Inverse Learning of the Altruism and Cost Level in Mixed-Individual Mean Field Games

Haoyang Cao, Gökçe Dayanıklı, and Xiaofei Shi

***Abstract*— Understanding how humans respond to incentives, both at the individual and collective levels, is crucial to the design of effective policies. Within the continuous-time stochastic framework for large interacting populations, mean field games (MFGs) model populations of non-cooperative agents, whereas mean field control (MFC) describes the fully cooperative benchmark, interpreted in our setting as fully altruistic behavior. Mixed-individual MFGs interpolate between these two extremes through a parameter governing the degree of altruism. A central challenge for regulators and policymakers, however, is that intrinsic altruism levels and other private structural parameters, such as individual labor costs, are typically unobservable. To address this challenge, we develop an inverse learning framework for mixed-individual MFGs. Our approach enables the recovery of (latent) altruism and labor cost levels from noisy observations, with experiments demonstrating the feasibility and accuracy of our method. These findings underscore the promise of inverse MFG methodologies for uncovering latent preference structures in large populations, with important implications for incentive design, empirical behavioral modeling, and data-driven policy analysis.**

I. Introduction

Mean field control (MFC) [4], [8] and mean field games (MFGs) [17], [20], [21] have increasingly become a tool for modeling and analyzing large populations of interacting agents in different applications such as modeling of energy markets [1], [6], [15], financial systems [9], [19], or traffic management [12], [16]. These models approximate solutions to finite-agent systems by considering the limit in which the number of agents goes to infinity, focusing on the interactions between a representative agent and the population distribution of the states or controls of the agents. Although related, the two frameworks differ in their underlying strategic settings: MFC models cooperative agents and approximates the social optimum, whereas MFG models non-cooperative agents and approximates a Nash equilibrium. The distinction in their solutions is a consequence of how individuals perceive their influence on the mean field term. In MFC, each agent endogenizes their impact on the mean field and optimizes accordingly. In contrast, in MFG, each agent treats the mean field as exogenous, optimizes her best response to it, and the equilibrium is achieved by a fixed point argument.

However, in many real life applications especially the ones including the modeling of the human behavior, agents may display a mixture of non-cooperative and cooperative behaviors. Following this motivation mixed individual MFGs (MI-MFGs) have been introduced [14], [13]. In this model, each agent sees their effect on the mean field term depending on an altruism parameter. Varying this altruism parameter allows the model to smoothly transition from fully non-cooperative behavior to fully cooperative control. This modeling flexibility enables a more realistic representation of agent behavior in socio-economic and engineering systems.

From a mechanism design perspective, improving the societal outcomes requires understanding the underlying preferences and objectives of interacting individual agents. In practice, model parameters or their level of altruism are typically unknown and cannot be directly observed, but they can be inferred from observed equilibrium behavior. While MI-MFGs provide a flexible framework to model the combination of cooperative and non-cooperative behavior, existing works have mainly focused on the equilibrium analysis for a known model and in contrast learning of characteristics in a data-driven way remains unexplored. In this paper, we address this gap by studying the inverse learning of both the cost parameter and the altruism parameter in the MI-MFG. One of the motivations for proposing a data-driven learning method is the fact that they can be easily generalized to the higher dimensions.

Related literature on inverse learning in mean-field models is still sparse and is considerably more developed for MFGs than for MFC. For MFGs, [22] studied a class of discrete-time models in which the inverse problem can be reduced to an induced Markov decision process and addressed using deep inverse reinforcement learning. [10] later showed that this reduction is essentially limited to cooperative settings and proposed an individual-level inverse-learning formulation that directly captures mean-field interactions; see also [11] for an adversarial extension designed to improve robustness to imperfect demonstrations. More recently, [3], [2] developed maximum-causal-entropy and kernel-based formulations for stationary MFGs, providing more structured optimization frameworks and richer reward classes. By contrast, inverse problems for MFC have received much less attention and have mostly appeared in the forms of inverse optimal control and data-driven identification; representative examples include the inverse traffic-type MFC model in [18] and the data-driven inverse approach for potential mean-field models in [23].

Our contributions are three-fold. The first contribution is related to the analysis of MI-MFGs. Different from [14] and [13] that use forward-backward stochastic differential equations, we prove characterization of mixed

H. Cao is with Department of Applied Mathematics and Statistics, Johns Hopkins University hycao@jhu.edu

G. Dayanıklı is with Department of Statistics, University of Illinois, Urbana-Champaign, gokced@illinois.edu

X. Shi is with Department of Statistical Sciences, University of Toronto, xf.shi@utoronto.ca

individual mean-field Nash equilibrium (MI-MFNE) via Hamilton-Jacobi-Bellman (HJB) and Kolmogorov-Fokker-Planck (KFP) forward-backward partial differential equations (FBPDEs). We also reduce the FBPDE system to a system of coupled forward-backward ordinary differential equations (FBODEs) and give the existence and unique results for the MI-MFNE. Our second contribution is to introduce the *first inverse learning method* for the MI-MFGs with the motivation of understanding the altruism levels and cost parameters of individuals which may not be known explicitly from the observed behaviors. Finally, we validate the theoretical inverse learning findings with experiments. In these experiments, we demonstrate how altruism influences equilibrium behavior.

The paper is structured as follows. Section II introduces the mathematical setting and related equilibrium definition for an MI-MFG model of our interest and formalize our problem statement. We emphasize the perspectives of both the regulator and individual agents in this game setting. Section III presents the theoretical analysis of the MI-MFG model and the inverse learning method. Section IV introduces the inverse learning algorithm to learn the altruism levels and cost parameters and presents the experiment results that validate the theoretical findings. Finally in Section V, we summarize our findings to conclude and give our future directions.

II. Model

A. Mixed Individual Mean-Field Game

Inspired by [13], we introduce the formulation of a mixed individual mean-field game (MI-MFG) over a finite time horizon $[0,T]$ with fixed $T>0$. Given a filtered probability space $(\Omega,\mathscr{F},\mathbb{P},\mathbb{F}=(\mathscr{F}_t)_{t\geq 0})$ satisfying the usual conditions that supports a one-dimensional standard Brownian motion $W=(W_t)_{t\geq 0}$, we let $\xi\in\mathscr{F}_0$ such that $\xi\sim\mu_0$ and $\mathbb{E}[\xi^2]<\infty$. For a representative agent i, with b_α and b_x being constant parameters and $\sigma>0$ as the constant volatility, her state process evolves according to

$$dX_t^\alpha = (b_\alpha\alpha_t + b_x X_t)\,dt + \sigma dW_t, \qquad X_0\sim\xi, \tag{1}$$

under a control $\alpha=(\alpha_t)_{t\in[0,T]}\in\mathscr{A}$. Here, $\mathscr{A}$ denotes the set of admissible controls such that

$$\mathscr{A} := \Big\{ (\alpha_t)_{t\in[0,T]} \Big| \exists [0,T]\times\mathbb{R}\overset{\alpha}{\mapsto}\mathbb{R} \text{ s.t. } \alpha_t=\alpha(t,X_t^\alpha) \text{ and (1) has a unique strong solution with } \mathbb{E}\big[\sup_{t\in[0,T]}(X_t^\alpha)^2\big]<\infty \Big\}. \tag{2}$$

Given a flow of probability distributions $m=(m_t)_{t\in[0,T]}$, Agent i faces the following expected total cost

$$J(\alpha;m) = \frac{1}{2}\mathbb{E}\left[\int_0^T c_\alpha\alpha_t^2 + c_x(X_t^\alpha)^2 + c_\mu f(\mu_t^\alpha,m_t;\lambda)\,dt\right], \tag{3}$$

subject to (1), where $c_\alpha,c_x,c_\mu>0$ are constant coefficients. Here, the cost of interaction f is given by

$$f(\mu^\alpha,m;\lambda) = \big(\lambda\mathbb{E}_{Z\sim m}[Z] + (1-\lambda)\mathbb{E}_{Z\sim\mu^\alpha}[Z]\big)^2, \tag{4}$$

with $\mu^\alpha=(\mu_t^\alpha)_{[0,T]}=(\mathrm{Law}(X_t^\alpha))$ and $\lambda\in[0,1]$. Our model is presented under the one-dimensional state, control, and noise case for the sake of simplicity in presentation and notation. The technical analysis and learning approach can be extended to the higher dimensional setting in a straightforward manner. Define the following notion of equilibrium:

Definition 1. *A pair $(\hat\alpha,\hat m)$ is said to be a mixed individual mean-field Nash equilibrium (MI-MFNE) provided that*

1) *Given the distribution flow $\hat m$, $\hat\alpha\in\arg\min_{\alpha\in\mathscr{A}}J(\alpha;\hat m)$ subject to* (1)*, and in the meantime,*
2) *$\hat m_t=\mu_t^{\hat\alpha}=\mathscr{L}(X_t^{\hat\alpha})$ for all $t\in[0,T]$.*

From the above definition, we can see that the agents in this MI-MFG are interacting with each other through the cost structure in the minimization problem. In particular, the presence of the cost of interaction (4) encompasses an intermediate status between a non-cooperative game (i.e., an MFG) and a fully cooperative setting (i.e., an MFC). In this way, intuitively, the altruism parameter λ reflects the level of altruism of the agent: when λ is close to 1, an agent is more of a non-cooperative player in the game, competing with her opponents via the mean field m in the cost; when λ is close to 0, on the other hand, the agent is more collaborative with the rest of the population through μ^α. We note that, strictly speaking, $(1-\lambda)\in[0,1]$ represents the *altruism level*; for simplification, we will call λ the altruism parameter.

B. Regulator versus Individual Agents

As suggested by [13], a general MI-MFG can be used to study the phenomenon of "tragedy of the commons", such as overfishing, deforestation, traffic congestion, and so on in a more realistic manner. In these models, agents' state processes can be interpreted as individual wealth from exploiting common goods. The mean information models exhaustion of common goods. The individual overall cost functions usually take into account both the cost of labor and the regulatory cost to prevent the tragedy of the commons; agents' level of altruism will then impact the effectiveness of such regulatory costs.

Back to the specific game setting introduced in Section II-A, we now put it into the perspectives of regulator and individual agents, respectively. We assume that the state process (1) is public to both the regulator and the agents, that is, $b_x,b_\alpha\in\mathbb{R}$ are given. In the total cost (3), the individual agents determine the coefficient in the quadratic labor cost $c_\alpha>0$ as well as the level of altruism via the altruism parameter $\lambda\in[0,1]$; the regulator determines the $c_x,c_\mu\in(0,+\infty)$ in the regulatory cost. From the regulator's perspective, it is crucial to understand how agents respond to the regulatory cost when the altruism parameter λ and labor cost structure c_α are private to the agent. For the rest of this paper, we study how to infer λ, or jointly infer (λ,c_α) by observing agents' equilibrium behavior under a given game setting.

III. Method and Results

A. Analysis of the Mixed Individual MFG

We first characterize the equilibrium strategy for the representative agent i under a given game structure. The

characterization is given by a coupled forward-backward partial differential equations (FBPDE).

Theorem 1 (FBPDE characterization of MI-MFNE)**.** *A control $\hat{\alpha}$ is an MI-MFNE control profile if and only if:*

$$\hat{\alpha}_t = -\frac{b_\alpha \partial_x u(t,x)}{c_\alpha}, \tag{5}$$

where $(u,m) = (u_t, m_t)_{t\in[0,T]}$ solves the following forward-backward partial differential equation (FBPDE) system of Kolmogorov-Fokker-Planck (KFP) and Hamilton-Jacobi-Bellman (HJB) equations:

$$\begin{cases} -\partial_t u_t = \frac{1}{2}\sigma^2\partial^2_{xx}u(t,x) - \frac{b_\alpha^2}{2c_\alpha}\big(\partial_x u(t,x)\big)^2 + b_x x\partial_x u(t,x) \\ \qquad\qquad + \frac{c_x}{2}x^2 + \frac{c_\mu}{2}\bar{x}_t^2 + c_\mu(1-\lambda)\bar{x}_t x, \\ \partial_t m_t = \frac{\sigma^2}{2}\partial^2_{xx}m(t,x) - \partial_x\big(m(t,x)\big(-\frac{b_\alpha^2}{c_\alpha}\partial_x u(t,x) + b_x x\big)\big), \\ u_T = 0, \qquad m(0,x) = \mu_0(x), \end{cases} \tag{6}$$

where $\bar{x}_t = \int x m(t,x)dx$ is a function of time.

Proof. The first step of the proof is to realize that given the population distribution $\hat{m}$, the problem becomes a mean field control problem. Following the proof in [4] and by introducing an exogenous population distribution $\hat{m}$, we first write the Hamiltonian:

$$H(x,\alpha,\hat{m},\mu,y;\lambda) := \tfrac{c_\mu}{2}\left(\lambda\int\eta\hat{m}(\eta)d\eta + (1-\lambda)\int\xi\mu(\xi)d\xi\right)^2 + (b_\alpha\alpha + b_x x)y + \tfrac{c_\alpha}{2}\alpha^2 + \tfrac{c_x}{2}x^2.$$

Due to the convexity, the Hamiltonian is minimized if and only if $\hat{\alpha} = -\frac{b_\alpha y}{c_\alpha}$. The minimized Hamiltonian is therefore:

$$\hat{H}(x,\hat{m},\mu,y;\lambda) := \tfrac{c_\mu}{2}\left(\lambda\int\eta\hat{m}(\eta)d\eta + (1-\lambda)\int\xi\mu(\xi)d\xi\right)^2 - \tfrac{b_\alpha^2}{2c_\alpha}y^2 + b_x xy + \tfrac{c_x}{2}x^2.$$

Then the HJB is given by:

$$-\partial_t u(t,x) = \tfrac{1}{2}\sigma^2\partial^2_{xx}u(t,x) + \hat{H}(x,\hat{m},\mu,\partial_x u(t,x);\lambda) + \int_{\mathbb{R}}\frac{\partial\hat{H}}{\partial\mu}(\xi,\hat{m},\mu,\partial_x u(t,\xi);\lambda)(x)\mu(\xi)d\xi$$

where we have:

$$\frac{\partial\hat{H}}{\partial\mu}(\xi,\hat{m},\mu,\partial_x u(t,\xi);\lambda)(x) = c_\mu\Big((1-\lambda)\lambda\big(\int\eta\hat{m}(\eta)d\eta\big)x + (1-\lambda)^2\big(\int\eta\mu(\eta)d\eta\big)x\Big),$$

with $u(T,x) = 0$ due to zero terminal cost. The coupled forward equation is written by using KFP equation and plugging in the minimizer of the Hamiltonian.

The second step is to ensure the consistency condition holds. In other words, at the equilibrium $\hat{m}_t = \mu_t^{\hat{\alpha}} = \mu_t$. By introducing $\int\eta\hat{m}d\eta = \int\eta\mu(\eta)d\eta = \bar{x}$, we can conclude the final system. □

Remark III.1. It is worth emphasizing that the backward equation in (6) is not from dynamic programming principle but rather a variational approach introduced in [4]. Therefore, the function $u(t,x)$ does not refer to the game value in equilibrium. Note that in the definition of admissible control set $\mathscr{A}$ in (2), the feedback control depends only on the state, not the augmented state-law space. Therefore, the u function corresponds to an adjoint state in the variational approach.

We can further specify the analytical solution of the equilibrium strategy.

Proposition 1 (FBODE characterization of MI-MFNE)**.** *Let $(r_t, q_t, s_t, \bar{x}_t)$ denote the solution of the following forward-backward ordinary differential equation (FBODE) system:*

$$\begin{cases} -\dot{r}_t = -\frac{2b_\alpha^2}{c_\alpha}r_t^2 + 2b_x r_t + \frac{c_x}{2}, & r_T = 0, \quad (7) \\ -\dot{q}_t = -\frac{2b_\alpha^2}{c_\alpha}r_t q_t + b_x q_t + c_\mu(1-\lambda)\bar{x}_t, & q_T = 0, \quad (8) \\ -\dot{s}_t = \sigma^2 r_t - \frac{b_\alpha^2}{2c_\alpha}q_t^2 + \frac{c_\mu}{2}(\bar{x}_t)^2, & s_T = 0, \quad (9) \\ \dot{\bar{x}}_t = -\frac{b_\alpha^2}{c_\alpha}(2r_t\bar{x}_t + q_t) + b_x\bar{x}_t, & \bar{x}_0 = \bar{\mu}_0, \quad (10) \end{cases}$$

where $\bar{\mu}_0 = \int_{\mathbb{R}}\xi\mu_0(\xi)d\xi$. Then

$$\hat{\alpha}_t = -\frac{b_\alpha}{c_\alpha}(2r_t X_t + q_t) \tag{11}$$

is the MI-MFNE control.

Proof. Following the linear quadratic form of the model, we propose the following ansatz $u(t,x) = r_t x^2 + q_t x + s_t$ where r_t, q_t, s_t are deterministic functions of time. Plugging in $\partial_t u(t,x) = \dot{r}_t x^2 + \dot{q}_t x + \dot{s}_t$, $\partial_x u(t,x) = 2r_t x + q_t$, and $\partial^2_{xx}u(t,x) = 2r_t$ in the backward differential equation and matching the terms including x^2, x, and without any x, we get the backward ODE system presented in (7)-(9). The forward ODE is found by plugging in the forward PDE below:

$$\begin{aligned}\dot{\bar{x}}_t &= \frac{d}{dt}\int x m(t,x)dx = \int x\partial_t m(t,x)dx, \\ &= \int x\tfrac{\sigma^2}{2}\partial^2_{xx}m(t,x) - \nabla_x\big(m(t,x)\big(-\tfrac{b_\alpha^2}{c_\alpha}\partial_x u(t,x) + b_x x\big)\big)dx.\end{aligned}$$

By plugging in $\partial_x u(t,x) = 2r_t x + q_t$ in the above equation and using integration by parts we conclude:

$$\dot{\bar{x}}_t = -\frac{b_\alpha^2}{c_\alpha}(2r_t\bar{x}_t + q_t) + b_x\bar{x}_t.$$

□

Finally, we emphasize the uniqueness of such an equilibrium strategy.

Theorem 2. *There exists a unique MI-MFNE under small time condition.*

Proof. Following Theorem 1 and Proposition 1, finding MI-MFNE is equivalent to solving the FBODE system in Proposition 1. We first emphasize that (7) is a scalar Riccati equation and it has the following explicit unique solution:

$$r_t = \frac{\frac{c_x}{2}\big(e^{2R(T-t)} - 1\big)}{(R-b_x)e^{2R(T-t)} + (b_x + R)}, \tag{12}$$

where $R = \sqrt{b_x^2 + \frac{b_\alpha^2 c_x}{c_\alpha}}$. Given $r = (r_t)_{t\in[0,T]}$ process, the backward ODE for $q = (q_t)_{t\in[0,T]}$ and the forward ODE for $\bar{x} = (\bar{x}_t)_{t\in[0,T]}$ are coupled.

Given any q, we can write another mapping $\bar{x} = \phi_1(q)$ and given any $\bar{x}$ we can write a mapping $q' = \phi_2(\bar{x})$. In order to show that the coupled system of q and $\bar{x}$ has a unique solution, we need to show that the composition mapping $\phi =$

$\phi_2 \circ \phi_1 : q \mapsto q'$ is a contraction. We define the sup norm for $f \in C([0,T])$ as $||f||_T := \sup_{t\in[0,T]} |f(t)|$, furthermore we define notation $\Delta g = g^1 - g^2$ where $g \in \{q_t, \bar{x}_t, \dot{q}_t, \dot{\bar{x}}_t\}$.

Step 1: We first fix q^1, q^2 and can write:

$$\Delta\dot{\bar{x}}_t = \big(-\frac{2b_\alpha^2 r_t}{c_\alpha} + b_x\big)\Delta\bar{x}_t - \frac{b_\alpha^2}{c_\alpha}\Delta q_t, \quad \Delta\bar{x}_0 = 0.$$

Then we have:

$$\begin{aligned}|\Delta\bar{x}_t| &= \int_0^t \Big|\Big(\big(-\frac{2b_\alpha^2 r_s}{c_\alpha} + b_x\big)\Delta\bar{x}_s - \frac{b_\alpha^2}{c_\alpha}\Delta q_s\Big)\Big| ds \\ &\le \int_0^t \Big|\big(-\frac{2b_\alpha^2 r_s}{c_\alpha} + b_x\big)\Big| |\Delta\bar{x}_s| + \frac{b_\alpha^2}{c_\alpha}|\Delta q_s| ds\end{aligned}$$

By using Grönwall's inequality we have:

$$\begin{aligned}|\Delta\bar{x}_t| &\le \int_0^t \frac{b_\alpha^2}{c_\alpha}|\Delta q_s| ds \times \exp\left(\int_0^t \Big|-\frac{2b_\alpha^2 r_s}{c_\alpha} + b_x\Big| ds\right) \\ &\le C_{\bar{x}} \int_0^T |\Delta q_s| ds\end{aligned}$$

where the constant $C_{\bar{x}} := \frac{b_\alpha^2}{c_\alpha}\exp\big(T\big(\frac{2b_\alpha^2|r|_T}{c_\alpha} + |b_x|\big)\big)$ with $|r|_T = \sup_{s\in[0,T]}|r_s| < c_x e^{2RT}/4R$ only depends on $b_\alpha, b_x, c_\alpha, c_x$ and T.

Step 2: Next, given $\bar{x}^1, \bar{x}^2$, we have:

$$-\Delta\dot{q}'_t = \Big(-\frac{2b_\alpha^2}{c_\alpha} r_t + b_x\Big)\Delta q'_t + c_\mu(1-\lambda)\Delta\bar{x}_t, \quad \Delta q'_T = 0$$

By plugging in the bound we found in **Step 1** and by using Grönwall's inequality, we have:

$$\begin{aligned}|\Delta q'_t| &= \int_t^T \Big|\big(\frac{2b_\alpha^2}{c_\alpha} r_s - b_x\big)\Delta q'_s - c_\mu(1-\lambda)\Delta\bar{x}_s\Big| ds \\ &\le \int_t^T \Big|\big(\frac{2b_\alpha^2}{c_\alpha} r_s + |b_x|\big)\Big| |\Delta q'_s| + c_\mu(1-\lambda)|\Delta\bar{x}_s| ds \\ &\le \int_t^T \Big|\big(\frac{2b_\alpha^2}{c_\alpha} r_s + |b_x|\big)\Big| |\Delta q'_s| ds + T c_\mu(1-\lambda) C_{\bar{x}} \int_0^T |\Delta q_s| ds \\ &\le \exp\Big(T\big(\tfrac{2b_\alpha^2}{c_\alpha}|r|_T + |b_x|\big)\Big) T c_\mu(1-\lambda) C_{\bar{x}} \int_0^T |\Delta q_s| ds.\end{aligned}$$

Then we have $||\Delta q'||_T \le C_q ||\Delta q||_T$ where $C_q := \exp\Big(T\big(\frac{2b_\alpha^2}{c_\alpha}|r|_T + |b_x|\big)\Big) T^2 c_\mu(1-\lambda)C_{\bar{x}}$. When T is small enough, we have $C_q < 1$ hence leads to the mapping $q \mapsto q'$ being a contraction mapping. Then by Banach fixed point theorem, there exists a unique solution to the coupled q and $\bar{x}$ system given r. Finally, s_t process is determined uniquely given the unique $(q_t, r_t, \bar{x}_t)$ processes. □

The contraction argument above establishes existence and uniqueness for a sufficiently small time horizon. Extensions to arbitrary finite horizons may be possible under additional structural conditions; related continuation methods for forward-backward systems can be found in [7, Section 4.1.2]. Alternatively, existence alone may be approached through Schauder's fixed-point theorem, provided that the required compactness and invariant-set conditions are verified.

B. Inverse Learning for the Level of Altruism and the Labor Cost

Next, we take the regulator's viewpoint and study the inference of altruism level via altruism parameter λ and the labor cost parameter c_α. The method is inspired by an inverse learning approach, where the inference procedure is given by the following identifiability result.

Theorem 3. *Suppose the regulator observes a linear equilibrium strategy $a(t,x) = k(t)x + b(t)$ for $t \in [0,T]$, where $k, b : [0,T] \to \mathbb{R}$ satisfy*

1) *there exists $\rho > b_x$ such that for any $t \in [0,T)$,*
$$k(t) = \frac{b_x^2 - \rho^2}{b_\alpha} \frac{e^{2\rho(T-t)} - 1}{(\rho - b_x)e^{2\rho(T-t)} + b_x + \rho} < 0;$$
2) *there exists $c_0 \le 0$ such that for any $t \in [0,T]$,*
$$\begin{cases} b(t) = b_\alpha c_\mu c_0 \int_t^T \tilde{x}(s)\exp\{b_x(s-t) + b_\alpha\int_t^s k(u)du\} ds, \\ \tilde{x}(t) = \bar{\mu}_0 \exp\{b_x t + b_\alpha \int_0^t k(s)ds\} \\ \qquad\qquad + b_\alpha \int_0^t b(s)\exp\{b_x(t-s) + b_\alpha\int_s^t k(u)du\} ds. \end{cases}$$

If the representative agent i reveals her labor cost parameter $c_\alpha > 0$, then the level of altruism $1-\lambda$ can be uniquely identified by using

$$\widehat{\lambda} = 1 + \frac{c_\alpha b(t)\exp\{b_x t + b_\alpha\int_0^t k(s)ds\}}{b_\alpha c_\mu \int_t^T \tilde{x}(s)\exp\{b_x s + b_\alpha\int_0^s k(u)du\} ds}, \tag{13}$$

for any given $t \in [0,T)$ such that the denominator in (13) is nonzero. Otherwise, assume $b_x, b_\alpha > 0$, then (c_α, λ) can be uniquely identified as

$$\widehat{c}_\alpha = \frac{b_\alpha^2 c_x}{\rho_0^2 - b_x^2}, \tag{14}$$

$$\widehat{\lambda} = 1 + \frac{\widehat{c}_\alpha b(t)\exp\{b_x t + b_\alpha\int_0^t k(s)ds\}}{b_\alpha c_\mu \int_t^T \tilde{x}(s)\exp\{b_x s + b_\alpha\int_0^s k(u)du\} ds}, \tag{15}$$

where ρ_0 is the unique root on $(b_x, +\infty)$ for

$$f(\rho) = k(t_0) + \frac{\rho^2 - b_x^2}{b_\alpha}\cdot\frac{e^{2\rho(T-t_0)} - 1}{(\rho - b_x)e^{2\rho(T-t_0)} + \rho + b_x} = 0, \tag{16}$$

for any fixed $t_0 \in [0,T)$.

Proof. Under the observed linear policy $a(t,x) = k(t)x + b(t)$, the state dynamics (1) implies that

$$\mathbb{E}[X_t^a] = \bar{\mu}_0 + \int_0^t [b_x + b_\alpha k(s)]\mathbb{E}[X_s^a] + b_\alpha b(s) ds, \quad t \in [0,T].$$

Then $\mathbb{E}[X_t^a] = \tilde{x}(t)$ for all $t \in [0,T]$. Also, note that conditions 1) and 2) are necessary for strategy a to be MI-MFNE specified in Proposition 1.

Suppose $c_\alpha > 0$ is given. Then from (8), we know that

$$b'(t) + [b_x + b_\alpha k(t)]b(t) - \frac{b_\alpha c_\mu}{c_\alpha}(1-\lambda)\bar{x}(t) = 0, \quad t \in [0,T).$$

Condition 2) implies that (13) is time-invariant thus the conclusion follows.

Suppose $c_\alpha > 0$ is not known. We can write

$$f(\rho) = k(t_0) + f_1(\rho)\cdot f_2(\rho),$$

with

$$f_1(\rho) = \frac{\rho + b_x}{b_\alpha}, \quad f_2(\rho; t_0) = 1 - \frac{2\rho}{(\rho - b_x)e^{2\rho(T-t_0)} + \rho + b_x}.$$

Notice that $f(b_x) = k(t_0) < 0$, and $f(\rho) \to +\infty$ as $\rho \to +\infty$. Then on $(b_x, +\infty)$ and for every $t_0 \in [0,T)$, f_1 is strictly increasing, and

$$f_2'(\rho) = 2\frac{2\rho(T-t_0)e^{2\rho(T-t_0)}(\rho-b_x) + b_x(e^{2\rho(T-t_0)}-1)}{[(\rho-b_x)e^{2\rho(T-t_0)}+\rho+b_x]^2} > 0.$$

Hence, f is strictly increasing on $(b_x, +\infty)$ for every $0 \leq t_0 < T$, and with condition 1), $f = 0$ admits a unique root ρ_0. Following (12), (14) uniquely holds; (15) together with its uniqueness follows from same arguments for (13) with $c_\alpha = \widehat{c}_\alpha$. □

Remark III.2. 1) Conditions 1) and 2) in Theorem 3 ensure that the strategy in the linear feedback form is indeed an equilibrium strategy in the MI-MFG specified in Section II. Any violation will lead to the issue of model misspecification.

2) Theorem 3 tells us that the unknown parameters (c_α, λ) can be uniquely determined if the equilibrium policy is given. That is, our inverse learning problem does not suffer from nonidentifiability issue which is common in inverse reinforcement learning; see for instance [5]. This is likely due to the structural assumptions we have made to the game setting. The identifiability guarantee for general inverse learning problems for mean-field game/control remains to be explored in detail.

IV. Learning Algorithm and Results

Theorem 3 characterizes the inverse learning result for the altruism level and the labor cost parameter if the linear feedback form of the MI-MFNE is explicitly known by the regulator. In practice, however, the equilibrium strategy is only available as sample paths of actions. Moreover, the regulator is often only able to observe agents' state and action trajectories in the discrete time. Lastly, the regulator can collect trajectories for only finite number of agents, though this number can be large. To address the gap between the theoretical and the practical settings, we now propose the corresponding learning approach.

A. Algorithm Setting

With pre-specified $(c_x, c_\mu) \in (0,+\infty)^2$, the regulator is observing an N-agent mixed-individual game in Nash equilibrium over the finite time horizon $[0,T]$, $T > 0$. The regulator records agents' states and actions every $\Delta t > 0$ time. Therefore, at each round of observation, the regulator collects a set of N trajectories $\boldsymbol{\tau} = \{\tau^{(i)}\}_{i=1}^N$, where $\tau^{(i)} = \left\{(X_t^{(i)}, a_t^{(i)})\right\}_{t\in\mathbb{T}}$ with $\mathbb{T} = \left\{T \wedge i\Delta t : i = 0,\dots,N_T, \Delta t = \frac{T}{N_T}\right\}$. From $\boldsymbol{\tau}$, the regulator can estimate the population mean process $\tilde{X} = \{\tilde{X}_t\}_{t\in\mathbb{T}}$ where for any $t \in \mathbb{T}$,

$$\tilde{X}_t = \frac{1}{N}\sum_{i=1}^N X_t^{(i)}.$$

We assume that the trajectories follow a time-homogeneous Gaussian transition kernel,

$$X_{T\wedge(t+\Delta t)}^{(i)} | (X_t^{(i)}, a_t^{(i)}) \sim N\left((1+b_x\Delta t)X_t^{(i)} + b_\alpha a_t^{(i)}\Delta t, \sigma^2(\Delta t \wedge (T-t))\right), \tag{17}$$

where parameters $b_x, b_\alpha > 0$ are explicitly known by the regulator; initial distribution for $X_0^{(i)}$ as well as the volatility parameter σ remain unknown.

B. Inverse Learning Procedure

a) Learning the linear feedback form of equilibrium strategy: Since the trajectories $\boldsymbol{\tau}$ are actually collected from the MI-MFNE, then there exists a true underlying linear form $[0,T]\times\mathbb{R} \overset{a}{\mapsto} k(t)x + b(t)$ such that for any $i = 1,\dots,N$ and any $t \in \mathbb{T}$,

$$a_t^{(i)} = k(t)X_t^{(i)} + b(t).$$

Moreover, conditions 1) and 2) in Theorem 3 are satisfied. Therefore, we directly learn (k,b) from linear regression,

$$\begin{aligned}(k,b) &= \text{Regression}(\boldsymbol{\tau}) \\ &= \underset{(\kappa,\beta)=\{(\kappa_t,\beta_t)\}_{t\in\mathbb{T}}}{\arg\min} \frac{1}{N}\sum_{i=1}^N\sum_{t\in\mathbb{T}}\left(\kappa_t X_t^{(i)} + \beta_t - a_t^{(i)}\right)^2.\end{aligned} \tag{18}$$

b) Learning the altruism level $1-\lambda \in [0,1]$ when $c_\alpha > 0$ is given: From Proposition 1 and Theorem 3, we know that

$$k_t = -\frac{2b_\alpha}{c_\alpha}r_t, \quad b_t = -\frac{b_\alpha}{c_\alpha}q_t, \quad t \in \mathbb{T},$$

where the paths r and q satisfy (7) and (8), respectively. When $c_\alpha > 0$ is given, we construct inference result of λ according to (13). Notice that, by linearity of expectation, $\tilde{X}$ is an unbiased estimator for $\tilde{x}$ specified in condition 2) of Theorem 3. We also introduce the following approximations of integrals in discrete time,

$$E_t = \exp\left\{b_x t + b_\alpha\Delta t \sum_{u\in\mathbb{T},u<t} k_u\right\}, \tag{19}$$

$$\mathscr{E}_t = \Delta t \sum_{u\in\mathbb{T},u\in[t,T)} \tilde{X}_u E_u, \tag{20}$$

for any $t \in \mathbb{T}\setminus\{T\}$. Fix any $t \in \mathbb{T}\setminus\{T\}$, construct the inferred λ as follows,

$$\widehat{\lambda} = 1 + \frac{c_\alpha b_t E_t}{b_\alpha c_\mu \mathscr{E}_t}; \tag{21}$$

equivalently, the inferred altruism level $1-\lambda$ is given by

$$1-\widehat{\lambda} = -\frac{c_\alpha b_t E_t}{b_\alpha c_\mu \mathscr{E}_t}.$$

c) Jointly learning the altruism level $1-\lambda \in [0,1]$ and the labor cost parameter $c_\alpha > 0$: Now that $c_\alpha > 0$ is to be estimated as well, according to Theorem 3, we construct the estimated results of c_α and λ according to (14) and (15) sequentially. First, we fix any $t_0 \in \mathbb{T}\setminus\{0,T\}$. Following (16), define

$$f(\rho;t_0) = k_{t_0} + \frac{\rho^2 - b_x^2}{b_\alpha}\cdot\frac{e^{2\rho(T-t_0)}-1}{(\rho-b_x)e^{2\rho(T-t_0)}+\rho+b_x}.$$

Then, find the root $\hat{R} \in (b_x, +\infty)$ such that $f(\hat{R};t_0) = 0$. By (14), construct the learning result for c_α as

$$\widehat{c}_\alpha = \frac{b_\alpha^2 c_x}{\hat{R}^2 - b_x^2}. \tag{22}$$

Subsequently, plugging $\widehat{c}_\alpha$ into by (21), construct the learning result for λ as

$$\widehat{\lambda} = 1 + \frac{\widehat{c}_\alpha b_t E_t}{b_\alpha c_\mu \mathscr{E}_t}, \tag{23}$$

for any $t \in [0,T)$; equivalently, the altruism level $1-\lambda$ is given by

$$1-\widehat{\lambda} = -\frac{\widehat{c}_\alpha b_t E_t}{b_\alpha c_\mu \mathscr{E}_t}.$$

d) Algorithm: We present the above learning procedure as in the following Algorithm 1. To reduce the variance related to population mean estimate, we let the regulator conduct M rounds of observations over this N-agent system. The output is then the average over M observations (i.e., samples).

Algorithm 1: Estimation of altruism level and labor cost

Setting: number of total observations M for N-agent system; time horizon $T>0$, time steps $N_T \in \mathbb{N}^+$, discretization precision $\Delta t = \frac{T}{N_T}$, time index $\mathbb{T} = [0, \Delta t, 2\Delta t, \ldots, N_T \Delta t]$, time point used in learning $t_0 \in \mathbb{T} \setminus \{0, T\}$; boolean variable $I = \mathbb{1}\{c_\alpha \text{ is given}\}$
Initialization: null lists C_α and Λ; $m = 0$;
while $m < M$ **do**
  $E =$ all 0 list of length t_0, $\mathscr{E} = 0$;
  Sample trajectories $\boldsymbol{\tau}$ for the N agents under MI-MFNE
  $(k, b) = \text{Regression}(\boldsymbol{\tau})$
  $\tilde{X} = \frac{1}{N}\sum_{i=1}^N X^{(i)}$
  for t is in $[t_0, \ldots, (N_T-1)\Delta t]$ **do**
    E.append(E_t) according to (19)
  end for
  $\mathscr{E} = \mathscr{E}_{t_0}$ according to (20) with $E_u = E[u]$ for $u \in [t_0, \ldots, (N_T-1)\Delta t]$
  if $I = 1$ **then**
    C_α.append(c_α), Λ.append$\left(1 + \frac{c_\alpha b_{t_0} E[t_0]}{b_\alpha c_\mu \mathscr{E}}\right)$
  else
    Solve $f(\rho; t_0) = 0$ for $\hat{R}$ on $(b_x, +\infty)$
    $\widehat{c}_\alpha = \frac{b_\alpha^2 c_x}{\hat{R}^2 - b_x^2}$
    C_α.append$(\widehat{c}_\alpha)$, Λ.append$\left(1 + \frac{\widehat{c}_\alpha b_{t_0} E[t_0]}{b_\alpha c_\mu \mathscr{E}}\right)$
  end if
  $m{+}{+}$;
end while
return $\texttt{mean}(C_\alpha)$, $\texttt{mean}(1-\Lambda)$, $\texttt{std}(C_\alpha)$, $\texttt{std}(\Lambda)$.

C. Experimental Results

We test Algorithm 1 with the following parameter settings:

$$T = 1,\ b_\alpha = 1,\ b_x = 0.5,\ c_\mu = 1,\ c_x = 1.$$

We generate our observations of $\{\alpha^{(i)}, X^{(i)}\}_{t\in\mathbb{T}}$ through (17), in which the initial distribution ξ of each agent follows an i.i.d normal distribution with mean $\bar{\mu}_0 = 0.5$ and variance the volatility $\sigma_0^2 = 0.01$ and where $\sigma = 0.1$. With the ground truth for the altruism level and labor costs being

$$1-\lambda = 0.1, \qquad c_\alpha = 1,$$

we obtain simulated 100-agent system action-state pairs using $N_T = 1000$ time steps so that the discretization precision is $\Delta t = 0.001$. We compare learned altruism levels and labor costs from the proposed algorithm to the ground truth.

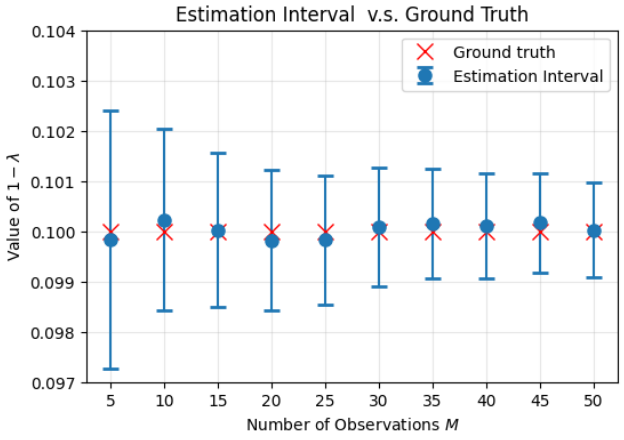

Fig. 1. Inverse learning for altruism level $1-\lambda$ when c_α is known.

a) Inference for altruism level: When the labor cost $c_\alpha = 1$ is known, we choose $t_0 = T/2$ to apply Algorithm 1. In a 100-agent model, we can obtain the following estimation for the altruism level $1-\lambda$ as the number of observations M of the system increases from 1 to 50 as illustrated in Figure 1. The estimated interval is constructed by

$$[\texttt{mean}(1-\Lambda) - 2.58\texttt{std}(\Lambda), \texttt{mean}(1-\Lambda) + 2.58\texttt{std}(\Lambda)].$$

Here, 2.58 corresponding to the 99.5% upper quantile of standard normal distribution. We can clearly see that, the overall estimation is very accurate, and the estimation interval becomes smaller when the number of observations increases, as expected.

b) Inference for labor cost and altruism level: When the labor cost c_α is unknown, we fix $t_0 = T/2$ to apply Algorithm 1. In a 100-agent model, the estimation error for c_α via Algorithm 1 is $|c_\alpha - \widehat{c}_\alpha| \approx 2.2 \times 10^{-10}$, with a standard deviation $O(10^{-16})$. Thus we can claim the success of recovering the labor costs. Figure 2 shows the estimation of the altruism level $1-\lambda$ with increasing number of observations, where the estimation interval is constructed in the same manner as before. As the number of observations M of the system increases from 1 to 50, the estimation gets more and more accurate as in the previous case, with only slightly larger variance (where the difference of variance is of order 10^{-4}).

V. Conclusion and Future Work

In this paper, we study inverse learning in MI-MFGs, with the goal of identifying both labor cost parameters and the altruism levels in a data-driven way from observed equilibrium behavior. We introduced and proved forward-backward PDE-based characterization for the MI-MFNE and established its existence and uniqueness. Furthermore, we analyzed the inverse learning approach and established identifiability results. Our results constitute the first inverse learning framework for MI-MFGs and offer theoretical guarantees for altruism and cost parameter recovery. We further provided learning approach and our experiments illustrate the effectiveness of the proposed approach. Our findings

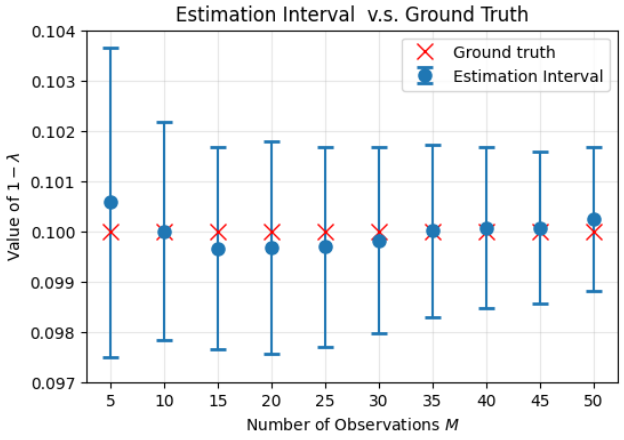

Fig. 2. Inverse learning for altruism level $1-\lambda$ when c_α is unknown.

contribute to the understanding of the agent behavior in large systems in which both cooperative and non-cooperative tendencies can occur.

Several directions remain open for future research. In the present model, the relationship between the cost structure and the equilibrium strategy is established through a variational approach. A natural next step is to develop an MI-MFG formulation that admits a dynamic programming principle, thereby enabling inverse *reinforcement* learning methods for recovering altruism levels from more general forms of observed behavior. Such a formulation would require equilibrium feedback strategies to be defined on an augmented state-and-distribution space, reflecting the dependence of the agents' decisions on both their individual states and the population distribution. Another important direction is to formulate a genuine Stackelberg game between a regulator and a population of mixed-individual agents. This would allow the regulator not only to infer latent preferences, but also to use the inferred model to design *optimal* incentives and anticipate the resulting equilibrium response.

We also plan to extend the proposed learning methodology to settings in which the regulator has access only to state trajectories, only to control trajectories, or to aggregate population-level observations. Further extensions may incorporate multidimensional state dynamics, nonlinear cost structures, heterogeneous populations, and imperfect or partially observed equilibrium behavior. In these more general settings, establishing identifiability will remain a fundamental challenge, since different combinations of preferences and structural parameters may generate observationally similar behavior. Developing finite-sample guarantees, uncertainty quantification, and computationally efficient learning algorithms will therefore be essential for translating inverse MI-MFG methods into reliable tools for empirical behavioral modeling and data-driven policy design.

References

[1] René Aïd, Roxana Dumitrescu, and Peter Tankov. The entry and exit game in the electricity markets: A mean-field game approach. *Journal of Dynamics and Games*, 2021.

[2] Berkay Anahtarci, Can Deha Kariksiz, and Naci Saldi. Kernel based maximum entropy inverse reinforcement learning for mean-field games. *arXiv preprint arXiv:2507.14529*, 2025.

[3] Berkay Anahtarci, Can Deniz Kariksiz, and Naci Saldi. Maximum causal entropy IRL in mean-field games and GNEP framework for forward RL. *Journal of Machine Learning Research*, 26:1–72, 2025.

[4] Alain Bensoussan, Jens Frehse, Phillip Yam, et al. *Mean field games and mean field type control theory*, volume 101. Springer, 2013.

[5] Haoyang Cao, Samuel Cohen, and Lukasz Szpruch. Identifiability in inverse reinforcement learning. *Advances in Neural Information Processing Systems*, 34:12362–12373, 2021.

[6] René Carmona, Gökçe Dayanıklı, and Mathieu Laurière. Mean field models to regulate carbon emissions in electricity production. *Dynamic Games and Applications*, 12(3):897–928, 2022.

[7] René Carmona and François Delarue. *Probabilistic theory of mean field games with applications. I*, volume 83 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2018. Mean field FBSDEs, control, and games.

[8] René Carmona, François Delarue, and Aimé Lachapelle. Control of mckean–vlasov dynamics versus mean field games. *Mathematics and Financial Economics*, 7(2):131–166, 2013.

[9] René Carmona, Jean-Pierre Fouque, and Li-Hsien Sun. Mean field games and systemic risk. *Communications in Mathematical Sciences*, 13(4):911–933, 2015.

[10] Yang Chen, Libo Zhang, Jiamou Liu, and Shuyue Hu. Individual-level inverse reinforcement learning for mean field games. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 253–262, 2022.

[11] Yang Chen, Libo Zhang, Jiamou Liu, and Michael Witbrock. Adversarial inverse reinforcement learning for mean field games. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1088–1096, 2023.

[12] Geoffroy Chevalier, Jerome Le Ny, and Roland Malhamé. A micro-macro traffic model based on mean-field games. In *2015 American Control Conference (ACC)*, pages 1983–1988. IEEE, 2015.

[13] Gokce Dayanıklı and Mathieu Lauriere. Cooperation, competition, and common pool resources in mean field games. *arXiv preprint arXiv:2504.09043*, 2025.

[14] Gökçe Dayanıklı and Mathieu Laurière. How can the tragedy of the commons be prevented?: Introducing linear quadratic mixed mean field games. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, pages 6792–6798, 2024.

[15] Boualem Djehiche, Julian Barreiro-Gomez, and Hamidou Tembine. Price dynamics for electricity in smart grid via mean-field-type games. *Dynamic Games and Applications*, 10:798–818, 2020.

[16] Kuang Huang, Xu Chen, Xuan Di, and Qiang Du. Dynamic driving and routing games for autonomous vehicles on networks: A mean field game approach. *Transportation Research Part C: Emerging Technologies*, 128:103189, 2021.

[17] Minyi Huang, Roland P Malhamé, and Peter E Caines. Nash certainty equivalence in large population stochastic dynamic games: Connections with the physics of interacting particle systems. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 4921–4926. IEEE, 2006.

[18] Pushkin Kachroo, Shaurya Agarwal, and Shankar Sastry. Inverse problem for non-viscous mean field control: Example from traffic. *IEEE Transactions on Automatic Control*, 61(11):3412–3421, 2016.

[19] Aimé Lachapelle, Jean-Michel Lasry, Charles-Albert Lehalle, and Pierre-Louis Lions. Efficiency of the price formation process in presence of high frequency participants: a mean field game analysis. *Mathematics and Financial Economics*, 10:223–262, 2016.

[20] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. I – Le cas stationnaire. *Comptes Rendus. Mathématique*, 343(9):619–625, November 2006.

[21] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, March 2007.

[22] Jiachen Yang, Xiaojing Ye, Rakshit Trivedi, Huan Xu, and Hongyuan Zha. Learning deep mean field games for modeling large population behavior. In *International Conference on Learning Representations (ICLR)*, 2018.

[23] Jingguo Zhang, Xianjin Yang, Chenchen Mou, and Chao Zhou. Learning surrogate potential mean field games via gaussian processes: A data-driven approach to ill-posed inverse problems. *Journal of Computational Physics*, 543:114412, 2025.