# Distributionally Consistent Price Taking Equilibria in Stochastic Dynamic Games

Gürdal Arslan and Serdar Yüksel

***Abstract*— We consider a non-cooperative multi-stage game with discrete-time state dynamics. Players have their own decoupled state dynamics and each player wishes to minimize its own expected total cost. The salient aspect of our model is that each player's stage cost includes a payment (e.g., to a public utility) proportional to the magnitude of the player's decision. The coefficient multiplying each player's decision, called the price, is the same for all players and is determined as a function of the average of all player's decisions at that stage. Hence, each player's cost depends on the decisions of the other players only through the price. Here, we provide a stochastic and dynamic generalization of an equilibrium concept adopted in the economics literature, called the price-taking equilibrium, at which each player has no incentive to unilaterally deviate from its equilibrium strategy provided that the player ignores the effect of its own decisions on the price. In our setup, we allow for stochasticity in the price process and players observe only the past price realizations in addition to their own state realizations and their own past decisions. At a price-taking equilibrium, if players are given the distribution of the price process as if the price process is exogenous, they would have no incentive to unilaterally deviate from their equilibrium strategies. The main contribution of this paper is to establish such a stochastic and dynamic game generalization of price taking equilibria. We first derive the conditions for the existence of a price-taking equilibrium in the special case where the state dynamics are linear, the stage cost are quadratic, and the price function is linear. In this special case, our existence results are constructive for both finite-horizon and infinite horizon-problems. In the case where the number of players is taken to infinity, a price taking equilibrium exists which in turn is a mean-field equilibrium and is thus actually a Bayesian Nash equilibrium unlike the setup with a finite number of players. Finally, non-constructive existence results for price-taking equilibria and asymptotic equivalence with Nash equilibria are obtained for the case where the state and action sets are finite.**

## I. INTRODUCTION

In the economics literature, the concept of price-taking equilibria has been utilized [1] [2] for static or dynamic noise-free systems. Such an equilibria decouples competitive decision makers once they are given a sequence of *price variables* for different time stages and if the best responses give rise to the previously announced price, this would constitute a price-taking equilibria. An important contribution in this field due to Radner [1] which establishes the existence of equilibria for a class of static and dynamic setups.

In this paper, for stochastic dynamic games, we introduce a stochastic generalization of price-taking equilibrium. Unlike much of the existing literature (e.g. [1]), we do not impose the price process to be deterministic; the process realizations are causally given to the players, but its distribution is provided as common knowledge. Thus, at a price-taking equilibrium, unlike at a Nash equilibrium, players observe only the past price realizations in addition to their own state realizations and their own past decisions with the knowledge of the probability measure induced by the price process; but the optimal solutions in turn lead to a price process that is *distributionally consistent* with the price process given apriori. The main contribution of this paper is to establish such a stochastic and dynamic game generalization of price taking equilibria.

A motivating application area is the Autonomous Demand Response (ADR) problem on price-sensitive temperature regulation of multiple (possibly very large number of) units [3] [4]. In such a system, the real-time price of electricity $p_t$ depends on the system load [5] and there are two common approaches to determining $p_t$; ex-post pricing or ex-ante pricing [5]. In ex-post pricing, $p_t$ is determined at the end of the time stage $t$; whereas, in ex-ante pricing, $p_t$ is determined at the beginning of the time stage $t$. An example of ex-post pricing is $p_t = p\left(\sum_i(\ell_t^i + u_t^i) + L_t\right)$ where $L_t$ is the (random) total power drawn by all non-participating units during the period $t$ (kW) and $p$ is an increasing price function mapping the total system load to the unit price of energy. An example of ex-ante pricing would be $p_t = p\left(\sum_i(\ell_{t-1}^i + u_{t-1}^i) + L_{t-1}\right)$. In either case, since the real-time price of electricity depends on the decisions of all units, the units are facing coupled stochastic optimal control problems. Furthermore, in a realistic implementation each unit would have access only to its local information. In particular, a unit would not have access to any information about the other units, and may not even know the presence of the other units. In this context, recently [6] have generalized the findings in the economics literature in the context of energy systems where a linear quadratic Gaussian model is introduced and utilized to obtain optimal centralized and decentralized solutions under a class of technical assumptions requiring coordination among the decentralized players through a bidding process and assuming that the price sequence is deterministic.

### A. Stochastic Dynamic Price-taking Equilibrium

Consider a decentralized stochastic system with $N$ decision makers where the $i$-th decision maker is referred to as $\text{DM}^i$. Each $\text{DM}^i$ has its own scalar state, scalar control input, and random disturbance denoted by $x_t^i$, $u_t^i$, and $w_t^i$ at

G. Arslan is with the Department of Electrical Engineering, University of Hawaii at Manoa, 440 Holmes Hall, 2540 Dole Street, Honolulu, HI 96822, USA gurdal@hawaii.edu

S. Yüksel is with the Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, CANADA, K7L 3N6 yuksel@mast.queensu.ca

the discrete time step $t \geq 0$, respectively. Each DM$^i$'s state evolves as

$$x_{t+1}^i = f_t^i(x_t^i, u_t^i, w_t^i), \qquad t \geq 0$$

starting with some (possibly random) initial state $x_0^i$. The scalar price variable $p_t$ at time $t \geq 0$ is produced by

$$p_t = \kappa_t(u_t^1, \dots, u_t^N, \xi_t) \tag{1}$$

where $\kappa_t$ is some scalar-valued function and $\xi_t$ is some random variable, $t \geq 0$.

We assume that the random variables $x_0^i$, $w_t^i$, $\xi_t$, for all $i$, $t \geq 0$, are mutually independent.

Each DM$^i$ uses a strategy $\eta^i = (\eta_0^i, \eta_1^i, \dots)$, where $\eta_t^i$ is a mapping from DM$^i$'s information set to a control input at time $t \geq 0$. More precisely,

$$u_t^i = \eta_t^i(I_t^i, P_{t-1}), \qquad t \geq 0$$

where, $I_0^i := x_0^i$, $P_{-1} := \emptyset$, and for $t \geq 1$,

$$I_t^i := (x_0^i, \dots, x_t^i, u_0^i, \dots, u_{t-1}^i), \quad P_{t-1} := (p_0^i, \dots, p_{t-1}^i).$$

The strategies of all DMs other than DM$^i$ is denoted by $\eta^{-i}$.

Each DM$^i$ wishes to minimize its finite-horizon or infinite-horizon cost

$$J^i(\eta^i, \eta^{-i}) = E\left( \sum_{t=0}^{T-1} (\beta^i)^t \Big( p_t u_t^i + g_t^i(x_t^i, u_t^i) \Big) + g_T^i(x_T^i) \right)$$

where $T \geq 1$ is the length of the horizon, $\beta^i \in (0,1)$ is DM$^i$'s discount factor, and $g_t^i$ is some scalar-valued function which determines DM$^i$'s cost at each time $t \geq 0$. Note that each DM$^i$'s cost is influenced by the other DMs only through the price sequence.

Suppose now that the price sequence is an **exogenous** (that is, primitive, generated by nature) random sequence $(z_0, \dots, z_{T-1})$ with a given probability distribution $\zeta$ instead of the endogenous sequence $(p_0, \dots, p_{T-1})$ generated by (1). By a slight abuse of notation, let

$$J^i(\eta^i; \zeta) := E\left( \sum_{t=0}^{T-1} (\beta^i)^t \Big( z_t u_t^i + g_t^i(x_t^i, u_t^i) \Big) + g_T^i(x_T^i) \right)$$

where

$$u_t^i = \eta_t^i(I_t^i, Z_{t-1}), \qquad t \geq 0$$

with $Z_{-1} := \emptyset$, and for $t \geq 1$, $Z_{t-1}^i := (z_0^i, \dots, z_{t-1}^i)$.

In this case, no DM has any influence on the price sequence and the cost $J^i(\eta^i; \zeta)$ is independent of the strategies of all DMs other than DM$^i$.

*Definition 1:* Let $\eta = (\eta^1, \dots, \eta^N)$ be a joint strategy and $\zeta_\eta$ be the probability distribution of the price sequence generated by $\eta$ endogenously through (1). The joint strategy $\eta$ is called a *distributionally consistent price-taking equilibrium* if

$$J^i(\eta^i; \zeta_\eta) \leq J^i(\tilde{\eta}^i; \zeta_\eta), \quad \text{for all } i, \tilde{\eta}^i \tag{2}$$

where the price sequence is (incorrectly) assumed to be an exogenous random sequence with the probability distribution $\zeta_\eta$.

Note that, in a price-taking equilibrium, each DM$^i$ ignores the influence of its own strategy $\eta^i$ on the price sequence in minimizing its long-term cost. In contrast, the well-known concept of *Nash equilibrium* requires a joint strategy $\eta = (\eta^1, \dots, \eta^N)$ to satisfy

$$J^i(\eta^i, \eta^{-i}) \leq J^i(\tilde{\eta}^i, \eta^{-i}), \quad \text{for all } i, \tilde{\eta}^i.$$

where each DM$^i$ takes into account the entire influence of its own strategy $\eta^i$ on its long-term cost including through the price sequence.

## II. LINEAR QUADRATIC PROBLEMS WITH FINITELY MANY PLAYERS

### A. Finite Horizon Case

Here, each DM$^i$ wishes to minimize the finite-horizon quadratic-cost

$$E\left( \sum_{t=0}^{T-1} (\beta^i)^t \Big( p_t u_t^i + r_t^i (u_t^i)^2 + q_t^i (x_t^i - \bar{x}_t^i)^2 \Big) + (\beta^i)^T q_T^i (x_T^i - \bar{x}_T^i)^2 \right)$$

subject to linear price and state dynamics

$$\begin{aligned} p_t &= c_t^0 + \sum_j c_t^j u_t^j + \xi_t \\ x_{t+1}^i &= a_t^i x_t^i + b_t^i u_t^i + w_t^i. \end{aligned}$$

where $r_t^i > 0$, $q_t^i \geq 0$, $\bar{x}_t^i$, $c_t^0$, $c_t^i$, $a_t^i$, $b_t^i$ are given scalars, for all $i$, $t \geq 0$.

Suppose that DM$^i$ instead aims to minimize $J^i(\eta^i; \zeta)$ over $\eta^i$ for a given exogenous price sequence $(z_0, \dots, z_{T-1})$ with probability distribution $\zeta$. Let $\hat{z}_{t|t-1} := E(z_t | Z_{t-1})$, for $t \geq 0$. DM$^i$'s cost-to-go functions

$$\begin{aligned} V_k^i(I_k^i, Z_{k-1}) := \min_{\eta_k^i, \dots, \eta_{T-1}^i} E\Bigg( \sum_{t=k}^{T-1} (\beta^i)^{t-k} \Big( \hat{z}_{t|t-1} u_t^i \\ + r_t^i (u_t^i)^2 + q_t^i (x_t^i - \bar{x}_t^i)^2 \Big) \\ + (\beta^i)^{T-k} q_T^i (x_T^i - \bar{x}_T^i)^2 \Big| I_k^i, Z_{k-1} \Bigg) \end{aligned}$$

satisfy, for $k = 0, \dots, T-1$,

$$\begin{aligned} V_k^i(I_k^i, Z_{k-1}) := \min_{u_k^i} \big\{ \hat{z}_{k|k-1} u_k^i + r_k^i (u_k^i)^2 + q_k^i (x_k^i - \bar{x}_k^i)^2 \\ + \beta^i E\big( V_{k+1}^i (I_{k+1}^i, Z_k) \big| I_k^i, Z_{k-1} \big) \big\} \end{aligned} \tag{3}$$

with the boundary condition $V_T^i(I_T^i, Z_{T-1}) := q_T^i (x_T^i - \bar{x}_T^i)^2$.

As an induction hypothesis, assume that $V_{k+1}^i$ has the following form, for $k = 0, \dots, T-1$,

$$V_{k+1}^i(I_{k+1}^i, Z_k) = Q_{k+1}^i (x_{k+1}^i)^2$$

$$+\left(C_{k+1}^i + \sum_{t=k+1}^{T} S_{k+1|t}^i \hat{z}_{t|k}\right) x_{k+1}^i + K_{k+1}^i(Z_k)$$

where $Q_{k+1}^i \geq 0$, $C_{k+1}^i$, $S_{k+1|t}^i$, $t = k+1, \ldots, T$, are constants and $K_{k+1}^i$ is some function of $Z_k$ ($\hat{z}_{T|k} := 0$ is introduced for notational consistency). $V_T^i$ satisfies this hypothesis with

$$Q_T^i = q_T^i, \quad C_T^i = -2q_T^i \bar{x}_T^i \quad S_{T|T}^i = 0, \quad K_T^i = q_T^i \left(\bar{x}_T^i\right)^2.$$

The unique minimizing control in (3) is obtained as

$$u_k^i = f_k^i x_k^i + \sum_{t=k}^{T} n_{k|t}^i \hat{z}_{t|k-1} + o_k^i \tag{4}$$

where

$$\begin{aligned}
f_k^i &= \frac{-\beta^i Q_{k+1}^i b_k^i a_k^i}{r_k^i + \beta^i Q_{k+1}^i (b_k^i)^2} \\
n_{k|k}^i &= \frac{-1/2}{r_k^i + \beta^i Q_{k+1}^i (b_k^i)^2} \\
n_{k|t}^i &= \frac{-\beta^i b_k^i S_{k+1|t}^i/2}{r_k^i + \beta^i Q_{k+1}^i (b_k^i)^2}, \quad t = k+1, \ldots, T \\
o_k^i &= \frac{-\beta^i b_k^i (Q_{k+1}^i \hat{w}_k^i + \frac{1}{2} C_{k+1}^i)}{r_k^i + \beta^i Q_{k+1}^i (b_k^i)^2}.
\end{aligned}$$

This leads to

$$\begin{aligned}
V_k^i(I_k^i, Z_{k-1}) =& Q_k^i (x_k^i)^2 + \left(C_k^i + \sum_{t=k}^{T} S_{k|t}^i \hat{z}_{t|k-1}\right) x_k^i \\
&+ K_k^i(Z_{k-1})
\end{aligned}$$

where

$$\begin{aligned}
Q_k^i =& r_k^i (f_k^i)^2 + q_k^i + \beta^i Q_{k+1}^i (a_k^i + b_k^i f_k^i)^2 \\
C_k^i =& 2r_k^i f_k^i o_k^i - 2q_k^i \bar{x}_k^i \\
&+ 2\beta^i Q_{k+1}^i (a_k^i + b_k^i f_k^i)(b_k^i o_k^i + \hat{w}_k^i) \\
&+ \beta^i C_{k+1}^i (a_k^i + b_k^i f_k^i) \\
S_{k|k}^i =& f_k^i + 2r_k^i f_k^i n_{k|k}^i + 2\beta^i Q_{k+1}^i (a_k^i + b_k^i f_k^i) b_k^i n_{k|k}^i \\
S_{k|t}^i =& 2r_k^i f_k^i n_{k|t}^i + 2\beta^i Q_{k+1}^i (a_k^i + b_k^i f_k^i) b_k^i n_{k|t}^i \\
&+ \beta^i S_{k+1|t}^i (a_k^i + b_k^i f_k^i), \qquad t = k+1, \ldots, T \\
K_k^i(Z_{k-1}) =& \hat{z}_{k|k-1} \left(\sum_{t=k}^{T} n_{k|t}^i \hat{z}_{t|k-1} + o_k^i\right) \\
&+ q_k^i \left(\bar{x}_k^i\right)^2 + \beta^i Q_{k+1}^i \times \\
&E\left(b_k^i \sum_{t=k}^{T} n_{k|t}^i \hat{z}_{t|k-1} + b_k^i o_k^i + w_k^i\right)^2 \\
&+ \beta^i \left(C_{k+1}^i + \sum_{t=k+1}^{T} S_{k+1|t}^i \hat{z}_{t|k-1}\right) \times \\
&\left(b_k^i \sum_{t=k}^{T} n_{k|t}^i \hat{z}_{t|k-1} + b_k^i o_k^i + \hat{w}_k^i\right) \\
&+ \beta^i E\big(K_{k+1}^i(Z_k) | Z_{k-1}\big).
\end{aligned}$$

Since $V_k^i(I_k^i, Z_{k-1})$ satisfies the induction hypothesis, DM$^i$'s unique optimal strategy with respect to any given exogenous price sequence $(z_0, \ldots, z_{T-1})$ with probability distribution $\zeta$ is given by (4). Furthermore, if each DM$^i$ uses its optimal strategy (4), then the following price sequence is generated: for $t = 0, \ldots, T-1$,

$$p_t = c_t^0 + \sum_i c_t^i f_t^i x_t^i + \sum_i c_t^i \sum_{k=t}^{T} n_{t|k}^i \hat{z}_{k|t-1} + \sum_i c_t^i o_t^i + \xi_t \tag{5}$$

where

$$x_{t+1}^i = (a_t^i + b_t^i f_t^i) x_t^i + b_t^i \sum_{k=t}^{T} n_{t|k}^i \hat{z}_{k|t-1} + b_t^i o_t^i + w_t^i.$$

Suppose that a joint strategy $\eta = (\eta^1, \ldots, \eta^N)$ is a distributionally consistent price-taking equilibrium, and let $\zeta_\eta$ be the probability distribution of the price sequence generated by $\eta$. Each $\eta^i$ must be optimal with respect to $\zeta_\eta$ in the sense of (2). Therefore, each $\eta^i$ would be of the form (4) and the resulting price sequence $(p_0, \ldots, p_{T-1})$ would be of the form (5) where the conditional expectations $\hat{z}_{k|t-1}$ on the right-hand-sides of (4) and (5) would be replaced with $\hat{p}_{k|t-1} := E(p_k | P_{t-1})$, $k \geq t \geq 0$. Since the resulting price sequence has distribution $\zeta_\eta$, the unconditional expectations $\hat{p}_{0|-1}, \ldots, \hat{p}_{T-1|-1}$ satisfy the linear equations

$$\begin{aligned}
\hat{p}_{t|-1} =& c_t^0 + \sum_i c_t^i f_t^i \hat{x}_{t|-1}^i + \sum_i c_t^i \sum_{k=t}^{T} n_{t|k}^i \hat{p}_{k|-1} \\
&+ \sum_i c_t^i o_t^i + \hat{\xi}_t
\end{aligned} \tag{6}$$

$$\hat{x}_{t+1|-1}^i = (a_t^i + b_t^i f_t^i) \hat{x}_{t|-1}^i + b_t^i \sum_{k=t}^{T} n_{t|k}^i \hat{p}_{k|-1} + b_t^i o_t^i + \hat{w}_t^i \tag{7}$$

for $t = 0, \ldots, T-1$, where $\hat{x}_{0|-1} = \hat{x}_0$. Continuing recursively for $\ell = 1, \ldots, T-1$, it is straightforward to show that the conditional expectations $\hat{p}_{t|\ell-1}$ satisfy (6)-(7) for $t = \ell, \ldots, T-1$, i.e.,

$$\begin{aligned}
\hat{p}_{t|\ell-1} =& c_t^0 + \sum_i c_t^i f_t^i \hat{x}_{t|\ell-1}^i + \sum_i c_t^i \sum_{k=t}^{T} n_{t|k}^i \hat{p}_{k|\ell-1} \\
&+ \sum_i c_t^i o_t^i + \hat{\xi}_t
\end{aligned} \tag{8}$$

$$\begin{aligned}
\hat{x}_{t+1|\ell-1}^i =& (a_t^i + b_t^i f_t^i) \hat{x}_{t|\ell-1}^i + b_t^i \sum_{k=t}^{T} n_{t|k}^i \hat{p}_{k|\ell-1} \\
&+ b_t^i o_t^i + \hat{w}_t^i
\end{aligned} \tag{9}$$

for $t = \ell, \ldots, T-1$, where $\hat{x}_{\ell|\ell-1}^i$ is obtained from, $t = 0, \ldots, \ell-1$,

$$\begin{aligned}
\hat{x}_{t+1|\ell-1}^i =& (a_t^i + b_t^i f_t^i) x_{t|\ell-1}^i + b_t^i \sum_{k=t}^{T} n_{t|k}^i \hat{p}_{k|t-1} \\
&+ b_t^i o_t^i + E(w_t^i | P_{\ell-1})
\end{aligned}$$

$$
\begin{aligned}
p_t =& c_t^0 + \sum_i c_t^i f_t^i \hat{x}_{t|\ell-1}^i + \sum_i c_t^i \sum_{k=t}^{T} n_{t|k}^i \hat{p}_{k|t-1} \\
&+ \sum_i c_t^i o_t^i + E(\xi_t | P_{\ell-1})
\end{aligned}
$$

with $\hat{x}_{0|\ell-1}^i = E(x_0^i | P_{\ell-1})$.

*Theorem 1:* Consider the finite-horizon linear quadratic problem of this section.

1) A distributionally consistent price-taking equilibrium exists if and only if the linear equations (8)-(9) can be recursively solved for $\ell = 0, \dots, T-1$.
2) Any solution $\hat{p}_{t|\ell-1}$, $0 \le \ell \le t \le T-1$, of (8)-(9) corresponds to a distributionally consistent price-taking equilibrium $\eta$ where each $\eta^i$ is defined by

$$
u_k^i = f_k^i x_k^i + \sum_{t=k}^{T} n_{k|t}^i \hat{p}_{t|\ell-1} + o_k^i \tag{10}
$$

for $k = 0, \dots, T-1$.

*Proof:* Omitted due to space constraints. ■

Theorem 1 establishes that the existence of a distributionally consistent price-taking equilibrium in the finite-horizon linear quadratic case is equivalent to the solvability of the equations (8)-(9). Furthermore, Theorem 1 constructs a distributionally consistent price-taking equilibrium corresponding to any solution of (8)-(9). In the next section, we extend this result to infinite-horizon linear quadratic problems.

### *B. Infinite Horizon Case*

In this section, we consider a time-invariant system and let the time horizon $T$ approach $\infty$. To develop a stationary solution, we also assume that the information available at time $t$ with regard to the price process contains data from the infinite past, that is, each decision maker at time $t$ has access to $P_{t-1} = (\dots, p_{-2}, p_{-1}, p_0, \dots, p_{t-1})$ with regard to the price process. The price process before the initial time $t = 0$, which is $P_{-1}$, is generated by some exogenous distribution; whereas, the prices $(p_0, p_1, \dots)$ are generated endogenously as described below. Here, DM$^i$ wishes to minimize

$$
E \sum_{t \ge 0} (\beta^i)^t \Big( p_t u_t^i + r^i (u_t^i)^2 + q^i (x_t^i - \bar{x}^i)^2 \Big)
$$

subject to, for $t \ge 0$,

$$
\begin{aligned}
p_t &= c^0 + \sum_j c^j u_t^j + \xi_t \\
x_{t+1}^i &= a^i x_t^i + b^i u_t^i + w_t^i.
\end{aligned}
$$

where $r^i > 0$, $q^i > 0$, $\bar{x}^i$, $c^0$, $c^i$, $a^i$, $b^i \neq 0$ are given scalars, for all $i$. Moreover, we assume that

(i) Each sequence $\{\xi_t\}_{t \ge 0}, \{w_t^1\}_{t \ge 0}, \dots, \{w_t^N\}_{t \ge 0}$ is independent and identically distributed with finite second moments
(ii) $\{x_0^1, \dots, x_0^N, E(x_0^1 | P_{-1})), \dots, E(x_0^N | P_{-1})\}$, $\{\xi_t\}_{t \ge 0}$, $\{w_t^1\}_{t \ge 0}, \dots, \{w_t^N\}_{t \ge 0}$ are independent.

Consider an exogenous random price sequence $(\dots, z_{-1}, z_0, z_1, \dots)$ with probability distribution $\zeta$ such that the sequence of conditional expectations

$$
\hat{z}_{t|t-1} := E(z_t | Z_{t-1}), \qquad t \ge 0
$$

is stationary where $Z_{t-1} := (\dots, z_{-1}, z_0, \dots, z_{t-1})$. Suppose that DM$^i$ instead aims to minimize

$$
J^i(\eta^i, \zeta) := E \sum_{t \ge 0} (\beta^i)^t \Big( \hat{z}_{t|t-1} u_t^i + r^i (u_t^i)^2 + q^i (x_t^i - \bar{x}^i)^2 \Big)
$$

where

$$
x_{t+1}^i = a^i x_t^i + b^i u_t^i + w_t^i.
$$

This is a standard linear quadratic optimal control problem with an additional linear control cost. It can be shown that the value function

$$
\begin{aligned}
V^i(x_0^i) :=& \min_{\eta_0^i, \eta_1^i, \dots} \\
& E\left( \sum_{t \ge 0} (\beta^i)^t \Big( \hat{z}_{t|t-1} u_t^i + r^i (u_t^i)^2 + q^i (x_t^i - \bar{x}^i)^2 \Big) \Big| x_0^i \right)
\end{aligned}
$$

satisfies the Bellman equation. It is straightforward to show that the value function has the quadratic form

$$
V^i(x^i) = Q_i (x^i)^2 + (C^i + S^i E(z_0)) x^i + K^i
$$

where $Q^i \ge 0$, $C^i$, $S^i$, and $K^i$ are constants. The minimizing control is given by

$$
u^i = f^i x^i + m^i \hat{z}_{0|-1} + n^i \hat{z}_0 + o^i
$$

where

$$
\begin{aligned}
f^i &= -\frac{\beta^i Q^i b^i a^i}{r^i + \beta^i Q^i (b^i)^2}, \quad m^i = -\frac{1/2}{r^i + \beta^i Q^i (b^i)^2} \\
n^i &= -\frac{\beta^i S^i b^i / 2}{r^i + \beta^i Q^i (b^i)^2}, \quad o^i = -\frac{\beta^i b^i (Q^i \hat{w}_0^i + C^i/2)}{r^i + \beta^i Q^i (b^i)^2}
\end{aligned}
$$

and $Q^i$, $C^i$, $S^i$ satisfy

$$
\begin{aligned}
Q^i =& r^i (f^i)^2 + q^i + \beta^i Q^i (a^i + b^i f^i)^2 \\
C^i =& 2 r^i f^i o^i - 2 q^i \bar{x}^i + 2 \beta^i Q^i (a^i + b^i f^i)(b^i o^i + \hat{w}_0^i) \\
&+ \beta^i C^i (a^i + b^i f^i) \\
S^i =& f^i + 2 r^i f^i (m_i + n^i) + 2 \beta^i Q^i (a^i + b^i f^i) b^i (m^i + n^i) \\
&+ \beta^i S^i (a^i + b^i f^i).
\end{aligned}
$$

Therefore, DM$^i$'s optimal strategy $\eta^i$ with respect to $\zeta$ is given by

$$
u_t^i = \eta_t^i (I_t^i) = f^i x_t^i + m^i \hat{z}_{t|t-1} + n^i \hat{z}_0 + o^i.
$$

Let $\eta := (\eta^1, \dots, \eta^N)$. The price sequence generated by the joint strategy $\eta$ is

$$
p_t = c^0 + \sum_i c^i (f^i x_t^i + m^i \hat{z}_{t|t-1} + n^i \hat{z}_0 + o^i) + \xi_t \tag{11}
$$

where

$$
x_{t+1}^i = (a^i + b^i f^i) x_t^i + b^i (m^i \hat{z}_{t|t-1} + n^i \hat{z}_0 + o^i) + w_t^i.
$$

If the conditional expectations $\hat{p}_{t|t-1} = E(p_t|P_{t-1})$, $t \geq 0$, of the price sequence (11) is stationary and satisfies

$$\hat{z}_{t|t-1} = \hat{p}_{t|t-1} \tag{12}$$

then the joint strategy $\eta$ would be a price-taking equilibrium.

We now focus on the special case where $\{\xi_t\}_{t\geq 0}$, $\{w_t^i\}_{t\geq 0}$, $x_0^i$, $E(x_0^i|P_{-1})$, for all $i$, are Gaussian. We rewrite (11) in the vectorized form

$$p_t = c + H^T x_t + m^p \hat{z}_{t|t-1} + n^p \hat{z}_0 + \xi_t \tag{13}$$
$$x_{t+1} = F x_t + M^x \hat{z}_{t|t-1} + N^x \hat{z}_0 + O^x + w_t \tag{14}$$

where $x_t := (x_t^1, \dots, x_t^N)^T$, $w_t := (w_t^1, \dots, w_t^N)^T$, and $c$, $H$, $m_p$, $n_p$, $F$, $M^x$, $N^x$, $O^x$ are appropriate dimensional constants. The conditional means of (13)-(14) satisfy

$$\hat{p}_{t|t-1} = c + H^T \hat{x}_{t|t-1} + m^p \hat{z}_{t|t-1} + n^p \hat{z}_0 + \hat{\xi}_0$$
$$\hat{x}_{t+1|t} = F \hat{x}_{t|t} + M^x \hat{z}_{t|t-1} + N^x \hat{z}_0 + O^x + \hat{w}_0$$

where $\hat{x}_{t|t-1} = E(x_t|P_{t-1})$, $\hat{x}_{t|t} = E(x_t|P_t)$, for $t \geq 0$. After some analysis, we obtain

$$\begin{aligned}\hat{x}_{t+1|t} =& F\hat{x}_{t|t-1} + F\Sigma_{t|t-1}H\frac{p_t - \hat{p}_{t|t-1}}{H^T\Sigma_{t|t-1}H + \Psi}\\ &+ M^x \hat{z}_{t|t-1} + N^x \hat{z}_0 + O^x + \hat{w}_0\\ \Sigma_{t+1|t} =& F\left(\Sigma_{t|t-1} - \frac{\Sigma_{t|t-1}HH^T\Sigma_{t|t-1}}{H^T\Sigma_{t|t-1}H + \Psi}\right)F^T + \mathcal{W}\end{aligned}$$

where $\mathcal{W} = \mathrm{cov}(w_0, w_0)$.

To ensure the stationarity of $\hat{p}_{t|t-1}$, we initialize $\Sigma_{0|-1}$ as

$$\Sigma_{0|-1} = \bar{\Sigma} = F(\bar{\Sigma} - \bar{\Sigma}H(H^T\bar{\Sigma}H + \Psi)^{-1}H^T\bar{\Sigma})F^T + \mathcal{W} \tag{15}$$

which yields $\Sigma_{t|t-1} = \bar{\Sigma}$, for all $t \geq 0$. The steady-state Riccati equation in (15) has always a nonnegative definite solution $\bar{\Sigma}$ due to the stability of $F$ which follow from the controllability of $(a^i, b^i)$ and the observability of $(a^i, \sqrt{q^i})$, for all $i$. Let $G := \frac{F\bar{\Sigma}H}{H^T\bar{\Sigma}H+\Psi}$. We then rewrite the overall dynamics (generating $\hat{p}_{t|t-1}$) as

$$\hat{p}_{t|t-1} = c + H^T \hat{x}_{t|t-1} + m^p \hat{z}_{t|t-1} + n^p \hat{z}_0 + \hat{\xi}_0 \tag{16}$$

where

$$\begin{aligned}x_{t+1} - \hat{x}_{t+1|t} =& (F - GH^T)(x_t - \hat{x}_{t|t-1})\\ &-G(\xi_t - \hat{\xi}_0) + w_t - \hat{w}_0 \quad (17)\\ x_{t+1} =& F x_t + M^x \hat{z}_{t|t-1} + N^x \hat{z}_0 + O^x + w_t. \quad (18)\end{aligned}$$

The time-invariant dynamics (17)-(18) are stable[1] and driven by stationary processes. To obtain stationary behavior, we attempt to initialize (17)-(18) at steady state as well. First, we initialize the mean dynamics as

$$\hat{x}_0 = \bar{x} := (I - F)^{-1}((M^x + N^x)\hat{z}_0 + O^x + \hat{w}_0) \tag{19}$$

which yields $\hat{x}_t = E(\hat{x}_{t|t-1}) = \bar{x}$, for all $t \geq 0$. Note that enforcing $\hat{z}_0 = \hat{p}_0$ in (16) and (19) leads to

$$\bar{x} = \left(I - F - \frac{(M^x + N^x)H^T}{1 - m^p - n^p}\right)^{-1} \times \left((M^x + N^x)\frac{c + \hat{\xi}_0}{1 - m^p - n^p} + O^x + \hat{w}_0\right) \tag{20}$$

assuming that $1 - m^p - n^p \neq 0$ and

$$\left(I - \frac{(I-F)^{-1}(M^x + N^x)H^T}{1 - m^p - n^p}\right)^{-1} \text{ exists.}$$

We wish to initialize to initialize the covariance dynamics at steady-state. Suppose that the following holds.

$$\Phi_0 = \bar{\Phi} = F\bar{\Phi}F^T + \frac{M^x H^T(\bar{\Phi} - \bar{\Sigma})H(M^x)^T}{(1 - m^p)^2} + \mathcal{W} \tag{21}$$

where $\bar{\Phi} - \bar{\Sigma}$ is nonnegative definite. Then, for all $t \geq 0$,

$$\begin{pmatrix} x_t - \hat{x}_{t|t-1} \\ x_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \bar{x} \end{pmatrix}, \begin{pmatrix} \bar{\Sigma} & \bar{\Sigma} \\ \bar{\Sigma} & \bar{\Phi} \end{pmatrix}\right) \tag{22}$$

provided (15), (20), (21) hold. This results in the stationarity of $\hat{p}_{t|t-1}$, $t \geq 0$, due to the stability of the time-invariant dynamics (17)-(18) driven by stationary processes.

Finally, enforcing (12) leads to the price-taking equilibrium strategy, for all $i$,

$$u_t^i = \eta_t^i(I_t^i) = f^i x_t^i + m^i \hat{p}_{t|t-1} + n^i \hat{p}_0 + o^i \tag{23}$$

where

$$\hat{p}_{t|t-1} = \frac{c + H^T \hat{x}_{t|t-1} + n^p \hat{p}_0 + \hat{\xi}_0}{1 - m^p} \tag{24}$$

$$\begin{aligned}\hat{x}_{t+1|t} =& F\hat{x}_{t|t-1} + G(p_t - \hat{p}_{t|t-1})\\ &+ M^x \hat{p}_{t|t-1} + N^x \hat{p}_0 + O^x + \hat{w}_0 \quad (25)\end{aligned}$$

and

$$\hat{p}_0 = \frac{c + H^T \bar{x} + \hat{\xi}_0}{1 - m^p - n^p}. \tag{26}$$

*Theorem 2:* Consider the infinite-horizon linear quadratic problem of this section. Assume that the dynamics are initialized at steady-state, i.e., (22) holds at $t = 0$. Then, the joint strategy $\eta$ defined by (23)-(26) is a distributionally consistent price-taking equilibrium.

## III. LINEAR QUADRATIC PROBLEMS WITH INFINITELY MANY IDENTICAL DMs

In this section, we consider the case of infinitely many identical DMs ($N \uparrow \infty$) where the price at each time $t \geq 0$ is determined by the average of the decisions as

$$p_t = c_t^0 + \limsup_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} u_t^j + \xi_t.$$

We will assume that $(x_0^1, x_0^2, \dots)$ is independent and identically distributed (iid). We will only consider the infinite horizon case and we further restrict our attention to a time-invariant system as in subsection II-B. In parallel to the development in subsection II-B, suppose that each DM$^i$ instead aims to minimize its long-term cost with respect to an exogenous random price sequence $(z_0, z_1, \dots)$ which is iid. Each DM$^i$'s optimal strategy $\eta^i$ with respect to this price sequence is obtained as

$$u_t^i = \eta_t^i(I_t^i) = f x_t^i + (m + n)\hat{z}_0 + o$$

[1] The stability of $F$ implies the stability of $F - GH^T$.

where $f$, $m$, $n$, $o$ are as in subsection II-B except the superscript $i$ is dropped throughout. The joint strategy $\eta = (\eta^1, \eta^2, \dots)$ generates the price sequence

$$p_t = c^0 + f \limsup_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_t^i + (m+n)\hat{z}_0 + o + \xi_t.$$

If $(p_0, p_1, \dots)$ is an iid sequence and $\hat{z}_0 = \hat{p}_0$, then the joint strategy $\eta$ would be a price-taking equilibrium. Note that, under $\eta$, each DM$^i$'s state evolves as

$$x_{t+1}^i = (a+bf)x_t^i + b(m+n)\hat{z}_0 + bo + w_t^i.$$

Since $(x_0^1, x_0^2, \dots)$ is iid, $(x_t^1, x_t^2, \dots)$ is iid, for any $t \geq 0$. Thus, $\limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_t^i = \hat{x}_t := E(x_t^1)$, for any $t \geq 0$, and

$$p_t = c^0 + f\hat{x}_t + (m+n)\hat{z}_0 + o + \xi_t. \tag{27}$$

If $\hat{x}_t$ is initially at steady-state, i.e.,

$$\hat{x}_0 = \bar{x} = \frac{b(m+n)\hat{z}_0 + bo + \hat{w}_0}{1-a-bf} \tag{28}$$

then $\hat{x}_t = \bar{x}$, for all $t \geq 0$, which would make $(p_0, p_1, \dots)$ an iid sequence. Enforcing $\hat{z}_0 = \hat{p}_0$ in (27), (28) yields

$$\bar{x} = \frac{b(m+n)\frac{c^0+o+\hat{\xi}_0}{1-m-n} + bo + \hat{w}_0}{1-a-\frac{bf}{1-m-n}}$$

assuming that

$$1-m-n \neq 0 \quad \text{and} \quad 1-a-\frac{bf}{1-m-n} \neq 0.$$

This leads to the price-taking equilibrium strategy, for all $i$,

$$u_t^i = \eta_t^i(I_t^i) = fx_t^i + (m+n)\hat{p}_0 + o \tag{29}$$

where

$$\hat{p}_0 = \frac{c^0 + f\bar{x} + o + \hat{\xi}_0}{1-m-n}. \tag{30}$$

*Theorem 3:* Consider the infinite-horizon linear quadratic problem of this section with infinitely many identical DMs. Assume that the dynamics are initialized at steady-state, i.e., $\hat{x}_0 = \bar{x}$. Then, the joint strategy $\eta$ defined by (29)-(30) is a distributionally consistent price-taking equilibrium.

A remark is in order at this point. In the case where the price is determined by the average decisions of infinitely many identical DMs, no individual DM can change the price sequence by unilaterally deviating to an alternative strategy; therefore, a price-taking equilibrium is also a Nash equilibrium, which in fact is a mean-field equilibrium [7] [8] [9].

## IV. FINITE-STATE ACTION SYSTEMS

### A. Finite Number of Players

The results presented for the linear case apply to finite state and action problems, which indeed include many applications including energy systems. Consider the following setup

$$x_{t+1}^i = f(x_t^i, u_t^i, w_t^i)$$

where $x^i$ is $\mathbb{X}$-valued, $u_t^i$ is $\mathbb{U}$-valued with $\mathbb{X}$ and $\mathbb{U}$ are finite sets, each viewed as a subset of $\mathbb{R}$. Furthermore, the price variable $p_t$ is $\mathbb{P}$-valued with dynamics given by $p_t^i = Q(\frac{1}{N}\sum_{i=1}^{N} u_t^i)$, where $Q : \mathbb{R} \to \mathbb{P}$ is a quantization operation with $\mathbb{P}$ being a finite set. The cost per stage is, as before, $g^i(x_t^i, u_t^i, p_t) = (x_t^i)^2 + (u_t^i)^2 + p_t u_t^i$. The goal of each DM is to minimize $E[\sum_{k=0}^{T-1} g^i(x_k^i, u_k^i, p_k)]$. For such a problem, the following can be established.

*Theorem 4:* A distributionally consistent price-taking equilibrium exists.

### B. Infinite Number of Players

Here, we consider a discounted cost setup with infinite horizons; i.e., the goal of each DM is to minimize $E[\sum_{k=0}^{\infty} \beta^k g^i(x_t^i, u_t^i, p_t)]$ for some $\beta \in (0,1)$. The following holds.

*Theorem 5:* There exists a Nash equilibrium which coincides with a distributionally consistent price-taking equilibrium.

## V. CONCLUSION

We presented a stochastic and dynamic generalization of an equilibrium concept, called the distributionally consistent price-taking equilibrium, at which each player has no incentive to unilaterally deviate from its equilibrium strategy provided that the player ignores the effect of its own decisions on the price variables.

## REFERENCES


[1] R. Radner, "Competitive equilibrium under uncertainty," *Econometrica: Journal of the Econometric Society*, pp. 31–58, 1968.

[2] E. J. Green, "Noncooperative price taking in large dynamic markets," *Journal of Economic Theory*, vol. 22, no. 2, pp. 155–182, 1980.

[3] P. Constantopoulos, F. C. Schweppe, and R. C. Larson, "Estia: A real-time consumer control scheme for space conditioning usage under spot electricity pricing," *Computers & Operations Research*, vol. 18, no. 8, pp. 751–765, 1991.

[4] M. Roozbehani, M. A. Dahleh, and S. K. Mitter, "Volatility of power grids under real-time pricing," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1926–1940, 2012.

[5] M. Ilić and F. Galiana, "Power systems operation: old vs. new," in *Power Systems Restructuring*. Springer, 1998, pp. 15–107.

[6] R. Singh, P. R. Kumar, and L. Xie, "Decentralized control via dynamic stochastic prices: The independent system operator problem," *arXiv preprint arXiv:1605.08926*, 2016.

[7] M. Huang, P. E. Caines, and R. P. Malhamé, "Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized $\epsilon$-Nash equilibria," *IEEE Transactions on Automatic Control*, vol. 52, pp. 1560–1571, 2007.

[8] J. M. Lasry and P. L. Lions, "Mean field games," *Japanese J. of Mathematics*, vol. 2, pp. 229–260, 2007.

[9] B. B. Jovanovic and R. W. Rosenthal, "Anonymous sequential games," *Journal of Mathematical Economics*, vol. 17, no. 1, pp. 77–87, 1988.