2$  emissions. Finally, a “preferential” communication mechanism can endogenously emerge without a third party due to extra advantages available to a given player (e.g. due to preferential regulatory treatment or other externalities).

*Remark 2.4.* A variety of correlated decision-making is possible in sequential games [29]. Here we focus on the stepwise weak communication device whereby the players and the regulator communicate before each stage; such a formulation allows the most flexibility and fits our economic description. However, in practice much weaker correlation could suffice. For instance, players can agree at date 0 to use the preferential- $i$  correlation law which means that in any “tie-break” case, player  $i$  “wins”. Once this rule is fixed, no further communication would be necessary. Similarly, if  $\gamma$  is such that the implied strategy  $\vec{\pi}_{-i}(t, \gamma) | \mu_i(t, \gamma)$  of the other player is always pure, then a public randomization is sufficient at each step and no private signals are needed. Any mixture of NEP’s is also a CEP and therefore except for the strictly-competitive games, one may always find correlation devices that correspond to pure Nash equilibria, obviating the need for randomization (either by players or regulator).

### 3. SEQUENTIAL STOPPING GAME

Our analysis of the switching game will consist of building up the solution in several steps. We begin with analyzing the single-agent objective. Next, in Section 3.2 we move on to the one-shot non-zero-sum stopping game that is built iteratively from the one-period  $2 \times 2$  games, following the methods of [35]. Finally, in Section 3.3 we describe the sequential stopping game that in the limit is shown in Section 3 to coincide with our original model in (3).

**3.1. Single Producer Problem.** Before tackling the stochastic duopoly game, let us briefly review the solution of the single-player model. Since the control  $u_i(t)$  takes on a finite number of values, we have an optimal switching model that can be viewed as a sequence of optimal stopping problems. Such models (including price impact) were studied in [8, 28].

Let us consider the optimization for producer 1. For the remainder of this section we fix a production schedule  $u_2$  of the second producer, as well as a communication device  $\gamma$  that sends private signals  $\mu_1(t, \gamma)$  to player 1. In the single-producer problem, the objective is to maximize



the expected profit

$$(11) \quad \sup_{(u(t)) \in \mathcal{U}_1(\gamma)} \mathbb{E}^{(u, u_2, \gamma)} \left[ \sum_{t=0}^{T-1} \left\{ (a_1 P_t - b_1 X_t - c_1) u(t) - K_{\{1, u(t-), u(t)\}} \right\} \right].$$

Consider initial conditions  $P_s = p, X_s = x, u_2(s) = \zeta_2$  and let  $V(s, p, x, \zeta_2)$  be the value function corresponding to (11) conditional on starting in the “on”-production regime, and  $W(s, p, x, \zeta_2)$  the value function starting offline. Furthermore, using same initial conditions define recursively

$$(12) \quad \left\{ \begin{array}{l} V^0(s, p, x, \zeta_2) = \mathbb{E}^{(1, u_2, \gamma)} \left[ \sum_{t=s}^{T-1} (a_1 P_t - b_1 X_t - c_1) \right], \quad \text{as well as} \quad W^0(s, p, x, \zeta_2) = 0; \\ V^n(s, p, x, \zeta_2) = \sup_{\tau \in \mathcal{D}^1(s)} \mathbb{E}^{(1, u_2, \gamma)} \left[ \sum_{t=s}^{\tau-1} (a_1 P_t - b_1 X_t - c_1) + (W^{n-1}(\tau, P_\tau, X_\tau, u_2(\tau)) - K_{\{1, 1, 0\}}) \right]; \\ W^n(s, p, x, \zeta_2) = \sup_{\tau \in \mathcal{D}^1(s)} \mathbb{E}^{(0, u_2, \gamma)} [V^{n-1}(\tau, P_\tau, X_\tau, u_2(\tau)) - K_{\{1, 0, 1\}}], \quad n \geq 1 \end{array} \right.$$

where under  $\mathbb{P}^{(i, u_2, \gamma)}$  the drift of the carbon allowance price is  $f(i, u_2(t))$ .

**Proposition 3.1.** *Let  $\mathcal{U}_1^n \triangleq \{u \in \mathcal{U}_1 : u \text{ has at most } n \text{ switches}\}$ . Then,*

$$V^n(s, p, x, \zeta_2) = \sup_{(u(t)) \in \mathcal{U}_1^n, u(s-)=1} \mathbb{E}^{(u, u_2, \gamma)} \left[ \sum_{t=s}^{T-1} \left\{ (a_1 P_t - b_1 X_t - c_1) u(t) - K_{\{1, u(t-), u(t)\}} \right\} \right],$$

and as  $n \rightarrow \infty$ ,  $V^n(s, p, x, \zeta_2) \rightarrow V(s, p, x, \zeta_2)$ ,  $W^n(s, p, x, \zeta_2) \rightarrow W(s, p, x, \zeta_2)$  uniformly on compacts.

*Proof.* This is an analogue of [8, Theorem 1]. Compared to our earlier work, the only new feature is that the payoffs to producer 1 are randomized. Indeed, from her perspective, the strategy of player 2 (implied through the private signal  $\mu_1(t, \gamma)$ ) may be mixed. Consequently, her continuation value is unknown at decision time, depending as it is on the action of player 2. Formally, allowing for a relaxed switching control  $p_s^1$  at date  $s$  (representing probability of being on) the dynamic programming principle implies that in (12)

$$\begin{aligned} V^n(s, p, x, \zeta_2) = & \mathbb{E}^{\gamma(s)} \left[ \sup_{p_s^1 \in [0, 1]} \left\{ p_s^1 (a_1 p - b_1 x - c_1) - (1 - p_s^1) K_{\{1, 1, 0\}} \right. \right. \\ & + \mathbb{E}^{\mu_1(s, \gamma)} \left[ p_s^1 p_s^2 V^n(s+1, P_{s+1}, X_{s+1}^{(0,1)}, 1) + p_s^1 (1 - p_s^2) V^n(s+1, P_{s+1}, X_{s+1}^{(0,0)}, 0) \right. \\ & \left. \left. + (1 - p_s^1) p_s^2 W^{n-1}(s+1, P_{s+1}, X_{s+1}^{(0,1)}, 1) + (1 - p_s^1) (1 - p_s^2) W^{n-1}(s+1, P_{s+1}, X_{s+1}^{(0,0)}, 0) \right\} \right]. \end{aligned}$$

The outer expectation is averaging over the signal  $\mu_1$  whose law is specified by the communication device  $\gamma$ ; however the decision-maker has access to  $\mu_1(t, \gamma)$  and therefore makes the switching decision  $p_s^1$  based on the conditional strategy  $(p_s^2) | \mu_1(s, \gamma)$  of player 2. The inner optimization is linear in  $p_s^1$  and therefore the optimizer must be an endpoint of  $[0, 1]$ . Thus, as expected, given the

signal we can work with pure controls,  $u_1(t) \in \mathcal{F}_t \vee \sigma(\mu_1(t, \gamma))$ . Note that from the perspective of an observer who has access only to  $\mathcal{F}$ , the strategy of both players appears randomized.

The rest of the proof proceeds exactly as in [8] by iterating over the control decisions of producer 1 using the strong Markov property of  $(P, X)$  and the Snell envelope characterization of optimal stopping problems.  $\square$

Proposition 3.1 shows that the solution to (11) can be represented in terms of the sequence  $(V^n, W^n)$  which correspond to optimal stopping problems defined in (12). Taking the limit  $n \rightarrow \infty$  we obtain

**Corollary 3.1.**  *$(V, W)$  satisfy the coupled dynamic programming equation:*

$$\begin{cases} V(s, p, x, \zeta_2) = \sup_{\tau \in \mathcal{D}^1(s)} \mathbb{E}^{(1, u_2, \gamma)} \left[ \sum_{t=s}^{\tau-1} (a_1 P_t - b_1 X_t - c_1) + (W(\tau, P_\tau, X_\tau, u_2(\tau)) - K_{\{1,1,0\}}) \right], \\ W(s, p, x, \zeta_2) = \sup_{\tau \in \mathcal{D}^1(s)} \mathbb{E}^{(0, u_2, \gamma)} \left[ V(\tau, P_\tau, X_\tau, u_2(\tau)) - K_{\{1,0,1\}} \right]. \end{cases}$$

Moreover, an optimal strategy  $u_1^* \in \mathcal{U}_1$  exists.

**3.2. Correlated Equilibria in Non-Zero-Sum Stopping Games.** In this section we recall existing results on two-player non-zero sum stopping games in discrete time and finite horizon. Let  $\mathcal{Z} \equiv (Z_i^{jk}(t))$ ,  $i \in \{1, 2\}$ ,  $j, k \in \{0, 1\}$  be a octuple of bounded  $(\mathcal{F}_t)$ -adapted stochastic processes. Player  $i \in \{1, 2\}$  optimizes the reward

$$(13) \quad \tilde{J}_i(s, \tau_1, \tau_2) \triangleq \left( \sum_{t=s}^{\tau_i \wedge \tau_{-i} - 1} Z_i^{00}(t) \right) + Z_i^{10}(\tau_i) 1_{\{\tau_i < \tau_{-i}\}} + Z_i^{01}(\tau_{-i}) 1_{\{\tau_{-i} < \tau_i\}} + Z_i^{11}(\tau_i) 1_{\{\tau_i = \tau_{-i}\}},$$

by choosing the (randomized)  $(\mathcal{F}_t)$ -stopping time  $\tau_i \leq T$ . In words,  $Z_i^{00}$  is the ongoing reward for staying in the game,  $Z_i^{10}$  is the reward if the player stops first;  $Z_i^{01}$  is the reward if the other player stops first and  $Z_i^{11}$  is the reward if both players stop simultaneously. Thus, continuing is associated with action ‘0’ and stopping with action ‘1’.

The Dynkin zero-sum stopping game corresponds to  $Z_1^{10} = -Z_2^{01}$ ,  $Z_1^{01} = -Z_2^{10}$ ,  $Z_1^{11} = -Z_2^{11}$  and was recently fully analyzed by [16]. Also, the monotone cases  $Z_i^{01} \leq Z_i^{11} \leq Z_i^{10}$   $\mathbb{P}$ -a.s. (where both players prefer to stop late) and  $Z_i^{01} \geq Z_i^{11} \geq Z_i^{10}$  were considered by Ohtsubo [32]. In these special cases, a unique pure Markov NEP exists. The fundamental result of [32, 33] characterizes game value functions  $(V_1, V_2)$  for  $\mathcal{Z}$  as a pair of  $\mathcal{F}$ -adapted processes satisfying  $\mathbb{E}[\sup_{0 \leq t \leq T} V_i(t)] < \infty$ ,  $V_i(T) = Z_i^{11}(T)$  and for all  $0 \leq t \leq T$

$$(14) \quad (V_1(t), V_2(t)) \in \mathcal{E} \begin{pmatrix} (\mathbb{E}[V_1(t+1)|\mathcal{F}_t] + Z_1^{00}(t), \mathbb{E}[V_2(t+1)|\mathcal{F}_t] + Z_2^{00}(t)) & (Z_1^{01}(t), Z_2^{01}(t)) \\ (Z_1^{10}(t), Z_2^{10}(t)) & (Z_1^{11}(t), Z_2^{11}(t)) \end{pmatrix},$$

where  $\mathcal{E}(H)$  is the set of game values corresponding to NEPs of  $H$ . This reduces computation of game values to iterative solution of one-shot  $2 \times 2$  games, in complete analogy to standard dynamic programming. We seek a similar result for the switching game, see (23) below.

Without any assumptions on the structure of  $\mathcal{Z}$  appearing in (13), the existence of a pure NEP is not guaranteed. However, as shown by [19] (see also [37] and references therein) a two-person stopping game always admits a mixed NEP. Again, there is no uniqueness and we might need equilibrium selection. Let  $\gamma$  be an  $(\mathcal{F}_t)$ -adapted stochastic process taking values in  $\Delta_4$ . Define the dependent randomized stopping rules (cf. (8))

$$\begin{cases} \tau_1(\gamma) \triangleq \inf\{t : \eta'(t) \leq \gamma_{10}(t) + \gamma_{11}(t)\}, \\ \tau_2(\gamma) \triangleq \inf\{t : \eta'(t) \leq \gamma_{01}(t) + \gamma_{11}(t)\}, \end{cases} \quad \eta'(t) \sim \text{Unif}[0, 1] \quad \text{i.i.d..}$$

Thus, conditional on the game still continuing, the stage- $t$  payoff to player  $i$  is  $\sum_{j,k} \gamma_{jk}(t) Z_i^{jk}(t)$  and total expected payoff is

$$(15) \quad \mathbb{E}^\gamma \left[ \tilde{J}_i(s, \tau_1(\gamma), \tau_2(\gamma)) \right] = \mathbb{E} \left[ \sum_{t=s}^{T-1} \sum_{j,k} \left\{ \left( \prod_{r=s}^{t-1} \gamma_{00}(r) \right) \gamma_{jk}(t) Z_i^{jk}(t) \right\} \right].$$

As before, correlation is implemented through private signals  $\mu_i(t, \gamma)$  and a correlated equilibrium of  $\mathcal{Z}$  is a communication device  $\gamma$  inducing a stopping strategy profile  $\vec{\tau}(\gamma) \triangleq (\tau_1(\gamma), \tau_2(\gamma)) \in \mathfrak{D}^1 \times \mathfrak{D}^2$  such that for  $i = 1, 2$  and all  $0 \leq t < T$

$$(16) \quad V_i(t; \gamma, \mathcal{Z}) \triangleq \mathbb{E}^\gamma[\tilde{J}_i(t, \vec{\tau}(\gamma)) | \mathcal{F}_t] \geq \mathbb{E}^\gamma[\tilde{J}_i(t, \tilde{\tau}_i, \tau_{-i}(\gamma)) | \mathcal{F}_t], \quad \forall \tilde{\tau}_i \in \mathfrak{D}^i(t).$$

Observe that given a device  $\gamma$  leading to a CEP, it must be that

$$(17) \quad V_i(t; \gamma, \mathcal{Z}) = \sup_{\tau \in \mathfrak{D}^i(t)} \mathbb{E}^\gamma \left[ \left( \sum_{s=t}^{(\tau \wedge \tau_{-i}(\gamma)) - 1} Z_i^{00}(s) \right) + Z_i^{10}(\tau) 1_{\{\tau < \tau_{-i}\}} \right. \\ \left. + Z_i^{01}(\tau_{-i}) 1_{\{\tau_{-i} < \tau\}} + Z_i^{11}(\tau) 1_{\{\tau = \tau_{-i}\}} \mid \mathcal{F}_t \right]$$

which is a standard optimal stopping problem for player  $i$  in the enlarged filtration  $\mathcal{G}^i$ .

**Lemma 3.2.** [35, Theorem 2.3] *Consider a CEP with communication device  $\gamma$  of a stopping game  $\mathcal{Z}$ . Then for all  $t \in \{0, 1, \dots, T-1\}$  we have*

$$(18) \quad \begin{cases} \gamma_{00}(t)(\mathbb{E}[V_1(t+1) | \mathcal{F}_t] + Z_1^{00}(t)) + \gamma_{01}(t)Z_1^{01}(t) \geq \gamma_{00}(t)Z_1^{10}(t) + \gamma_{01}(t)Z_1^{11}(t); \\ \gamma_{00}(t)(\mathbb{E}[V_2(t+1) | \mathcal{F}_t] + Z_2^{00}(t)) + \gamma_{10}(t)Z_2^{10}(t) \geq \gamma_{00}(t)Z_2^{01}(t) + \gamma_{10}(t)Z_2^{11}(t); \\ \gamma_{10}(t)Z_1^{10}(t) + \gamma_{11}(t)Z_1^{11}(t) \geq \gamma_{10}(t)(\mathbb{E}[V_1(t+1) | \mathcal{F}_t] + Z_1^{00}(t)) + \gamma_{11}(t)Z_1^{01}(t); \\ \gamma_{01}(t)Z_2^{01}(t) + \gamma_{11}(t)Z_2^{11}(t) \geq \gamma_{01}(t)(\mathbb{E}[V_2(t+1) | \mathcal{F}_t] + Z_2^{00}(t)) + \gamma_{11}(t)Z_2^{10}(t). \end{cases}$$

Lemma 1 shows that a CEP of the stopping game is rational at each stage. For instance, the first inequality in (18) means that conditional on player 1 signal being ‘continue’, the expected payoff to player 1 from continuing (the right-hand-side) is better than the expected payoff from stopping. In either scenario, player 2 implements the conditional strategy  $\vec{\pi}_2(t, \gamma) |_{\mu_1(t, \gamma)=0} = \left( \frac{\gamma_{00}(t)}{\gamma_{00}(t) + \gamma_{01}(t)}, \frac{\gamma_{01}(t)}{\gamma_{00}(t) + \gamma_{01}(t)} \right)$ .

As shown by [35, Theorem 2.4], any finite-horizon stopping game with bounded payoffs admits a CEP; in fact outside the zero-sum and monotone cases we expect that a large number of CEPs

are possible. It is convenient to think of communication device  $\gamma$  leading to a CEP in (16) as a measurable selector of local correlated equilibrium points in the one-shot  $2 \times 2$  games. Thus, let  $\Gamma : \mathbb{T} \times \Omega \times \mathbb{R}^{2 \times 2 \times 2} \rightarrow \Delta_4$  be a measurable map such that for any  $2 \times 2$  game  $H$ ,  $\Gamma(t, \omega, H)$  is a CEP of  $H$ . Then using  $\Gamma$ , one may construct a communication device  $\gamma$  by inductively using the CEP  $\Gamma(t, \omega, H(t, \omega))$ , where  $H(t, \omega)$  is the right-hand-side in (14), and proceeding back in time. Observe that for most  $H$ 's,  $\Gamma(\cdot, H)$  is simply the unique NEP available, so that the selection feature is “silent”, and the device is only really activated when considering the coordination game. With this perspective in mind, we call a *correlation law*  $\Gamma$  a communication device which is based on the same local criterion (for instance “minimize today’s emissions” or “maximize today’s value of player 1”).

**3.3. Recursive Construction.** We return to the emissions market duopoly setup. The emission schedules of the two agents are interpreted as a sequence of regime-changes. Thus, the single-stopping game in the previous section is viewed as the sub-game for making the next regime-switch. The stopping game in Section 3.2 is accordingly denoted as a  $(1, 1)$ -fold switching game and we now will consider  $(n, m)$ -fold switching games with game value functions  $V^{n,m}$ . These games have a restricted set of possible production strategies; namely the total number of regime switches over the game horizon is bounded by  $n$  and  $m$ , respectively. Using the Markov property of the game state and actions it is not surprising that these various switching games are related to each other.

In terms of the notation of Section 3.2, we identify the running profit with  $Z_i^{00}(t) = (a_i P_t - b_i X_t - c_i)u_i(t)$  and the other  $Z_i^{jk}$ 's with various game continuation-values. For the remainder of the section, we make a standing assumption that a communication device  $\gamma$  is chosen and fixed. Let us fix an initial state  $P_s = p, X_s = x$  and initial production regime  $\vec{\zeta} = (\zeta_1, \zeta_2)$ . Define a double cascade of stopping games indexed by  $n$  and  $m$  via

$$(19) \quad V_i^{n,m}(s, p, x, \vec{\zeta}) \triangleq V_i(s; \gamma, \tilde{Z}^{n,m}(\vec{\zeta})), \quad n, m \geq 1$$

which uses the notation of (16) based on the recursive payoff structure

$$(20) \quad \begin{cases} (\tilde{Z}^{n,m})_i^{00}(t, \vec{\zeta}) = (a_i P_t - b_i X_t - c_i)\zeta_i; \\ (\tilde{Z}^{n,m})_i^{01}(t, \vec{\zeta}) = V_i^{n,m-1}(t, P_t, X_t, \zeta_1, 1 - \zeta_2) - 1_{\{i=2\}}K_{\{2, \zeta_2, 1-\zeta_2\}}; \\ (\tilde{Z}^{n,m})_i^{10}(t, \vec{\zeta}) = V_i^{n-1,m}(t, P_t, X_t, 1 - \zeta_1, \zeta_2) - 1_{\{i=1\}}K_{\{1, \zeta_1, 1-\zeta_1\}}; \\ (\tilde{Z}^{n,m})_i^{11}(t, \vec{\zeta}) = V_i^{n-1,m-1}(t, P_t, X_t, 1 - \zeta_1, 1 - \zeta_2) - K_{\{i, \zeta_i, 1-\zeta_i\}}. \end{cases}$$

The boundary cases are first

$$V_i^{0,0}(s, p, x, \vec{\zeta}) \triangleq \mathbb{E}^{\vec{\zeta}} \left[ \sum_{t=s}^{T-1} (a_i P_t - b_i X_t - c_i)\zeta_i \right];$$

next,  $V_1^{n,0}(s, p, x, \vec{\zeta})$  and  $V_2^{0,m}(s, p, x, \vec{\zeta})$  are identified with the single-player optimization problems as in (12) (keeping the emission regime of the other player fixed at  $\zeta_{-i}$ ). Finally, we take

$$V_2^{n,0}(s, p, x, \vec{\zeta}) = \mathbb{E}^{(u_1^{n,*}, \zeta_2, \gamma)} \left[ \sum_{t=s}^{T-1} (a_2 P_t - b_2 X_t - c_2)\zeta_2 \right]$$

where  $u_1^{n,*}$  is an optimal control for the problem defining  $V_1^{n,0}$ , and similarly for  $V_1^{0,m}(s, p, x, \vec{\zeta})$ .

**3.4. Switching Game Equilibrium as Sequential Stopping Game Equilibrium.** We now proceed to *glue* the sequential stopping games of  $V_i^{n,m}$  and re-interpret the latter as value functions of a switching game. For  $n \geq 0$  denote by  $\mathcal{U}_i^n \subset \mathcal{U}_i$  the set of all production strategies for player  $i$  with at most  $n$  switches. Consider the restricted repeated game with payoffs (3) where we require  $u_1 \in \mathcal{U}_1^n$  and  $u_2 \in \mathcal{U}_2^m$ , so that the first producer may change her production regime at most  $n$  times, and the second producer at most  $m$  times.

Our first task is to obtain a switching-game CEP that matches the definition of  $V^{n,m}$ . To do so we pick a correlation law  $\Gamma$ ;  $\Gamma$  gives rise to a CEP of any stopping game, in particular it leads to well-defined game values  $V_i^{n,m}$  in (19). We now construct a communication device  $\gamma^{n,m}$  for the  $(n, m)$ -switching game. Let  $k_i(t)$  be the number of production switches used by player  $i$  by stage  $t$ . The device  $\gamma^{n,m}(t)$  at stage  $t$  is taken to be  $\Gamma\left(t, \omega, \tilde{Z}^{n-k_1(t), m-k_2(t)}(\vec{u}(t))\right)$  defined in terms of (20) and the latest regime  $\vec{u}(t)$ . Note that the overall  $\gamma^{n,m}$  is no longer Markovian since it has memory of the number of switches made by each player, which is necessary in the constrained game. The above construction is well-defined for all paths of  $(P, X, \vec{u})$ , even outside equilibrium.

Using  $\gamma^{n,m}$  we proceed to construct switching controls  $u_i^{n,m}$  for the  $(n, m)$ -switching game. To simplify notation we write  $\underline{\tau}^{n,m} = \tau_1(\gamma^{n,m}) \wedge \tau_2(\gamma^{n,m})$  which is interpreted as the equilibrium first stopping time for the game defined by (19) under the correlation law  $\Gamma$ . Given the starting production regime  $\vec{\zeta} = (\zeta_1, \zeta_2)$ , let us define the switching controls  $u_i^{n,m}(s)$  for this game by

$$(21) \quad u_1^{n,m}(s) = \begin{cases} \zeta_1 & \text{for } s < \underline{\tau}^{n,m}; \\ \begin{cases} 1 - \zeta_1 & \text{for } \underline{\tau}^{n,m} \leq s < \underline{\tau}^{n-1,m} \\ \zeta_1 & \text{for } \underline{\tau}^{n,m} \leq s < \underline{\tau}^{n,m-1} \\ 1 - \zeta_1 & \text{for } \underline{\tau}^{n,m} \leq s < \underline{\tau}^{n-1,m-1} \end{cases} & \begin{cases} \text{when } \tau_1^{n,m} < \tau_2^{n,m}; \\ \text{when } \tau_2^{n,m} < \tau_1^{n,m}; \\ \text{when } \tau_1^{n,m} = \tau_2^{n,m}, \end{cases} \end{cases}$$

... and so on,

and similarly for  $u_2^{n,m}(t)$ . In words,  $u_i^{n,m}$  keeps track of the production regime of the  $i$ -th agent following the decision rules defined sequentially by descending through the family of the  $V^{n,m}$ -stopping subgames (one stopping game at a time). Then by definition of (19) we have  $u_1^{n,m} \in \mathcal{U}_1^n$  and  $u_2^{n,m} \in \mathcal{U}_2^m$ . It can also be seen through an easy induction argument that

$$(22) \quad V_i^{n,m}(s, p, x, \vec{\zeta}) = V_i(s, p, x, \vec{\zeta}; \vec{u}^{n,m}),$$

so that the switching control  $\vec{u}^{n,m}$  of (21) allows to achieve the game values  $V^{n,m}$  defined recursively in (19). Moreover, the next theorem shows that the pair  $(u_1^{n,m}, u_2^{n,m})$  is in fact a correlated equilibrium (using correlation device  $\gamma^{n,m}$ ) for the game (3) over the control set  $\mathcal{U}_1^n \times \mathcal{U}_2^m$ .

**Theorem 3.3.** *For all  $n > 0$  and  $u_1 \in \mathcal{U}_1^n$  we have  $V_1(t, \cdot; u_1^{n,m}, u_2^{n,m}) \geq V_1(t, \cdot; u_1, u_2^{n,m})$ . Similarly for all  $m > 0$  and  $u_2 \in \mathcal{U}_2^m$  we have  $V_2(t, \cdot; u_1^{n,m}, u_2^{n,m}) \geq V_2(t, \cdot; u_1^{n,m}, u_2)$ .*

*Proof.* The idea of the proof is to make use of the Markov structure of our problem and apply induction. The other key tool is that given  $\gamma^{n,m}$ , we can look at one player at a time which essentially reduces to a single-player problem studied before, see (17).

Due to symmetry, it suffices to prove the result for player 1. When  $m = 0$  the other player cannot act, the game becomes trivial and Theorem 3.3 is just a re-statement of Proposition 3.1. Conversely, when  $n = 0$ , the first player cannot act and there is nothing to prove. Using induction we assume that the theorem has been shown for the pairs  $(n-1, m-1)$ ,  $(n, m-1)$  and  $(n-1, m)$ ; let us show it for the case  $(n, m)$ . Given an arbitrary  $u_1 \in \mathcal{U}_1^n$ , write it as  $u_1 = (\tau^1, \hat{u}_1)$  where  $\hat{u}_1 \in \mathcal{U}_1^{n-1}$  denotes the remainder of  $u_1$  after the first switch time  $\tau^1$ . Let  $\tau^{2,*} \equiv \tau^2(\gamma^{n,m})$  be the first switch for the second player dictated through  $\gamma^{n,m}$ . Define  $\underline{\tau} = \tau^1 \wedge \tau^{2,*}$ . Also for notational convenience we omit all the arguments of  $V^{n,m}$  except for the time variable. Then the strong Markov property of  $(P, X)$  and the way  $u_2^{n,m}(t)$  was constructed show that

$$\begin{aligned} \mathbb{E}^{(u_1, u_2^{n,m}, \gamma^{n,m})} \left[ \sum_{s=\underline{\tau}}^{T-1} (a_1 P_s - b_1 X_s - c_1) \hat{u}_1(s) \right] &= \mathbb{E}^{(u_1, u_2^{n,m}, \gamma^{n,m})} \left[ V_1(\underline{\tau}; \hat{u}_1, u_2^{n-1,m}) 1_{\{\tau^1 < \tau^{2,*}\}} \right. \\ &\quad \left. + V_1(\underline{\tau}; \hat{u}_1, u_2^{n,m-1}) 1_{\{\tau^1 > \tau^{2,*}\}} + V_1(\underline{\tau}; \hat{u}_1, u_2^{n-1,m-1}) 1_{\{\tau^1 = \tau^{2,*}\}} \right]. \end{aligned}$$

Conditioning on  $\tau^1$  and  $\tau^{2,*}$  we therefore have  $V_1(t; u_1, u_2^{n,m}) =$

$$\begin{aligned} &\mathbb{E}^{(u_1, u_2^{n,m}, \gamma^{n,m})} \left[ \left( \sum_{s=t}^{\tau-1} (a_1 P_s - b_1 X_s - c_1) u_1(t) \right) + \left( \sum_{s=\tau^1}^{T-1} (a_1 P_s - b_1 X_s - c_1) \hat{u}_1(s) \right) 1_{\{\tau^1 < \tau^{2,*}\}} \right. \\ &\quad \left. + \left( \sum_{s=\tau^{2,*}}^{T-1} (a_1 P_s - b_1 X_s - c_1) \hat{u}_1(s) \right) 1_{\{\tau^1 > \tau^{2,*}\}} + \left( \sum_{s=\tau^1}^{T-1} (a_1 P_s - b_1 X_s - c_1) \hat{u}_1(s) \right) 1_{\{\tau^1 = \tau^{2,*}\}} \right] \\ &= \mathbb{E}^{(u_1, u_2^{n,m}, \gamma^{n,m})} \left[ \left( \sum_{s=t}^{\tau-1} (a_1 P_s - b_1 X_s - c_1) u_1(t) \right) + V_1(\underline{\tau}; \hat{u}_1, u_2^{n-1,m}) 1_{\{\tau^1 < \tau^{2,*}\}} \right. \\ &\quad \left. + V_1(\underline{\tau}; \hat{u}_1, u_2^{n,m-1}) 1_{\{\tau^1 > \tau^{2,*}\}} + V_1(\underline{\tau}; \hat{u}_1, u_2^{n-1,m-1}) 1_{\{\tau^1 = \tau^{2,*}\}} \right] \end{aligned}$$

by induction hypothesis we have the inequality

$$\begin{aligned} &\leq \mathbb{E}^{(u_1, u_2^{n,m}, \gamma^{n,m})} \left[ \left( \sum_{s=t}^{\tau-1} (a_1 P_s - b_1 X_s - c_1) u_1(t) \right) + V_1(\tau^1; u_1^{n-1,m}, u_2^{n-1,m}) 1_{\{\tau^1 < \tau^{2,*}\}} \right. \\ &\quad \left. + V_1(\tau^{2,*}; u_1^{n,m-1}, u_2^{n,m-1}) 1_{\{\tau^1 > \tau^{2,*}\}} + V_1(\tau^1; u_1^{n-1,m-1}, u_2^{n-1,m-1}) 1_{\{\tau^1 = \tau^{2,*}\}} \right] \\ &\leq \sup_{\tau^1 \in \mathcal{D}^1(t)} \mathbb{E}^{(u_1(t), u_2^{n,m}, \gamma^{n,m})} \left[ \left( \sum_{s=t}^{\tau-1} (a_1 P_s - b_1 X_s - c_1) u_1(t) \right) + V_1(\tau^1; \bar{u}^{n-1,m}) 1_{\{\tau^1 < \tau^{2,*}\}} \right. \\ &\quad \left. + V_1(\tau^{2,*}; \bar{u}^{n,m-1}) 1_{\{\tau^1 > \tau^{2,*}\}} + V_1(\tau^1; \bar{u}^{n-1,m-1}) 1_{\{\tau^1 = \tau^{2,*}\}} \right] \\ &= V_1(t; \bar{u}_1^{n,m}), \end{aligned}$$

where the last line uses the relationship (22), the construction of the stopping game defining  $V^{n,m}$  in (20), and property (17).  $\square$

The above construction leads to the key result of this section that characterizes CEP of switching games, establishes their existence, and gives a recursive formula for the resulting game value functions. For a  $2 \times 2$  game  $H$  defined in (9) and correlation device  $\gamma$  we denote the respective game values as

**Theorem 3.4.** *Fix a correlation law  $\Gamma$ . Then  $\Gamma$  gives rise to a CEP of the switching game (3). Moreover, the corresponding value functions  $V_i(t, P_t, X_t, \vec{\zeta}; \Gamma)$  solve*

$$(23) \quad \begin{cases} V_1(t, P_t, X_t, \vec{\zeta}) = \gamma_{00}(t)Y_1(\zeta_1, \zeta_2) + \gamma_{01}(t)Y_1(\zeta_1, 1 - \zeta_2) \\ \quad + \gamma_{10}(t)(Y_1(1 - \zeta_1, \zeta_2) - K_{1,\zeta_1,1-\zeta_1}) + \gamma_{11}(t)(Y_1(1 - \zeta_1, 1 - \zeta_2) - K_{1,\zeta_1,1-\zeta_1}) \\ V_2(t, P_t, X_t, \vec{\zeta}) = \gamma_{00}(t)Y_2(\zeta_1, \zeta_2) + \gamma_{01}(t)(Y_2(\zeta_1, 1 - \zeta_2) - K_{2,\zeta_2,1-\zeta_2}) \\ \quad + \gamma_{10}(t)Y_2(1 - \zeta_1, \zeta_2) + \gamma_{11}(t)(Y_2(1 - \zeta_1, 1 - \zeta_2) - K_{2,\zeta_2,1-\zeta_2}) \end{cases}$$

where  $Y_i(t, \vec{\zeta}) \triangleq \mathbb{E}^{\vec{\zeta}}[V_i(t+1, \vec{\zeta}) | \mathcal{F}_t] + (a_i P_t - b_i X_t - c_i) \zeta_i$  and we have omitted the dependence on  $t$ . The equilibrium controls can be taken as  $u_i^* \equiv u_i^{T,T}$ , as defined in (21).

*Proof.* We wish to take  $n, m \rightarrow \infty$  in Theorem 3.3. Because for  $n > m$ ,  $\mathcal{U}^m \subseteq \mathcal{U}^n$ , it follows that for a fixed  $m$ ,  $V_1^{n,m}$  is increasing in  $n$  (and for a fixed  $n$ ,  $V_2^{n,m}$  is increasing in  $m$ ). For our discrete-time game, at most  $T$  regime switches are possible for each player. Therefore  $u_i^* \in \mathcal{U}_i^T$  and it follows that  $V_i^{n,n} \equiv V_i$  for all  $n > T$ . In particular, a switching CEP based on  $\Gamma$  results by using  $\gamma^{T,T}$ .

Moreover, at equilibrium at most one switch is made at any given stage due to the triangle condition on  $K_{i,j,k}$ . Therefore, if it is optimal to switch at stage  $t$  from  $\vec{\zeta}$  to  $\vec{u}$ , then already starting at regime  $\vec{u}$  at  $t$  and same state variables it is optimal to make no changes, so that  $V_i(t, \vec{u}) = \mathbb{E}^{\vec{u}}[V_i(t+1, \vec{u}) | \mathcal{F}_t] + (a_i P_t - b_i X_t - c_i) u_i$  for that scenario. Combining these facts with the form of (14) and dropping the constraints on the number of switches, we may express all payoffs in terms of next-stage game values. The recursion (23) is now obtained by making this substitution in (14).  $\square$

#### 4. NUMERICAL IMPLEMENTATION

Theorem 3.4 shows that a game value and equilibrium strategy profile can be obtained recursively by solving the 1-period 2-by-2 games in (23). The payoffs of those games are given in terms of *conditional expectations* of next-stage game values. Therefore, a numerical implementation hinges on accurate evaluation of these expectations. Since our state-space in  $(P, X)$  is continuous, it is impossible to make this computation exactly. Below we present two possible approximation approaches.

**4.1. Markov Chain Approximation Algorithm.** Our model would be simplified if the continuous state space of  $(P, X)$  is discretized. Let  $(\tilde{P}, \tilde{X})$  be an approximating discrete-state process with  $(\tilde{P}_t, \tilde{X}_t)$  living on a finite subset  $D_t \subset \mathbb{R}_+^2$ . If the pair  $(\tilde{P}, \tilde{X})$  is furthermore chosen to be again Markov, this is known as the Markov Chain Approximation (MCA) method of [25]. With such

$(\tilde{P}, \tilde{X})$ , a conditional expectation  $\mathbb{E}[f(P_{t+1}, X_{t+1}) | P_t = p, X_t = x] \simeq \mathbb{E}[f(\tilde{P}_{t+1}, \tilde{X}_{t+1}) | \tilde{P}_t = p, \tilde{X}_t = x]$  for any measurable function  $f$  is just a weighted sum based on the transition probability matrix of  $(\tilde{P}, \tilde{X})$ . The backward recursion in (23) for  $\tilde{V}_i$ , the corresponding approximation of  $V_i$ , can now be implemented directly for each stage  $t$  and each possible state of  $(\tilde{P}_t, \tilde{X}_t) \in D_t$ . A well-known procedure constructs  $(\tilde{P}, \tilde{X})$  by taking  $D_t$  to be a 2-dimensional regular grid or lattice and allowing state transitions only between neighboring grid points. Moreover, the transition probabilities of  $(\tilde{P}, \tilde{X})$  are chosen so as to have *local consistency* in the first two moments with the 1-step transition densities of  $(P, X)$ ; see [25, Chapter 5].

To use this approach in our model, one must take into account the price impact. Therefore, we construct four approximations  $(\tilde{P}, \tilde{X}^{\vec{\zeta}})$  indexed by the possible joint production regimes  $\vec{\zeta} \in \{0, 1\}^2$  that induce different local dynamics of  $\tilde{X}^{\vec{\zeta}}$ , see (2). In other words, our effective state variables are  $(\tilde{P}, \tilde{X}, \vec{\zeta})$ . For every possible combination  $(t, \tilde{p}, \tilde{x}, \vec{\zeta}) \in \mathbb{T} \times D_t \times \{0, 1\}^2$  the relation (23) is then solved through backward recursion. A generic convergence proof (as the grid spacing tends to zero) of this procedure for finite-horizon non-zero-sum stochastic games was obtained in [24]. Note that in our model the controls  $\vec{u}(t)$  are discrete and finite-valued and therefore all the compactness conditions in [24] for the control space are automatically satisfied.

**4.2. Least Squares Monte Carlo Approach.** Like classical dynamic programming, the MCA method above suffers from the curse of dimensionality. Indeed, the size of the approximating grid grows exponentially in the dimension of the state variables. In our basic model  $(P, X)$  are two-dimensional; however realistic implementations are likely to take multi-dimensional factor models for  $P$  and (possibly)  $X$ . Thus, it is helpful to seek a more robust algorithm.

A seminal idea due to [9, 15, 27] is to use a cross-sectional regression combined with a Monte Carlo simulation to compute the relevant conditional expectations. The key step is a global approximation of the maps  $(t, p, x, \vec{\zeta}) \mapsto V_i(t, p, x, \vec{\zeta})$  and equilibrium one-step strategies  $(t, p, x, \vec{\zeta}) \mapsto \vec{u}(t, p, x, \vec{\zeta})$  (based on a fixed correlation law  $\Gamma$ ) via a random sample of  $(P_t, X_t)$ . The construction is iterative and backward in time.

Suppose that the current date is  $t$  and we already know all the approximations  $v_i(s, p, x, \vec{\zeta}) \simeq V_i(s, p, x, \vec{\zeta})$  for  $s > t$  and the corresponding equilibrium strategy profiles. Given a collection of initial points  $(p_t^n, x_t^n)$ , for  $n = 1, \dots, N$ , and an arbitrary starting emission regime  $\vec{\zeta} = \vec{u}^n(t)$  we first simulate the future *cashflows* on  $[t + 1, T]$  for each scenario  $n$ . This is done by iteratively updating  $(p_{s+1}^n, x_{s+1}^n)$  through an independent draw from the conditional law  $\mathbb{P}^{\vec{u}^n(s)}$  and then computing the equilibrium actions  $u_i^n(s)$  of each player for  $s = t + 1, \dots, T$  based on the estimated future game values  $v_i(s, p_s^n, x_s^n, \vec{u}^n(s))$  and the chosen communication device  $\Gamma$ . If  $\Gamma$  leads to randomized strategies, such a randomization is naturally implemented as part of this simulation. The realized pathwise cashflow  $\vartheta_i^n(t + 1, \vec{\zeta})$  represents an empirical draw from  $V_i(t + 1, P_{t+1}, X_{t+1}^{\vec{\zeta}}, \vec{\zeta})$  conditional on  $P_t = p_t^n, X_t = x_t^n$ . We now perform a cross-sectional regression of  $(\vartheta_i^n(t + 1, \vec{\zeta}))_{n=1}^N$  against  $(p_t^n, x_t^n)_{n=1}^N$  by using a collection of basis functions  $B_\ell(t, p, x)$ ,  $\ell = 1, \dots, r$ . The regression yields



the predicted continuation values

$$\hat{v}_i(t, p_t^n, x_t^n, \vec{\zeta}) \simeq \mathbb{E}^{\vec{\zeta}} \left[ V_i(t+1, P_{t+1}, X_{t+1}, \vec{\zeta}) \mid P_t = p_t^n, X_t = x_t^n \right].$$

Finally, using  $\hat{v}_i$  together with the current payoffs and switching costs and the correlation law  $\Gamma$  we solve for the equilibrium game values  $v_i(t, p_t^n, x_t^n, \vec{u})$  for each production regime  $\vec{u}$  by applying the stage- $t$  sub-game of Theorem 3.4. The computed game equilibrium also provides the map  $(t, p_t^n, x_t^n, \vec{u}) \mapsto \vec{u}^*(t)$  for the equilibrium strategies. The regression results allow to further extend this to arbitrary initial condition  $(t, p, x, \vec{u})$ . Working back to  $t = 0$ , the final answer (which is a random variable depending on the Monte Carlo sample) is simply the average  $V_i(0, p_0, x_0, \vec{\zeta}_0) \simeq \frac{1}{N} \sum_n v_i(0, p_0^n, x_0^n, \vec{u}_0)$ .

The initial collection  $(p_t^n, x_t^n)_{t=1}^T$  is obtained by simulation. Since,  $X$  is affected by the price impact of  $\vec{u}$ , to simulate  $(p_t^n, x_t^n)$  we need to select some anterior auxiliary strategy profile  $\vec{u}^0$ . While in theory  $\vec{u}^0$  can be arbitrary, in practice it should be close to the equilibrium  $\vec{u}^*$ . Indeed, the collection  $(v_i(t, p_t^n, x_t^n, \vec{\zeta}))_{n=1}^N$  is supposed to approximate  $V_i(t, P_t, X_t^*, \vec{u}(t))$  where  $X_t^*$  is the equilibrium  $CO_2$  allowance price. Because  $v_i$ 's are computed by employing regression, the resulting approximation cannot be uniformly good on  $\mathbb{R}_+^2$ . From the point of view of accurate solutions, it needs to be good around the region of interest for  $X_t^*$ . Thus, we need most of the  $x_t^n$ 's to be in that (*a priori* unknown) neighborhood. To overcome this difficulty, as the algorithm works back through time, the future paths  $(p_s^n, x_s^n)$ ,  $s > t$  are re-computed using the now-available (approximately) equilibrium strategies  $u^*(s)$ . To further mitigate the problem, we iteratively re-do the whole simulation and subsequent backward recursion a few times (in practice three iterations suffice), using the computed  $\vec{u}^*$  from one iteration as the anterior  $\vec{u}^0$  in the next one. The Appendix summarizes the above scheme in pseudo-code.

Selection of basis functions should reflect the expected shape of  $(p, x) \mapsto V_i(t, p, x, \vec{\zeta})$ . A typical choice is to use low-degree polynomial basis functions, such as  $p, p^2, x, x^2$ , etc. In practice,  $r = 5 - 7$  basis functions and  $N = 32000 - 50000$  paths suffice. A large degree of customization, such as time-varying bases, constrained least-squares regression, variance reduction methods, etc., is possible to speed up the computations. The Appendix summarizes the above scheme in pseudo-code in Algorithm 2. It calls as a sub-routine Algorithm 1 that carries out the forward simulations of  $\vartheta_i^n$ . The cost of simulations in Algorithm 2 is  $\mathcal{O}(N \cdot T^2)$  which consists of re-simulating  $N$  paths on  $[t, T]$  as  $t$  goes from  $T - 1$  to zero (see Algorithm 1). The cost of doing regression against  $r$  basis functions on each path and for each stage is  $\mathcal{O}(N \cdot T \cdot r^3)$  and the cost of computing continuation values is  $\mathcal{O}(N \cdot T^2 \cdot r)$ . The memory requirements from storing all the simulation paths are  $\mathcal{O}(N \cdot T)$ .

**4.3. Numerical Examples.** In this section we illustrate our analysis with a numerical case-study. The selected model parameters are listed in Table 1. The example represents emission scheduling of two producers over one calendar year; all the parameters of  $(P, X)$  are in annualized units and we use bi-weekly periods  $T' = 26$  to model the scheduling flexibility. Note that the electricity price  $P_t$  is more volatile than the  $CO_2$  allowance price  $X_t$ ; also the mean-reversion parameter  $\kappa_X$  of  $X$  is quite large, implying a significant price impact. In (2) we take  $f(\zeta_1, \zeta_2) = \log(12 + 8\zeta_1 + 4\zeta_2)$ ,

$\kappa_X$	3	$\sigma_X$	0.25		
$\kappa_P$	2	$\sigma_P$	0.4		
$T$	1	$\rho$	0.6		
$\bar{P}$	45	$\bar{X}$	12		
$P_0$	45	$X_0$	15		

Producer 1		Producer 2	
$a_1$	1	$a_2$	2
$b_1$	2	$b_2$	1
$c_1$	10	$c_2$	80
$g_1$	8	$g_2$	4
$K_1$	0.2	$K_2$	0.2

TABLE 1. Model Parameters for the Examples in Section 4.3.

Correlation Law	$V_1(0, P_0, X_0)$	$V_2(0, P_0, X_0)$
Utilitarian	5.30	4.14
Egalitarian	5.33	4.20
Preferential 1	5.39	4.11
Preferential 2	5.02	4.24

TABLE 2. Comparison of equilibrium game values for different correlation laws  $\Gamma$ . Standard errors of the Monte Carlo scheme are about 1%. Parameters are as given in Table 1.

so that the mean-reversion level of  $\log X$  is linear in the production regimes of producers 1 and 2, with producer 1 having more influence due to emitting twice as much carbon,  $b_1 = 2b_2 \Rightarrow g_1 = 2g_2$ . The stylized production/emission parameters represent a dirty “coal” producer 1 who has low input costs but needs lots of allowances, and a clean “natural gas” producer 2 who has high fixed costs but small sensitivity to allowance prices (and can generate twice as much electricity). Observe that if both producers emit simultaneously for a long period of time, then we expect  $P_t \sim \bar{P} = 45, X_t \sim f(1, 1) = 24$  meaning that everyone will be losing money. Therefore, extended joint emissions are not *sustainable*.

A large variety of CEPs are possible in our model; Table 2 shows the game values corresponding to four representative correlation laws. These values were obtained by running Algorithm 2 discussed in Section 4.2 using  $N = 40000$  paths, and the basis functions  $\{1, p, x, x^2, (2p - x - 80)_+, (p - 2x - 10)_+\}$ . We find that different correlation laws modify the expected profit of the producers by 3%–5%. As expected, individual producer values are maximized by the preferential equilibria that always favor the respective player. Counter-intuitively, the egalitarian CEP (which maximizes at each stage the minimum continuation value) produces larger game values to both producers than the utilitarian CEP (which maximizes the sum of continuation values). This occurs because the correlation law is applied *stage-wise* and optimizes a local criterion; there is no guarantee that the corresponding global criterion is respected. A similar phenomenon was observed in [35, Section 5.4].

To illustrate the equilibrium strategy profiles, Figure 1 shows the empirical regions in the  $(P, X)$ -space corresponding to different equilibrium strategies at a fixed date  $t = 7$  (i.e. about three

months into the year) using the Preferential-1 correlation law that always favors producer 1. As expected, when the current P&L of both producers is strongly negative (upper-left corner), the equilibrium action is  $\bar{u}^*(t) = (0, 0)$ ; when it is strongly positive (large  $P_t$ ) the equilibrium is to generate electricity  $\bar{u}^*(t) = (1, 1)$ . Because of the differing carbon-efficiencies of the producers, there are also large regions where exactly one producer can generate profit (e.g. around  $\{P_t \in [40, 45], X_t^* \in [10, 12]\}$  only producer 2 is profitable). However at the border regions, the price impact and competition create new effects. In Figure 1, we observe the emergence of a local anti-coordination game around  $\{(P_t, X_t^*) = (50, 15)\}$ , and a competitive game around  $\{(P_t, X_t^*) = (50, 12)\}$ . We cannot analytically verify whether a particular type of game may emerge locally; thus the competitive game region in Figure 1 could be either a true phenomenon or an aberration due to numerical errors (e.g. poor regression fit in that region). Note that most simulated equilibrium paths for  $X_t^*$  stay above  $x = 13$ , so the competitive game scenario at  $t = 7$  is very unlikely to be realized (i.e. very few paths hit that region).

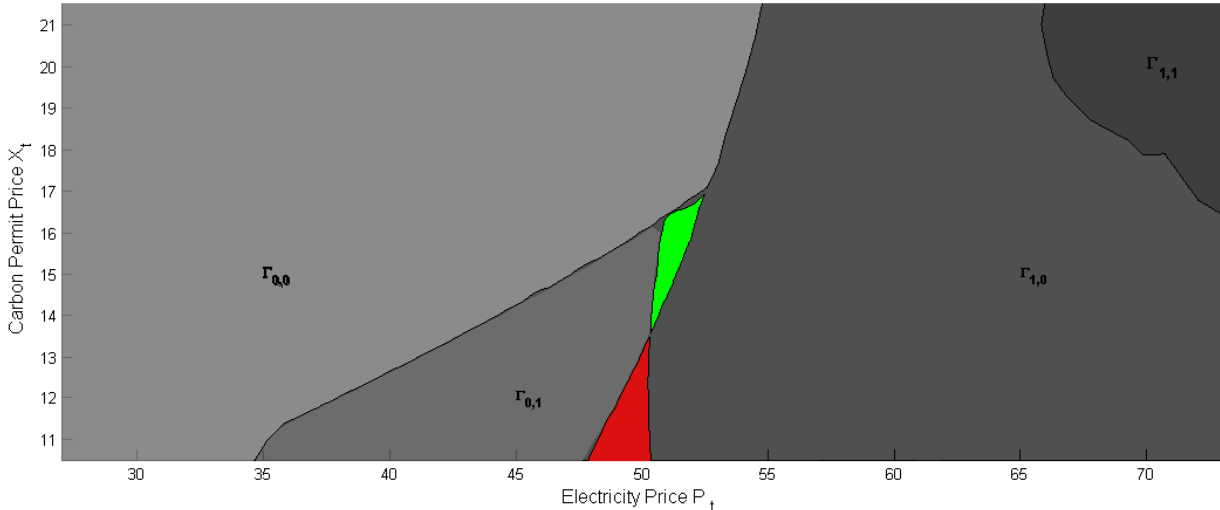


FIGURE 1. Equilibrium game strategy  $\bar{u}^*(t)$  as a function of  $(P_t, X_t^*)$  for  $t = 7$  and  $\vec{\zeta} = (0, 0)$ . The green region denotes the anti-coordination game-type where the Preferential-1 correlation law is used, and the red region denotes the competitive game-type where the unique mixed NEP is chosen. Elsewhere, we label the regions according to the unique pure NEP implemented.

To better illustrate the optimal strategy over time, Figure 2 shows a sample path of the equilibrium price ( $X_t^*$ ) for one  $\omega$ . Analogously to single-player problems, the  $CO_2$  allowance price undergoes hysteresis cycles [13]. Thus, when ( $X_t^*$ ) is low, production becomes profitable. This leads to increased emissions and  $X_t^*$  tends to rise through the price impact mechanism. In turn, the ensuing higher emission costs eventually curtail production and  $X_t^*$  falls back. The presence of switching costs  $K_i$  lowers the scheduling flexibility of the producers and further amplifies this cycle through inertia.

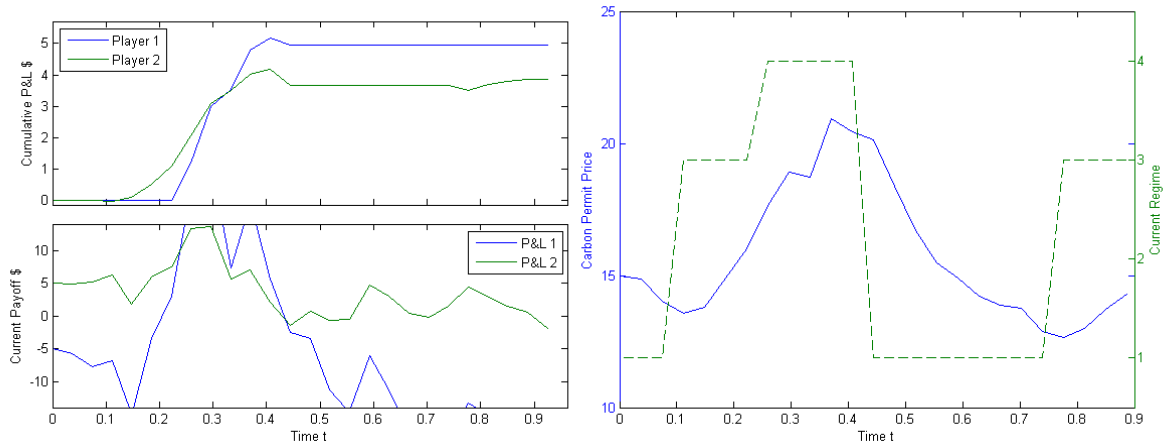


FIGURE 2. Sample equilibrium path of the emissions game. Top left panel: cumulative realized P&L of the players as a function of  $t$ . Bottom left panel: the electricity-carbon spread of each producer for the current time step. Right panel: evolution of the controlled equilibrium allowance price  $X_t^*$ , as well as the implemented strategy  $\bar{u}^*(t) \in \{00, 01, 10, 11\} \equiv \{1, \dots, 4\}$ . The panels were generated using Algorithm 1 given in the Appendix.

## 5. CONCLUSION

In this paper we studied a new type of stochastic games which were motivated by dynamic emission schedules of energy producers under cap-and-trade schemes. Because multiple game equilibria can emerge, we explored various correlated equilibria. It is an interesting economic policy question which equilibrium is likely/desirable to be implemented and how the regulator can steer market participants towards that choice. For example, putting a price on emissions is supposed to partially drive out “dirty” producers. It would be an intriguing exercise to study how much these effects depend on equilibrium selection and whether *blockading* of inefficient polluters is possible under some equilibria.

In our simplified model, the producers only made binary emission decisions at each stage. On a practical level, much finer granularity is available. It would be straightforward to extend our problem and allow a more general finite-state control set of size  $|\mathcal{A}|$ . The only modification would be to replace the  $2 \times 2$  bimatrix games with a more general  $\mathcal{A} \times \mathcal{A}$  bimatrix. The theory for more than two producers is incomplete and it is an open problem to establish existence of CEP/NEP for multi-player stopping games (see [38] for current state-of-the-art).

**5.1. Further Extensions.** Several aspects of our model merit further analysis. The dynamics for  $CO_2$  allowance prices in (2) were selected to capture succinctly the price impact of each producer, leaving out other important features. As described in the introduction, as the permit expiration date  $T$  approaches, the  $CO_2$  price should converge either to zero (if excess permits remain) or to a fixed upper bound  $\bar{x}$  (the penalty for emitting without an allowance). New (time-dependent) stochastic

models are needed to mimic this property, see [7, 18]. Also, some cap-and-trade proposals will allow trading of allowances by financial participants whence *no-arbitrage* restrictions might have to be imposed on the dynamics of  $X$ . All these possibilities can be handled straightforwardly, since the main construction is for arbitrary  $X$ -dynamics. Ideally, a fully endogenous model is desired for allowance prices; namely  $X_t$  should be a function of total expected emissions until  $T$  compared to total current supply, i.e. have a characterization in terms of conditional expectations of future equilibrium emission schedules. See [6, 7, 10] for such price-formation models and related general equilibrium frameworks. These extensions will be considered in forthcoming papers.

Our formulation was in discrete-time; while this is sufficient for practical purposes, it is of great theoretical interest to construct a continuous-time model counterpart. The overall structure of a switching game as a sequence of stopping games straightforwardly carries over to continuous-time. However, description of correlated stopping equilibria in continuous time has not been attempted so far. In fact, the only reference dealing with randomized continuous-time stopping games is [40] (see also [26] for the latest results on general continuous timing games). Note that in continuous-time one must work with Nash  $\epsilon$ -equilibria since all stopping strategies are defined only in the almost-sure sense. Second, to ensure the representation of  $V_i$  as iterative stopping games through  $V_i^{n,m}$ , it is necessary to *a priori* show that each player makes finitely many regime switches. At this point we are not able to state any conditions to guarantee this, except requiring mandatory “cool-off” periods between each emission regime switch.

In our Markovian setting, solutions of continuous-time single-player switching problems have representations in terms of reflected backward stochastic differential equations (BSDE) [20]. This representation should continue to hold in a game setting and will be explored in a separate paper. Related results have already been obtained for stochastic differential game analogues of our setup, whence  $\bar{u}(t)$  has continuous state-space, see [21, 22].

**Acknowledgment.** I am grateful to the anonymous referees for their valuable suggestions that improved the final presentation. Thanks also to René Carmona for many useful discussions and David Ramsey, Krzysztof Szajowski, Jianfeng Zhang and participants at the MSRI Workshop on Economic Games and Mechanisms to Address Climate Change (May 2009) and the IPAM New Directions in Financial Mathematics Workshop (January 2010) for their feedback on earlier versions of this paper.

## REFERENCES

- [1] L. H. R. Alvarez. A class of solvable stopping games. *Appl. Math. Optim.*, 58(3):291–314, 2008.
- [2] R. J. Aumann. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, 55(1):1–18, 1987.
- [3] S. Basar and G. J. Olsder. *Dynamic noncooperative Game Theory*. SIAM Classics in Applied Mathematics, 1999.
- [4] M. Brennan and E. Schwartz. Evaluating natural resource investments. *J. Business*, 58:135–157, 1985.
- [5] A. Calvó-Armengol. The set of correlated equilibria of  $2 \times 2$  games. Technical report, Universitat Autònoma de Barcelona, 2006.
- [6] R. Carmona, M. Fehr, and J. Hinz. Optimal stochastic control and carbon price formation. *SIAM J. Control Optim.*, 48(4):2168–2190, 2009.

- [7] R. Carmona, M. Fehr, J. Hinz, and A. Porchet. Market design for emission trading schemes. *SIAM Review*, to appear, 2009.
- [8] R. Carmona and M. Ludkovski. Pricing asset scheduling flexibility using optimal switching. *Appl. Math. Finance*, 15(4):405–447, 2008.
- [9] J. F. Carrière. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Math. Econom.*, 19:19–30, 1996.
- [10] M. Chesney and L. Taschini. The endogenous price dynamics of emission allowances: An application to  $CO_2$  option pricing. Technical Report Research Paper No. 08-02, Swiss Finance Institute, 2009. Available at SSRN: <http://ssrn.com/abstract=1090150>.
- [11] J. Cvitanić and I. Karatzas. Backward stochastic differential equations with reflection and Dynkin games. *Ann. Probab.*, 24(4):2024–2056, 1996.
- [12] S. Dayanik and M. Egami. A direct solution method for optimal switching problems of one-dimensional diffusions, 2005. working paper.
- [13] A. Dixit and R. S. Pindyck. *Investment Under Uncertainty*. Princeton University Press, 1994.
- [14] E. Dynkin. A game-theoretic version of an optimal stopping problem. *Soviet Math. Dokl.*, 10:270–274, 1969.
- [15] D. Egloff. Monte Carlo algorithms for optimal stopping and statistical learning. *Ann. Appl. Probab.*, 15(2):1396–1432, 2005.
- [16] E. Ekström and G. Peskir. Optimal stopping games for Markov processes. *SIAM J. Control Optim.*, 47(2):684–702, 2008.
- [17] A. Eydeland and K. Wolyniec. *Energy and Power Risk Management: New Developments in Modeling, Pricing and Hedging*. John Wiley&Sons, Hoboken, NJ, 2003.
- [18] M. Fehr and J. Hinz. A quantitative approach to carbon price risk modeling. Technical report, Institute for Operations Research, ETH Zurich, 2008.
- [19] E. Z. Ferencstein. Randomized stopping games and Markov market games. *Math. Methods Oper. Res.*, 66(3):531–544, 2007.
- [20] S. Hamadène and M. Jeanblanc. On the starting and stopping problem: application in reversible investments. *Math. Oper. Res.*, 32(1):182–192, 2007.
- [21] S. Hamadène, J.P.Lepeltier, and S.Peng. BSDE with continuous coefficients and application to Markovian nonzero sum stochastic differential games. In N. N. El-Karoui and L. Mazliak, editors, *Pitman Research Notes in Mathematics Series*, volume 364, pages 115–128. Longman, Harlow, 1997.
- [22] S. Hamadène and J. Lepeltier. Reflected BSDEs and mixed game problem. *Stochastic Process. Appl.*, 85(2):177–188, 2000.
- [23] S. Hamadène and J. Zhang. The continuous time nonzero-sum Dynkin game problem and application in game options. *SIAM Journal of Control and Optimization*, 48(5):3659–3669, 2010.
- [24] H. J. Kushner. Numerical approximations for nonzero-sum stochastic differential games. *SIAM J. Control Optim.*, 46(6):1942–1971, 2007.
- [25] H. J. Kushner and P. Dupuis. *Numerical methods for stochastic control problems in continuous time*, volume 24 of *Applications of Mathematics*. Springer-Verlag, New York, second edition, 2001.
- [26] R. Laraki, E. Solan, and N. Vieille. Continuous-time games of timing. *J. Econom. Theory*, 120(2):206–238, 2005.
- [27] F. Longstaff and E. Schwartz. Valuing American options by simulations: a simple least squares approach. *Rev. Finan. Studies*, 14:113–148, 2001.
- [28] M. Ludkovski. Financial hedging of operational flexibility. *Int. J. Theor. Appl. Finance*, 11 (8):799–839, 2008.
- [29] R. Myerson. Multistage games with communication. *Econometrica*, 54(2):323–358, 1986.
- [30] A. Nowak. On a new class of nonzero-sum discounted stochastic games having stationary Nash equilibrium points. *Intern. J. Game Theory*, 32 (1):121–132, 2003.
- [31] A. S. Nowak and T. E. S. Raghavan. Existence of stationary correlated equilibria with symmetric information for discounted stochastic games. *Math. Oper. Res.*, 17 (3):519–526, 1992.

- [32] Y. Ohtsubo. A nonzero-sum extension of Dynkin's stopping problem. *Math. Oper. Res.*, 12(2):277–296, 1987.
- [33] Y. Ohtsubo. On a discrete-time non-zero-sum Dynkin problem with monotonicity. *J. Appl. Probab.*, 28(2):466–472, 1991.
- [34] H. Pham and V. Ly Vath. Explicit solution to an optimal switching problem in the two-regime case. *SIAM J. Control Optim.*, 46(2):395–426, 2007.
- [35] D. M. Ramsey and K. Szajowski. Selection of a correlated equilibrium in Markov stopping games. *European J. Oper. Res.*, 184(1):185–206, 2008.
- [36] D. Rosenberg, E. Solan, and N. Vieille. Stopping games with randomized strategies. *Probab. Theory Related Fields*, 119(3):433–451, 2001.
- [37] E. Shmaya and E. Solan. Two-player nonzero-sum stopping games in discrete time. *Ann. Probab.*, 32(3B):2733–2764, 2004.
- [38] E. Solan and N. Vieille. Correlated equilibrium in stochastic games. *Games Econom. Behav.*, 38(2):362–399, 2002.
- [39] E. Solan and R. V. Vohra. Correlated equilibrium in quitting games. *Math. Oper. Res.*, 26 (3):601–610, 2001.
- [40] N. Touzi and N. Vieille. Continuous-time Dynkin games with mixed strategies. *SIAM J. Control Optim.*, 41(4):1073–1088, 2002.

## APPENDIX: NUMERICAL ALGORITHMS

(M. Ludkovski) DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY, UNIVERSITY OF CALIFORNIA, SANTA BARBARA, CA 93106-3110, USA

*E-mail address:* ludkovski@pstat.ucsb.edu, www.pstat.ucsb.edu/faculty/ludkovski

---

**Algorithm 1** Simulating one realized cashflow path  $\vartheta_i(s)$ ,  $0 \leq s \leq T$

---

**input:** Basis functions  $B_\ell(p, x)$ ,  $\ell = 1, \dots, r$ , regression coefficients  $\vec{\alpha}_i(t, \vec{\zeta})$ ; correlation law  $\Gamma$

**input:** Initial condition  $(p_0, x_0, \vec{u}(0))$ ; horizon  $T$

Initialize  $\vartheta_i(0) \leftarrow 0$  // Realized cashflows

**for**  $t = 0, \dots, T - 1$  **do**

**for** each  $\vec{\zeta} \in \{0, 1\}^2$  **do**

    // Evaluate the predicted continuation values from taking action  $\vec{\zeta}$

    Set  $\hat{q}_i(t, \vec{\zeta}) \leftarrow \sum_{\ell=1}^r \alpha_i^\ell(t, \vec{\zeta}) B_\ell(p_t, x_t) - K_{\{i, u_i(t), \zeta_i\}} + (a_i p_t - b_i x_t - c_i) \zeta_i$

**end for**

  Compute the stage- $t$  game values based on  $\hat{q}_i(t, \cdot)$ ,  $i = 1, 2$  and  $\Gamma$ , see (23)

  Obtain the correlated equilibrium strategy  $\vec{\mathbf{u}}(t)$ .

**if**  $\vec{\mathbf{u}}(t)$  is mixed **then**

    Perform randomization to obtain the realized action pair  $\vec{u}(t + 1)$

**else**

    Set  $\vec{u}(t + 1) \leftarrow \vec{\mathbf{u}}(t)$  //  $\vec{\mathbf{u}}(t)$  is pure

**end if**

  Update  $\vartheta_i(t + 1) \leftarrow \vartheta_i(t) - K_{\{i, u_i(t), u_i(t+1)\}} + (a_i p_t - b_i x_t - c_i) u_i(t + 1)$ ,  $i = 1, 2$

  Make an independent draw  $(p_{t+1}, x_{t+1}) \sim \mathbb{P}^{\vec{u}(t+1)}(\cdot | p_t, x_t)$

**end for**

---

---

**Algorithm 2** Computing Correlated Equilibrium Game Values
 

---

**input:**  $N > 0$  (number of paths);  $B_\ell(p, x)$ ,  $\ell = 1, \dots, r$  ( $r$  regression basis functions)

**input:** Correlation law  $\Gamma$ 

 Select anterior strategy profile  $\vec{u}^0$ 
**for** each regime  $\vec{\zeta} \in \{0, 1\}^2$  **do**

 Simulate  $N$  i.i.d. paths  $(p_t^n, x_t^{\vec{\zeta}, n})_{n=1}^N$  under  $\mathbb{P}^{\vec{u}^0}$  using Algorithm 1 and  $p_0^n = p_0$ ,  $x_0^{\vec{\zeta}, n} = x_0$ 
**end for**

 Initialize  $\vartheta_i^n(T, \vec{\zeta}) \leftarrow 0$ ,  $n = 1, \dots, N$ 
**for**  $t = (T - 1), \dots, 1, 0$  **do**
**for** each regime  $\vec{\zeta}$  **do**

 Evaluate  $B_\ell(p_t^n, x_t^{\vec{\zeta}, n})$  for  $\ell = 1, \dots, r$  and  $n = 1, \dots, N$ 

Regress

$$\vec{\alpha}_i(t, \vec{\zeta}) \leftarrow \arg \min_{\vec{\alpha} \in \mathbb{R}^r} \sum_{n=1}^N \left| \vartheta_i^n(t+1, \vec{\zeta}) - \sum_{\ell=1}^r \alpha^\ell B_\ell(p_t^n, x_t^{\vec{\zeta}, n}) \right|^2$$

**end for**
**for** each current regime  $\vec{u}$  **do**
**for** each  $\vec{\zeta} \in \{0, 1\}^2$ , and each  $n = 1, \dots, N$  **do**

 // Compute the predicted continuation value for each player from taking action  $\vec{\zeta}$ 

 Set  $\hat{q}_i^n(t, \vec{u}, \vec{\zeta}) \leftarrow \sum_{\ell=1}^r \alpha_i^\ell(t, \vec{\zeta}) B_\ell(p_t^n, x_t^{\vec{u}, n}) - K_{\{i, u_i, \zeta_i\}} + (a_i p_t^n - b_i x_t^{\vec{u}, n} - c_i) \zeta_i$ .

**end for**
**for** each path  $n = 1, \dots, N$  **do**

 Compute the stage- $t$  game values based on  $\hat{q}^n(t, \vec{u}, \cdot)$  and  $\Gamma$ , see (23)

 Obtain the equilibrium policy  $\vec{u}^{n,*}(t, \vec{u})$ 

 Recompute  $\vartheta_i^n(t, \vec{u})$  using  $\vec{u}^{n,*}(t, \vec{u})$  at stage- $t$  and Algorithm 1 for future stages

**end for**
**end for**
**end for**
**return**  $V_i(0, p_0, x_0, \vec{\zeta}) \simeq \frac{1}{N} \sum_{n=1}^N \vartheta_i^n(0, \vec{\zeta})$ 
**return** Regression coefficients  $\vec{\alpha}_i(t, \vec{\zeta})$  summarizing equilibrium strategies
 

---