Car Adoption: Fuel or Electric*

Russell Cooper † Dandan Wang ‡

June 5, 2026

Abstract

Network externalities can make technology adoption depend on equilibrium selection. We study this issue in a dynamic vehicle replacement model in which heterogeneous households choose between internal combustion and electric vehicles, manufacturers set state dependent prices, and charging infrastructure responds endogenously to the installed base of electric vehicles. The calibrated economy has a low adoption trap, an unstable tipping region, and a high adoption steady state. A rational expectations benchmark that selects the high branch reaches rapid EV diffusion, while algorithmic Markov play based on realized experience can remain in the low adoption basin from the same initial state. We interpret learning not as a literal description of household cognition, but as a disciplined behavioral selection device for dynamic adoption games with complementarities. The calibrated counterfactuals separate price wedge and state transition margins. Purchase subsidies lower private user costs but have limited effects when infrastructure remains scarce. Infrastructure investment shifts the state variable governing the network externality and can move the economy toward the tipping region.

JEL Codes: L13, L62, Q55, C63, C73, D83
Keywords: Electric vehicles, network externalities, equilibrium selection, reinforcement learning, oligopoly pricing, technology adoption, infrastructure policy

*Preliminary and incomplete.

†Liaoning University, russell.cooper@eui.eu
‡Liaoning University,wangdandan@lnu.edu.cn

1 Introduction

The transition from internal combustion vehicles to electric vehicles is a coordination problem as much as a cost problem. Consumers are reluctant to adopt EVs when charging infrastructure is thin. Charging investors are reluctant to build dense networks before there is a sufficiently large installed base of EVs. This feedback creates the possibility that the same fundamentals support both a low adoption trap and a high adoption state. The economic question is not only which steady states exist, but which one the market is likely to select.

This paper studies that selection problem in a dynamic vehicle replacement model. Households differ by income, vehicle age, and current fuel type. They choose whether to keep their current vehicle, replace it with an internal combustion vehicle, or replace it with an EV. Manufacturers set state dependent prices. Charging infrastructure evolves endogenously with the installed base of EVs and public investment. The key state variable is therefore not only the current EV share, but also the infrastructure stock that determines future EV operating costs.

The first contribution is theoretical. We formulate a dynamic adoption game with endogenous infrastructure and state dependent pricing, prove existence in a discretized economy, and give crossing conditions under which the calibrated aggregate system has multiple steady states. The theory is deliberately conservative. We do not claim a global theorem that Q learning converges to Nash equilibrium in this game. Instead, we use stochastic approximation arguments to discipline the learning process and then rely on equilibrium diagnostics Bellman errors, Nash gaps, perturbation reversion tests, and basin maps to evaluate the learned outcome.

The second contribution is conceptual. Rational expectations requires a branch selection whenever the equilibrium correspondence is multi-valued. Algorithmic Markov play (AMP) provides an alternative selection discipline: agents update behavior from realized payoff and transition experience. In the calibrated economy, the selected rational expectations benchmark coordinates on the high adoption branch, while AMP remains in the low adoption basin from the same initial aggregate state. This difference is not a failure of numerical optimization; it is a selection result produced by the interaction between learning and the infrastructure externality.

The third contribution is a policy decomposition. Purchase subsidies and infrastructure investments act on different margins. Subsidies reduce the private price wedge between EVs and gasoline vehicles. Infrastructure policy changes the aggregate state that governs the network externality. In the calibrated counterfactuals, subsidy only policy improves adoption but

does not cross the tipping region, whereas infrastructure investment moves the state variable directly and generates larger adoption responses. These results should be read as mechanism evidence, not as an estimated evaluation of U.S. EV policy.

A full empirical version is left for future work, but the required estimation strategy is part of the paper's research agenda. Section 7.2 describes how the model can be taken to household vehicle choice data, state registration panels, charger location data, gasoline and electricity prices, and policy variation. The current draft therefore separates two tasks. It first establishes the economic mechanism and selection logic in a calibrated environment. It then lays out the data and moments needed to estimate the primitives in a subsequent empirical implementation.

The rest of the paper proceeds as follows. Section 2 is the related literature and Section 3 presents the model. Sections 4–7 describe equilibrium and theoretical properties, the computation and algorithmic play, simulation design, Calibrated results and further work.

2 Related Literature

This paper connects five strands of work: dynamic vehicle demand and industry dynamics, network externalities and technology adoption, electric vehicle policy design, consumer valuation of vehicle operating costs, and reinforcement learning as an equilibrium selection device.

Dynamic demand and industry dynamics. The household decision problem builds on dynamic discrete choice models of durable goods replacement, beginning with Rust (1987). Because the empirical object is vehicle replacement, the paper is also closely related to structural work on automobile replacement and durable goods adjustment, especially Schiraldi (2011). On the demand side, the differentiated product environment follows the tradition of empirical automobile market models such as Berry et al. (1995). On the supply side, the formulation is related to the dynamic oligopoly literature in which firms' state variables, expectations, and continuation values shape pricing, investment, and entry decisions (Ericson and Pakes, 1995; Pakes and McGuire, 1994; Aguirregabiria and Mira, 2007; Bajari et al., 2007; Fershtman and Pakes, 2012). Relative to this literature, the present paper does not estimate a full structural dynamic game. It uses a dynamic game environment to isolate a particular mechanism: how the interaction between consumer learning and infrastructure accumulation selects among multiple adoption equilibria.

Network externalities, lock in, and technology adoption. The central economic force is a network feedback between EV adoption and charging infrastructure. This mechanism is related to the classic literature on network externalities, compatibility, and standardization (Katz and Shapiro, 1985; Farrell and Saloner, 1985), as well as to work on increasing returns and path dependence in technology adoption (Arthur, 1989). In those models, adoption incentives depend on the expected size of the network or the installed base of complementary users. In our environment, the network is physical: a larger EV installed base raises the return to charger deployment, while greater charger density lowers the effective cost of EV ownership. This two way feedback generates the possibility of multiple stable steady states and makes equilibrium selection a first order economic object. The use of a low dimensional aggregate state is also related to approximate aggregation in heterogeneous agent economies, as in Krusell and Smith (1998), although our aggregate state variables are adoption and infrastructure rather than capital and wealth distributions.

Electric vehicles and infrastructure policy. A growing empirical literature studies EV adoption, charging infrastructure, and the relative effectiveness of vehicle and infrastructure subsidies. The closest empirical counterpart is Li et al. (2017), who document indirect network effects between EV purchases and charging station deployment and show that infrastructure subsidies can outperform equally costly purchase subsidies in some settings. Springel (2021) estimates a two sided market model using Norwegian registry data and shows that the ranking of station and consumer subsidies depends on the subsidy level and the strength of network feedbacks. Evidence from U.S. state level adoption also points to the joint importance of demand side incentives and charging infrastructure (Narassimhan and Johnson, 2018). Muehlegger and Rapson (2022) estimate subsidy pass through and demand responses for low and middle income households in California, while Xing et al. (2021) show that the emissions benefits of EVs depend on which gasoline vehicles they replace. Holland et al. (2016) emphasize that the environmental benefits of EVs vary substantially across locations because electricity generation mixes differ across regions. Our paper is complementary to this empirical literature. Rather than estimating the causal effect of a specific subsidy program, it treats the infrastructure externality as a dynamic state variable and compares rational expectations selection with experience based learning selection.

Vehicle operating costs and consumer valuation. The calibrated mechanism depends on the effective user cost of EV and ICE ownership. This connects the paper to work on how households value future operating costs when purchasing vehicles. Gillingham et al. (2021) provide evidence on consumer myopia in vehicle purchases, while related work studies how gasoline prices are capitalized into vehicle choices and used car markets. In our model, the operating cost wedge is endogenous because charging infrastructure changes the effective cost of EV use. The policy counterfactuals therefore distinguish price wedge instruments, such as purchase subsidies, from state transition instruments, such as infrastructure investment.

Learning, reinforcement learning, and equilibrium selection. The learning component is related first to the foundational literature on Q learning and stochastic approximation. Watkins and Dayan (1992) introduce Q learning, while Tsitsiklis (1994) and Jaakkola et al. (1994) establish convergence properties for stochastic iterative dynamic programming algorithms in stationary single agent Markov environments. Borkar (2008) and Benaïm et al. (2005) provide the stochastic approximation and differential inclusion tools that motivate the weaker learning limit interpretation used here. These results do not imply global convergence to Nash equilibrium in our environment, because each household learns in an endogenous aggregate state generated by the simultaneous adoption and learning behavior of other households. This limitation is central in the multi agent learning literature (Hu and Wellman, 2003; Shoham et al., 2007). We therefore use reinforcement learning as an equilibrium selection device rather than as a proof of global convergence.

The paper is also related to a recent economics literature that uses reinforcement learning to study market behavior. Calvano et al. (2020) show that Q learning pricing algorithms can sustain supracompetitive outcomes in repeated oligopoly pricing games. Klein (2021) studies Q learning under sequential pricing, and Asker et al. (2024) shows that market outcomes depend sharply on the information and updating protocols embedded in pricing algorithms. Our object differs from algorithmic collusion among firms. The central issue here is equilibrium selection in a network adoption economy with household heterogeneity, endogenous infrastructure, and multiple stable aggregate states.

3 Model

We develop a dynamic discrete choice model of vehicle replacement with four key features: (i) heterogeneous agents differentiated by income and vehicle

age, (ii) network externalities in EV adoption through charging infrastructure, (iii) strategic complementarities arising from endogenous infrastructure investment, and (iv) oligopolistic manufacturers who set prices strategically in response to adoption dynamics and infrastructure availability. We characterize Markov Perfect Equilibria theoretically and compute them using multi-agent reinforcement learning.

3.1 Environment and Agents

Time and Agents. Time is discrete, indexed by $t = 0, 1, 2, \ldots$. There is a unit mass of infinitely lived agents. Each agent i has time invariant income $y_i \in \mathcal{Y} \equiv \{y_1, \ldots, y_J\}$ and, in each period t, owns a vehicle characterized by its age $a_i^t \in \mathcal{A} \equiv \{1, 2, \ldots, \bar{a}\}$ and fuel type $\tau_i^t \in \{\text{ICE}, \text{EV}\}$, where ICE denotes internal combustion engine. The individual state is $\omega_i^t = (y_i, a_i^t, \tau_i^t)$.

Aggregate State. The aggregate state in period t is given by $\Omega^t = (\pi^t, \eta^t)$, where $\pi^t \in [0, 1]$ is the EV adoption rate (share of agents owning EVs) and $\eta^t \in \mathbb{R}_+$ is the charging station density (stations per capita). We denote by μ^t the distribution of individual states $(\omega_i^t)_{i \in [0,1]}$ in period t.

3.2 Preferences and Technology

Instantaneous Utility. An agent with income y who drives a vehicle of age a and type τ in aggregate state $\Omega = (\pi, \eta)$ receives flow utility

$$u(x, s_a^\tau) = \frac{1}{1-\sigma} \left(x^{1-\rho} (s_a^\tau)^\rho \right)^{1-\sigma}, \qquad x = y - m_a^\tau(\eta), \tag{1}$$

where x is residual income after annual mobility cost, s_a^τ is the service flow, and $\sigma, \rho \in (0, 1)$ are preference parameters. The nested-CES specification allows for non-homotheticity in income and separates risk aversion from the elasticity of substitution. Choice level idiosyncratic noise is captured by Type-I Extreme Value taste shocks ε_d introduced below; the model contains no systematic intrinsic preference for either fuel type.

Annual Mobility Cost. The total annual mobility cost of operating a vehicle of age a and type τ at infrastructure level η is

$$m_a^\tau(\eta) \;=\; \underbrace{c_a^\tau(\eta)}_{\text{operating cost}} \;+\; \underbrace{\kappa^\tau}_{\text{capital user cost}}, \tag{2}$$

combining a flow operating cost $c_a^\tau(\eta)$ (fuel/electricity, maintenance, charging time) with a capital user cost κ^τ paying for the vehicle stock itself.

Service Flows. The service flow from a vehicle of age a and type τ is

$$s_a^\tau = s_0^\tau (1 - \delta_\tau)^{a-1}, \quad \tau \in \{\text{ICE}, \text{EV}\}, \tag{3}$$

where s_0^τ is the initial quality and $\delta_\tau \in (0, 1)$ is the depreciation rate. We allow for $s_0^{\text{EV}} \neq s_0^{\text{ICE}}$ and $\delta_{\text{EV}} \neq \delta_{\text{ICE}}$ to capture quality and durability differences.

Operating Costs. Operating costs depend on vehicle characteristics and, for EVs, on aggregate infrastructure:

$$c_a^{\text{ICE}} = \bar{c}^{\text{ICE}} + \alpha_{\text{ICE}} \cdot a, \tag{4}$$

$$c^{\text{EV}}(\eta^t) = \bar{c}^{\text{EV}} - \kappa \log(1 + \beta_\eta \eta^t), \tag{5}$$

where $\bar{c}^\tau$ is the base operating cost, $\alpha_{\text{ICE}} > 0$ captures increasing maintenance costs for aging ICE vehicles, and the function $-\kappa \log(1 + \beta_\eta \eta^t)$ captures the convenience gain from denser charging infrastructure. The parameters (κ, β_η) govern the strength of network externalities. (We use κ without a superscript for the network externality strength parameter; the capital user cost κ^τ defined below carries an explicit superscript.)

Capital User Cost (Hall Jorgenson). Because y_i is annual flow disposable income while vehicle prices $p^\tau(\Omega^t)$ are an order of magnitude larger, vehicle ownership cannot enter utility as a single period stock deduction. Instead, we treat the vehicle as a capital asset and assign it a Hall Jorgenson rental rate, the foregone interest plus economic depreciation per year of ownership:

$$\kappa^\tau(\Omega^t) \;=\; (r + \delta_v)\, p^\tau(\Omega^t), \tag{6}$$

where r is the real interest rate (foregone return on the funds tied up in the asset) and δ_v is the rate of capital depreciation of the vehicle. This is the standard durable goods formulation used in the structural literature (Berry et al. (1995); Schiraldi (2011); Gillingham et al. (2021)). With our calibration ($r = 0.05$, $\delta_v = 0.10$, $p^{\text{ICE}} = \$45{,}000$, $p^{\text{EV}} = \$55{,}000$) the capital user cost is $\kappa^{\text{ICE}} = \$6{,}750/\text{yr}$ and $\kappa^{\text{EV}} = \$8{,}250/\text{yr}$, generating a calibrated $\$1{,}500/\text{yr}$ capital premium for EVs that reflects their higher purchase price. By construction $y_i - m_a^\tau(\eta) > 0$ for every household type and every equilibrium state, so the CES utility is well defined throughout the state space.

3.3 Decision Problem and Value Functions

Choice Set. In each period, an agent chooses from three actions:

$$\mathcal{D} = \{\text{Keep},\ \text{Replace-ICE},\ \text{Replace-EV}\}. \tag{7}$$

Keeping the current vehicle advances its age by one period (subject to mandatory scrappage at age $\bar{a}$). Replacement involves purchasing a new (age 1) vehicle of the chosen type at price $p^{\tau}(\Omega^t)$, which is set by oligopolistic manufacturers (Section 3.5).

Idiosyncratic Shocks. Following Rust (1987), we introduce i.i.d. choice specific taste shocks $\{\varepsilon_d\}_{d\in\mathcal{D}}$ drawn from a Type-I Extreme Value distribution with scale parameter λ. These shocks are private information and independent across agents and time.

Value Functions. Let $V(\omega, \Omega; \sigma)$ denote the ex-ante value function for an agent in individual state $\omega = (y, a, \tau)$ facing aggregate state Ω, when all agents follow strategies σ. The Bellman equation is

$$V(\omega, \Omega; \sigma) = \mathbb{E}_{\varepsilon}\left[\max_{d\in\mathcal{D}} \{v_d(\omega, \Omega; \sigma) + \varepsilon_d\}\right], \tag{8}$$

where $v_d(\omega, \Omega; \sigma)$ is the choice-specific value function excluding the taste shock:

$$v_{\text{Keep}}(\omega, \Omega; \sigma) = u(y - m_a^{\tau}(\eta),\ s_a^{\tau}) + \beta \mathbb{E}_{\Omega'|\Omega,\sigma} V\big((y, a{+}1, \tau), \Omega'; \sigma\big), \tag{9}$$

$$v_{\text{Replace-}\tau'}(\omega, \Omega; \sigma) = u\Big(y - m_1^{\tau'}(\eta),\ s_1^{\tau'}\Big) + \beta \mathbb{E}_{\Omega'|\Omega,\sigma} V\big((y, 1, \tau'), \Omega'; \sigma\big),$$
$$\tau' \in \{\text{ICE}, \text{EV}\}, \tag{10}$$

where $\beta \in (0, 1)$ is the discount factor and the expectations $\mathbb{E}_{\Omega'|\Omega,\sigma}$ are taken over the aggregate state transition induced by σ. The prices $p^{\tau'}(\Omega)$ are set by oligopolistic manufacturers as described in Section 3.5.

Choice Probabilities. The Type-I EV assumption yields the standard logit formula:

$$\sigma_d(\omega, \Omega) = \frac{\exp\left(v_d(\omega, \Omega; \sigma)/\lambda\right)}{\sum_{d'\in\mathcal{D}} \exp\left(v_{d'}(\omega, \Omega; \sigma)/\lambda\right)}. \tag{11}$$

3.4 Aggregate Dynamics

Law of Motion for EV Adoption. Given the distribution of individual states μ^t and strategies σ, the share of vehicles that are EVs evolves as

$$\pi^{t+1} = \underbrace{(1 - \delta_s)\,\pi^t}_{\substack{\text{surviving}\\ \text{EVs}}} + \underbrace{\phi(\Omega^t; \sigma)\,(1 - \pi^t)}_{\substack{\text{ICE owners}\\ \text{switching to EV}}}, \tag{12}$$

where $\delta_s \in (0,1)$ is the exogenous vehicle scrappage rate and

$$\phi(\Omega^t;\sigma) \;:=\; \frac{1}{1-\pi^t}\int_{\omega:\tau=\text{ICE}} \sigma_{\text{Replace-EV}}(\omega,\Omega^t)\,d\mu^t(\omega) \tag{13}$$

is the conditional EV replacement rate among current ICE owners, aggregating the choice probabilities from the individual dynamic programming problem. The reduced form equation (12) treats EV→ICE voluntary replacement as a second order effect, which is consistent with our equilibrium output: EV owners choose Replace ICE with probability below 10^{-3} at all states visited in equilibrium.

Law of Motion for Infrastructure. Charging station density responds to adoption with economic depreciation:

$$\eta^{t+1} \;=\; \underbrace{(1-\delta_\eta)\,\eta^t}_{\text{depreciation}} \;+\; \underbrace{\bar{g}\,(\pi^t)^{\gamma_g}}_{\substack{\text{market-driven}\\ \text{investment}}} \;+\; \underbrace{I^t_{\text{gov}}}_{\substack{\text{government}\\ \text{investment}}}, \tag{14}$$

where $\delta_\eta \in (0,1)$ is the rate of economic depreciation (physical wear, technology obsolescence such as CHAdeMO→CCS→NACS connector transitions, and station closures), $\bar{g} > 0$ is the market driven investment scale, $\gamma_g > 1$ governs the convexity required for multiple equilibria (so that at low π the investment flow is small relative to depreciation, while at high π it dominates), and $I^t_{\text{gov}} \geq 0$ represents the NEVI-style government infrastructure investment. The depreciation term ensures that the high equilibrium is *barely stable* in our calibration, with spectral radius $\rho(J) \approx 1-\delta_\eta = 0.97$, so that transitions are economically persistent even under optimal coordination.

Aggregate Transition Map. Let $\mathcal{T}(\pi,\eta;\sigma) = (\mathcal{T}_1(\pi,\eta;\sigma),\,\mathcal{T}_2(\pi,\eta;\sigma))$ collect the two laws of motion (12)–(14). A steady state $(\pi^\star,\eta^\star)$ under strategy profile σ solves $(\pi^\star,\eta^\star) = \mathcal{T}(\pi^\star,\eta^\star;\sigma)$. Multiple steady states can coexist when the implicit map $f(\pi^\star)$ obtained by substituting $\eta^\star = [\bar{g}(\pi^\star)^{\gamma_g} + I_{\text{gov}}]/\delta_\eta$ into the household best response crosses the 45 degree line more than once; Proposition 2 gives sufficient conditions.

3.5 Oligopolistic Manufacturer Pricing

Market Structure. There are F manufacturers indexed by $f = 1,\ldots,F$, each producing vehicles of a single fuel type $\tau_f \in \{\text{ICE},\text{EV}\}$. Manufacturer f chooses price p^t_f in each period to maximize discounted profits, taking

the aggregate state Ω^t and rival strategies as given. The profit function of manufacturer f in period t is

$$\Pi_f^t = (p_f^t - mc_f) \cdot D_f(p^t, \Omega^t), \tag{15}$$

where mc_f is the constant marginal cost and $D_f(p^t, \Omega^t)$ is the demand for firm f's vehicle, aggregated from household choice probabilities:

$$D_f(p^t, \Omega^t) = \int \sigma_{\text{Replace-}\tau_f}(\omega, \Omega^t; p^t)\, d\mu^t(\omega). \tag{16}$$

Dynamic Pricing Problem. Forward looking manufacturers recognize that current prices affect adoption, which influences future infrastructure and hence future demand. Manufacturer f solves

$$\max_{\{p_f^t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta_F^t \, \Pi_f^t(\boldsymbol{p}^t, \Omega^t), \quad \text{s.t. } \Omega^{t+1} = \mathcal{T}(\Omega^t; \sigma(\boldsymbol{p}^t)), \tag{17}$$

where $\beta_F \in (0,1)$ is the manufacturer's discount factor and $\boldsymbol{p}^t = (p_1^t, \ldots, p_F^t)$ is the price vector. The first order condition for EV manufacturer f yields the mark up rule

$$\underbrace{p_f^t - mc_f}_{\text{static mark-up}} = \underbrace{-\frac{D_f}{\partial D_f / \partial p_f}}_{\text{inverse semi-elasticity}} + \underbrace{\beta_F \frac{\partial \Pi_f^{t+1}}{\partial \Omega^{t+1}} \cdot \frac{\partial \Omega^{t+1}}{\partial D_f} \cdot \left(-\frac{\partial D_f}{\partial p_f} \right)^{-1}}_{\text{dynamic investment motive}}, \tag{18}$$

where the last term captures the incentive to lower prices today in order to build the installed base and infrastructure. This *dynamic investment motive* is negative for EV manufacturers (they price below the static mark up) and vanishes for ICE manufacturers who do not benefit from infrastructure externalities.

Equilibrium Pricing. In the Markov Perfect Equilibrium of the pricing game, each manufacturer's pricing rule is a function $p_f^*(\Omega)$ of the aggregate state. The aggregate demand system inherits the logit structure from household choice probabilities, ensuring that the pricing game has a unique Nash equilibrium in prices for each Ω (Caplin and Nalebuff, 1991). Prices enter the household's value function through the choice specific values, closing the loop between demand and supply.

When $F \to \infty$ within each fuel type, the mark up vanishes and prices converge to marginal cost: $p^\tau(\Omega) \to mc^\tau + f$, where f is a per unit tax or

subsidy. This competitive benchmark nests the standard Rust (1987) framework and serves as a comparison for quantifying the effect of market power on adoption dynamics. In our calibration, the dynamic investment motive reduces the EV mark up by 18% relative to the static oligopoly benchmark, accelerating adoption by approximately 3 years.

4 Equilibrium and Theoretical Properties

4.1 Equilibrium Concept

Definition 1 (Markov Perfect Equilibrium (MPE))**.** *A Markov Perfect Equilibrium is a tuple* $(\sigma^{H,*}, p^*, V^{H,*}, \{V_f^*\}_{f=1}^F, \mathcal{T}^*)$ *such that:*

1. ***Household optimality:*** *For all household states* ω *and aggregate states* Ω*,* $\sigma^{H,*}(\omega, \Omega)$ *solves the household dynamic program given the pricing rule* $p^*(\Omega)$ *and aggregate law* $\mathcal{T}^*$*.*

2. ***Firm optimality:*** *For each manufacturer* f*,* $p_f^*(\Omega)$ *solves the firm's Markov pricing problem given rival pricing rules* $p_{-f}^*(\Omega)$*, household demand induced by* $\sigma^{H,*}$*, and the aggregate law* $\mathcal{T}^*$*.*

3. ***Value consistency:*** $V^{H,*}$ *and* $\{V_f^*\}_{f=1}^F$ *satisfy their respective Bellman equations under* $(\sigma^{H,*}, p^*, \mathcal{T}^*)$*.*

4. ***Aggregate consistency:*** $\mathcal{T}^*$ *is the law of motion generated by the household policy, the pricing rule, infrastructure investment, depreciation, and the cross sectional distribution of household states.*

5. ***Rational expectations:*** *Household and firm expectations over future aggregate states coincide with* $\mathcal{T}^*$*.*

Because agents are anonymous and interact only through the low dimensional aggregate state $\Omega = (\pi, \eta)$, each agent's decision problem reduces to a single agent dynamic program conditional on the aggregate transition law, despite the underlying population being a continuum of heterogeneous agents.

4.2 Existence, Multiplicity, and Stability

We establish a set of deliberately conservative theoretical results: existence (Proposition 1), multiplicity under explicit crossing conditions (Proposition 2),

local stability (Proposition 3), an exploration floor that can move the economy across a tipping region (Proposition 4), a stochastic approximation interpretation of two time scale learning (Proposition 5), and a diagnostic decomposition of the learning rational expectations gap (Proposition 6). Complete proofs are in Appendix A.

Assumption 1 (Regularity Conditions)**.** *(i)* ***Bounded Primitives:*** *$\mathcal{Y}, \mathcal{A}$ are finite sets; $c_a^\tau(\eta)$, κ^τ, s_a^τ, $p^\tau(\Omega)$ are bounded for all (a, τ, η, Ω), so that the total mobility cost $m_a^\tau(\eta) = c_a^\tau(\eta) + \kappa^\tau$ is bounded.*

(ii) ***Continuity:*** *The functions $c^{EV}(\eta), g(\pi), h(I)$ are continuous.*

(iii) ***Discount Factor:*** *$\beta \in (0, 1)$.*

(iv) ***Monotonicity:*** *$\partial c^{EV}(\eta)/\partial \eta < 0$ (better infrastructure lowers EV costs).*

(v) ***Positive residual income:*** *$y_i - m_a^\tau(\eta) > 0$ for all household types, vehicle ages, and reachable infrastructure levels, ensuring the CES utility is well defined throughout the state space.*

Proposition 1 (Existence)**.** *Under Assumption 1, a Markov Perfect Equilibrium exists.*

Proof Sketch. For any fixed aggregate strategy profile $\bar{\sigma}$, the individual optimization problem is a contraction mapping (Blackwell's conditions) since $\beta < 1$ and payoffs are bounded, yielding a unique best response $BR(\bar{\sigma})$. The mapping $\bar{\sigma} \mapsto BR(\bar{\sigma})$ is continuous; Brouwer's fixed point theorem delivers $\sigma^* = BR(\sigma^*)$. See Appendix A.1, A.2. □ □

Proposition 2 (Strategic Complementarities and Multiplicity)**.** *Let $H(\pi) = [\bar{g}\pi^{\gamma_g} + I_{\text{gov}}]/\delta_\eta$ denote the steady state infrastructure induced by adoption, and define the reduced steady state adoption map*

$$\Psi(\pi) \equiv \mathcal{T}_1(\pi, H(\pi); \sigma^*), \qquad f(\pi) \equiv \Psi(\pi) - \pi.$$

Suppose Ψ is continuous, $g'(\pi) > 0$, $\partial c^{\text{EV}}/\partial \eta < 0$, and there exist $0 < \pi_1 < \pi_2 < 1$ such that

$$f(0) > 0, \qquad f(\pi_1) < 0, \qquad f(\pi_2) > 0, \qquad f(1) < 0.$$

Then the model has at least three steady states: a low adoption steady state $\pi_L \in (0, \pi_1)$, an intermediate steady state $\pi_M \in (\pi_1, \pi_2)$, and a high adoption steady state $\pi_H \in (\pi_2, 1)$, with $\eta_k = H(\pi_k)$ for $k \in \{L, M, H\}$. If, in addition, $|\Psi'(\pi_L)| < 1$, $\Psi'(\pi_M) > 1$, and $|\Psi'(\pi_H)| < 1$, then the low and high steady states are locally stable and the intermediate steady state is locally unstable.

Proof Sketch. Strategic complementarity follows from $g'(\pi) > 0$ and $\partial c^{\text{EV}}/\partial \eta < 0$: higher adoption raises future infrastructure, which lowers EV operating costs and increases the value of EV ownership. The four crossing inequalities imply, by the intermediate value theorem, that f has one zero in each interval $(0, \pi_1)$, (π_1, π_2), and $(\pi_2, 1)$. The derivative restrictions give the stated local stability classification. See Appendix A.3. □ □

Proposition 3 (Local Stability)**.** *Let (π^*, η^*) be a steady state equilibrium with Jacobian $J(\pi^*, \eta^*)$. If the spectral radius $\rho(J) < 1$, then the equilibrium is locally stable.*

Proof Sketch. Standard linearization: $x^{t+1} \approx Jx^t$ for deviations $x^t = (\pi^t - \pi^*, \eta^t - \eta^*)$. The condition $\rho(J) < 1$ ensures $\|x^t\| \to 0$ exponentially. See Appendix A.4. □ □

4.3 Equilibrium Selection via Learning Dynamics

Propositions 2–3 show that the model can contain multiple locally stable steady states. They do not determine which steady state is selected. We use learning dynamics to discipline this selection problem.

Definition 2 (Basin of Attraction)**.** *For a stable equilibrium $\Omega^* = (\pi^*, \eta^*)$ under policy σ, the basin of attraction $\mathcal{B}(\Omega^*; \sigma) \subset [0, 1] \times \mathbb{R}_+$ is the set of initial conditions Ω_0 from which the deterministic dynamics $\Omega_{t+1} = \mathcal{T}(\Omega_t; \sigma)$ converge to Ω^*.*

Proposition 4 (Exploration Floor and Selection)**.** *Consider ϵ-greedy algorithmic play with action set $\mathcal{D}$ and exploration rate $\epsilon \in [0, 1)$. For any underlying Q function, the adoption law satisfies*

$$\mathcal{T}_1^{\epsilon}(\pi, \eta) \geq (1 - \delta_s)\pi + \frac{\epsilon}{|\mathcal{D}|}(1 - \pi). \tag{19}$$

Consequently, any invariant adoption rate under the induced policy satisfies

$$\pi \geq \pi_{\min}(\epsilon) \equiv \frac{\epsilon}{\epsilon + |\mathcal{D}|\delta_s}, \tag{20}$$

which is strictly increasing in ϵ. If there exists a tipping threshold $\bar{\pi}$ such that all states with $\pi \geq \bar{\pi}$ lie in the basin of the high adoption steady state under the post exploration policy, then any ϵ satisfying $\pi_{\min}(\epsilon) \geq \bar{\pi}$ moves the long run state into the high adoption basin. In the calibrated model, higher early exploration expands the measured high adoption basin, but this basin expansion is a quantitative result rather than a general monotonicity theorem.

Proof Sketch. Under ϵ-greedy play, every action is chosen with probability at least $\epsilon/|\mathcal{D}|$. In particular, each current ICE owner selects Replace EV with probability at least $\epsilon/|\mathcal{D}|$, which gives the lower bound on $\mathcal{T}_1^\epsilon$. Solving the resulting scalar inequality at an invariant adoption rate gives $\pi_{\min}(\epsilon)$. The final statement follows directly from the definition of the basin of attraction. See Appendix A.5 for details. □ □

Exploration should be interpreted as an early adoption or experimentation channel: test drives, leasing programs, information spillovers, or experimentation by consumers who do not yet know the private value of EV ownership. The proposition does not say that more random behavior is always welfare improving. It says that experimentation creates a lower bound on EV adoption; if that lower bound is high enough to cross the tipping region, complementarities can sustain the high adoption path after exploration fades.

4.4 Convergence of Learning to Equilibrium

The convergence claim is intentionally weaker than a global convergence theorem for Q learning in a non stationary dynamic game. The result below is a stochastic approximation statement: learning tracks a limiting ordinary differential equation, and convergence to an MPE obtains only under additional local stability and basin conditions.

Proposition 5 (Two Time Scale Learning Dynamics)**.** *Consider the two time scale Q learning scheme described in Section 5. The Q values are updated on the fast time scale with learning rates α_t satisfying $\sum_t \alpha_t = \infty$ and $\sum_t \alpha_t^2 < \infty$. The aggregate transition kernel $\mathcal{P}$ is updated on the slow time scale with rate γ_t satisfying $\sum_t \gamma_t = \infty$, $\sum_t \gamma_t^2 < \infty$, and $\gamma_t/\alpha_t \to 0$. Assume finite state and action spaces, bounded rewards, GLIE exploration for the theoretical learning process, and bounded simulation noise. Then the interpolated learning process tracks the stochastic approximation dynamics*

$$\dot{\mathcal{P}} = \Phi(BR(\mathcal{P})) - \mathcal{P}, \tag{21}$$

where $BR(\mathcal{P})$ is the household best response policy for a fixed aggregate law and Φ maps a policy into its induced aggregate law. The limit set of the slow time scale process is contained in the internally chain transitive set of this ODE. If the ODE has an isolated locally asymptotically stable fixed point $\mathcal{P}^*$ whose basin contains the initial aggregate law estimate, then the learning process converges locally to $\mathcal{P}^*$, and the associated policy is an MPE policy for the discretized economy.

Proof Sketch. For fixed $\mathcal{P}$, each household type faces a stationary discounted MDP, so the Bellman operator is a contraction and standard Q learning converges under Robbins Monro and GLIE conditions. Since $\gamma_t/\alpha_t \to 0$, the aggregate law is approximately fixed on the fast time scale. On the slow time scale, the Q values are close to the best response to the current aggregate law, so the aggregate update follows the ODE displayed above. The stochastic approximation theorem gives convergence of the interpolated process to an internally chain transitive invariant set. If the relevant fixed point is isolated and locally asymptotically stable, the invariant set reduces locally to that fixed point. See Appendix A.6 for details. □ □

This proposition is the formal basis for treating AMP as an equilibrium selection procedure. It does not rule out convergence to different locally stable equilibria from different initial aggregate beliefs, nor does it establish global convergence in every calibration. For that reason, the quantitative section reports equilibrium diagnostics rather than relying on the theorem alone.

4.5 Opening the Black Box: Learning Path vs. Rational Expectations

The final theoretical result characterizes the systematic discrepancy between the transition path under rational expectations and the learning based path, which we call the *implementation lag*.

Definition 3 (Implementation Lag)**.** *For initial condition Ω_0 and target adoption level $\bar{\pi}$, define the hitting times*

$$T^{RE}(\bar{\pi}) = \inf\{t : \pi_t^{RE} \geq \bar{\pi}\}, \qquad T^{AMP}(\bar{\pi}) = \inf\{t : \pi_t^{AMP} \geq \bar{\pi}\}. \tag{22}$$

The implementation lag is $L(\bar{\pi}) = T^{AMP}(\bar{\pi}) - T^{RE}(\bar{\pi})$.

Proposition 6 (Learning Rational Expectations Gap)**.** *Fix an initial state Ω_0, a rational expectations branch, and an AMP trajectory. For a target adoption level $\bar{\pi}$, define $L(\bar{\pi}) = T^{\mathrm{AMP}}(\bar{\pi}) - T^{\mathrm{RE}}(\bar{\pi})$. Then:*

(a) If $\pi_t^{\mathrm{RE}} \geq \pi_t^{\mathrm{AMP}}$ for all t before the target is reached, then $L(\bar{\pi}) \geq 0$.

(b) If AMP reaches the target and its adoption path has a per period growth lower bound $\underline{a} > 0$ in the relevant neighborhood, then

$$L(\bar{\pi}) \leq \left\lceil \frac{\sup_{t \leq T^{\mathrm{AMP}}} |\pi_t^{\mathrm{RE}} - \pi_t^{\mathrm{AMP}}|}{\underline{a}} \right\rceil.$$

(c) *The gap can be decomposed mechanically into an exploration component and an adaptation component by partitioning dates according to whether the exploration probability is above a chosen threshold and whether the Bellman error exceeds a chosen tolerance.*

In the baseline calibration, the RE branch that selects the high adoption equilibrium reaches the high state while AMP remains in the low basin without infrastructure intervention. The reported welfare difference and lag decomposition are therefore calibrated diagnostics, not general comparative statics theorems.

Proof Sketch. Part (a) follows immediately from the definition of hitting times. Part (b) follows because the maximum path distance divided by the minimum local growth rate bounds the number of additional periods required for AMP to close the adoption shortfall. Part (c) is an accounting identity once the date partition is fixed. See Appendix A.7 for details. □ □

The learning gap matters because the RE benchmark selects a branch by assumption, whereas AMP selects a branch through realized experience. Policies that reduce private EV costs operate on the price wedge; policies that raise charging density operate on the state variable governing the network externality. The calibrated exercise below studies which margin is quantitatively more important.

5 Computation and Algorithmic Play

5.1 Computational Framework

Computing equilibrium in our heterogeneous agent model with network externalities requires solving a high dimensional fixed point problem. With J income levels, $\bar{a}$ vehicle ages, and binary fuel type, the individual state space contains $2J\bar{a}$ elements, while the aggregate state (π^t, η^t) evolves endogenously through the population distribution μ^t. Strategic complementarities from infrastructure externalities can generate multiple equilibria, necessitating global solution methods that can locate and characterize all fixed points.

We adopt a dual approach. First, a *dynamic programming benchmark* computes the rational expectations MPE via nested aggregate consistency iteration, providing the equilibrium against which we evaluate behavioral deviations. Second, we characterize the stochastic process of *algorithmic Markov play* (AMP) generated by Q learning and deep Q learning algorithms.

Following Calvano et al. (2020), we treat these not merely as computational tools but as behaviorally plausible learning rules that may themselves represent how decentralized markets coordinate, consistent with the bounded rationality literature (Ellison and Fudenberg, 1995; Young, 1993).

Definition 4 (Algorithmic Markov Play)**.** *Let $\mathcal{A}$ denote a learning algorithm with hyperparameters $\theta = (\alpha_t, \epsilon_t, \mathcal{F})$ comprising learning rates, exploration schedule, and function class. Algorithmic Markov Play (AMP) is the stochastic process $\{(\pi_t^{\mathcal{A}}, \sigma_t^{\mathcal{A}})\}_{t=0}^{\infty}$ generated by:*

$$\sigma_{t+1}^{\mathcal{A}} = \mathcal{U}(\sigma_t^{\mathcal{A}}, \{(s_\tau, a_\tau, r_\tau, s_{\tau+1})\}_{\tau \le t}; \theta) \tag{23}$$

$$\pi_{t+1}^{\mathcal{A}} = \mathcal{T}(\pi_t^{\mathcal{A}}, \eta_t^{\mathcal{A}}; \sigma_t^{\mathcal{A}}) \tag{24}$$

where $\mathcal{U}$ is the algorithm's update rule and $\mathcal{T}$ is the aggregate transition mapping.

5.2 Dynamic Programming Benchmark

We compute a benchmark *Markov Perfect Equilibrium* via a nested fixed point algorithm on a discretized aggregate grid $\mathcal{G} = \{(\pi_k, \eta_\ell) : k, \ell = 1, \ldots, 50\}$.

For a conjectured aggregate transition operator $\mathcal{P} : \mathcal{G} \to \Delta(\mathcal{G})$, we solve the household's dynamic program given private state ω and aggregate state $\Omega \in \mathcal{G}$.

Let choices be $d \in \mathcal{D}$ and assume i.i.d. type-I extreme value taste shocks with scale $\lambda > 0$. The choice specific value is

$$v_d(\omega, \Omega) \;\equiv\; u_d(\omega, \Omega) \;+\; \beta\, \mathbb{E}[V(\omega', \Omega') \mid \omega, \Omega, d]\,, \qquad \Omega' \sim \mathcal{P}(\cdot \mid \Omega). \tag{25}$$

The value function satisfies the log-sum-exp formula

$$V(\omega, \Omega) \;=\; \lambda \log\left(\sum_{d \in \mathcal{D}} \exp\big[v_d(\omega, \Omega)/\lambda\big]\right) + \lambda\gamma, \qquad \gamma \approx 0.577 \text{ (Euler Mascheroni)}. \tag{26}$$

The induced policy has the multinomial logit form

$$\sigma_d^*(\omega, \Omega) \;=\; \frac{\exp\big(v_d(\omega, \Omega)/\lambda\big)}{\sum_{d' \in \mathcal{D}} \exp\big(v_{d'}(\omega, \Omega)/\lambda\big)}. \tag{27}$$

Inner loop (DP). We solve (25)–(26) by *modified policy iteration* (Howard) with $\kappa = 20$ evaluation sweeps per improvement step and tolerance $\|V^{(m+1)} - V^{(m)}\|_\infty < 10^{-8}$.

Outer loop (aggregate consistency). Given σ^*, the implied aggregate law of motion Φ on $\mathcal{G}$ is obtained by simulating $N = 10^5$ agents and estimating the induced transition kernel via Gaussian kernel density estimation (bandwidth $h = 0.02$, selected by Silverman's rule of thumb). We iterate on the outer fixed point until

$$\sup_{\Omega \in \mathcal{G}} \left\| \Phi^{(n+1)}(\Omega) - \Phi^{(n)}(\Omega) \right\|_1 \; < \; 10^{-4}, \tag{28}$$

at which point the ergodic distribution of Ω stabilizes. When strategic complementarities generate multiple equilibria, we initialize from high/low aggregate states and, if needed, use homotopy/relaxation to trace distinct fixed points.

5.3 Q Learning and Emergent Play

We model *algorithmic play* via tabular Q learning. Each agent of type (y, a, τ) learns action values on the joint state space $s_t = (\omega_{i,t}, \Omega^t) \in \mathcal{S}$, where $\omega_{i,t}$ is the private state and $\Omega^t \in \mathcal{G}$ is the aggregate state. The update is

$$Q_{y,a,\tau}(s_t, d_t) \leftarrow (1 - \alpha_t) Q_{y,a,\tau}(s_t, d_t) + \alpha_t \Big[r_t + \beta \max_{d' \in \mathcal{D}} Q_{y',a',\tau'}(s_{t+1}, d') \Big], \tag{29}$$

with $s_{t+1} = (\omega_{i,t+1}, \Omega^{t+1})$; the next private state (y', a', τ') follows the physical law of motion (e.g. aging and fuel switch upon replacement). For the theoretical convergence statement, we impose a GLIE schedule: $\alpha_t = \alpha_0/(1+\xi_\alpha t)$ with $\sum_t \alpha_t = \infty$ and $\sum_t \alpha_t^2 < \infty$, together with ϵ-greedy exploration $\epsilon_t = \epsilon_0/(1 + \xi_\epsilon t)$, which ensures $\epsilon_t \to 0$ while maintaining infinite exploration. In the finite horizon simulations we also report a slowly decaying exponential schedule; those runs are interpreted through the equilibrium diagnostics rather than through the asymptotic GLIE theorem.

The aggregate state evolves endogenously via the aggregate dynamics

$$\pi^{t+1} = (1 - \delta_s)\, \pi^t + \frac{1}{N} \sum_{i:\, \tau_i^t = \text{ICE}} \mathbf{1}\{d_i^t = \text{Replace-EV}\}, \tag{30}$$

$$\eta^{t+1} = (1 - \delta_\eta)\, \eta^t + g(\pi^t) + h(I^t) + \xi_t, \qquad \xi_t \sim \mathcal{N}(0, 0.01^2), \tag{31}$$

where $g(\cdot)$ captures market driven infrastructure investment responding to adoption, $h(I^t)$ represents government infrastructure investment, and the shock ξ_t prevents deterministic limit cycles. To cope with the non stationarity inherent in multi agent learning, we employ a *two-time-scale* scheme: within each outer iteration (approximately fixed aggregate law $\mathcal{P}$), agents update (29) on a fast time scale; the aggregate transition kernel is updated only between episodes on a slow time scale (cf. Sec. 5.2).

5.4 Deep Q Learning Architecture

We approximate Q by a neural network $Q_\theta : \mathbb{R}^{13} \times \mathcal{D} \to \mathbb{R}$ with state vector

$$x = \big[y,\, a,\, \mathbf{1}_{\mathrm{EV}},\, \pi,\, \eta,\, \pi^2,\, \eta^2,\, \pi\eta,\, \log(1+\eta),\, a \cdot \mathbf{1}_{\mathrm{ICE}},\, y/\bar{y},\, a/\bar{a},\, t/T\big]^\top, \quad (32)$$

which encodes private attributes, aggregates, and economically motivated nonlinearities (all features are standardized).

We use a feedforward network with hidden layers $[256, 128, 64]$, ReLU activations, and *LayerNorm* (preferred over BatchNorm in non-stationary RL); dropout $p = 0.2$ acts as regularization. Training uses experience replay ($|\mathcal{M}| = 10^6$, minibatch $B = 256$) and a target network Q_{θ^-}. Every $K = 100$ steps we apply Polyak averaging with rate $\tau = 0.005$:

$$\theta^- \leftarrow \tau\, \theta + (1-\tau)\, \theta^-. \quad (33)$$

We minimize the Huber loss

$$\mathcal{L}(\theta) = \mathbb{E}_{(s,a,r,s')\sim\mathcal{M}}\Big[\ell_\delta\Big(r + \beta \max_{a'} Q_{\theta^-}(s', a') - Q_\theta(s, a)\Big)\Big], \quad (34)$$

$$\ell_\delta(u) = \begin{cases} \tfrac{1}{2}u^2, & |u| \le \delta, \\ \delta\big(|u| - \tfrac{1}{2}\delta\big), & |u| > \delta, \end{cases} \quad (35)$$

optimized with Adam (learning rate 10^{-4}). We employ *Double DQN* targets to reduce overestimation bias:

$$\max_{a'} Q_{\theta^-}(s', a') \;\text{ is replaced by }\; Q_{\theta^-}\big(s',\, \arg\max_{a'} Q_\theta(s', a')\big).$$

When actions are occasionally infeasible (e.g. no EV inventory), we apply *action masking* so the argmax and policy respect feasibility.

5.5 Game Theoretic Diagnostics of Algorithmic Play

To assess whether algorithmic market play (AMP) constitutes economically meaningful behavior, we deploy six diagnostics that open the black box of learning and connect it to equilibrium concepts.

Distance to Rational Equilibrium We compare AMP to the benchmark MPE along policies, steady states, and transition paths. Let μ^{RE} denote the ergodic state distribution under the rational expectations benchmark, and $\pi^{\mathrm{AMP}}(\cdot \mid s)$, $\pi^{\mathrm{RE}}(\cdot \mid s)$ be the (stochastic) policies over $\mathcal{D}$.

Policy distance.

$$d_\Pi \;=\; \mathbb{E}_{s\sim\mu^{\mathrm{RE}}}\big[\tfrac{1}{2}\,\big\|\pi^{\mathrm{AMP}}(\cdot \mid s) - \pi^{\mathrm{RE}}(\cdot \mid s)\big\|_1\big], \quad (36)$$

i.e. total variation distance averaged over rational-expectations states.

Steady-state distance. Let $(\pi^\star, \eta^\star)_{\text{AMP}}$ and $(\pi^\star, \eta^\star)_{\text{RE}}$ be the stationary aggregates:

$$d_{\text{state}} \;=\; \|(\pi^\star, \eta^\star)_{\text{AMP}} - (\pi^\star, \eta^\star)_{\text{RE}}\|_2 \,. \tag{37}$$

Path distance. For a common initialization and horizon T,

$$d_{\text{path}} \;=\; \frac{1}{T}\sum_{t=1}^{T} \|(\pi_t, \eta_t)_{\text{AMP}} - (\pi_t, \eta_t)_{\text{RE}}\|_2 \,. \tag{38}$$

(Discounted or Wasserstein alternatives can be reported in robustness.)

Algorithm 1 Benchmark MPE, Algorithmic Play, and Diagnostics

Require: Grid $\mathcal{G}$ for $\Omega = (\pi, \eta)$; primitives $(u, c^\tau, s^\tau, p^\tau, g, h, \beta, \lambda)$; tolerances $\varepsilon_{\text{in}}, \varepsilon_{\text{out}}$; Howard sweeps κ; population size N.

(Outer init) Initialize aggregate kernel / law $\mathcal{P}^{(0)}$ (or $\mathcal{T}^{(0)}$) on $\mathcal{G}$; set $n \leftarrow 0$.

repeat ▷ Aggregate consistency outer fixed point

  (Inner DP) For each $\Omega \in \mathcal{G}$, solve the individual agent's DP: Initialize $V^{(0)}(\cdot, \Omega)$; repeat

  for $m = 0, 1, \ldots$ **do**

    Policy improvement: $\sigma^{(m+1)}(\cdot, \Omega) \leftarrow \arg\max_{d \in \mathcal{D}} v_d(\cdot, \Omega) + \varepsilon_d$

    κ-step policy evaluation (Howard): $V^{(m+1)}(\cdot, \Omega) \leftarrow \textsc{Evaluate}\big(\sigma^{(m+1)}, \mathcal{P}^{(n)}(\cdot \mid \Omega)\big)$ for κ sweeps

    if $\|V^{(m+1)} - V^{(m)}\|_\infty < \varepsilon_{\text{in}}$ **then break**

    end if

  end for

  (Implied policy) Record $\sigma^*(\cdot, \Omega) \leftarrow \sigma^{(m+1)}(\cdot, \Omega)$.

  (Law of motion) Simulate N agents under σ^* on $\mathcal{G}$ to obtain $\Phi^{(n+1)}$ (or estimate $\mathcal{P}^{(n+1)}$ via KDE).

  (Outer update) $n \leftarrow n + 1$.

until $\sup_{\Omega \in \mathcal{G}} \|\Phi^{(n)}(\Omega) - \Phi^{(n-1)}(\Omega)\|_1 < \varepsilon_{\text{out}}$

Output: Benchmark rational expectations policy $\pi^{\text{RE}} = \sigma^*$ and steady state $(\pi^\star, \eta^\star)_{\text{RE}}$.

(Algorithmic Play: Tabular Q learning / DQN)

for episodes $e = 1, \ldots, E$ **do**

  Fix (or slowly update) aggregate law $\mathcal{P}$ to ensure near stationarity (*two time scale*).

  for steps $t = 0, \ldots, T_e$ **do**

    Observe $s_t = (\omega_t, \Omega_t)$; pick $a_t \sim \epsilon$-greedy from Q (tabular) or Q_θ (DQN).

    Get reward r_t, next state s_{t+1}; store (s_t, a_t, r_t, s_{t+1}) in replay buffer $\mathcal{M}$ (DQN).

    Tabular: $Q \leftarrow (1 - \alpha_t)Q + \alpha_t\big[r_t + \beta \max_{a'} Q(s_{t+1}, a')\big]$.

    DQN: sample minibatch, minimize Huber loss; Double DQN target; Polyak update of θ^- every K steps.

  end for

  Update empirical aggregate map $\widehat{\Phi}$ from simulated play (optional); small step mix with Φ.

end for

Output: π^{AMP} from Q/Q_θ and induced steady state $(\pi^\star, \eta^\star)_{\text{AMP}}$.

(Diagnostics)

Compute policy TV distance d_Π, steady state distance d_{state}, path distance d_{path}.

Estimate Nash gap via best response to π^{AMP}; test ε-CCE inequalities on $\hat{P}_T$.

Map basins of attraction on a grid of Ω_0; estimate $J(\Omega^\star)$ by finite differences and $\rho(J)$.

Exploitability and Nash Gap Fix opponents at π_{-i}^{AMP}. Let $V_i^{\mathrm{AMP}}(s)$ be the state value for player i under π^{AMP} and $V_i^{\mathrm{BR}}(s)$ the value of a best response to π_{-i}^{AMP} (solved as a single agent DP with fixed aggregate law). The *Nash gap* is

$$\mathrm{Gap}_i \;=\; \mathbb{E}_{s\sim\mu^{\mathrm{AMP}}}\big[V_i^{\mathrm{BR}}(s) - V_i^{\mathrm{AMP}}(s)\big] \;\geq 0, \tag{39}$$

with $\mathrm{Gap}_i = 0$ iff π^{AMP} is a Nash equilibrium for i. Empirically we estimate the pointwise version $\mathrm{Gap}_i(s) = \max_a Q_i^{\mathrm{BR}}(s,a) - \sum_a \pi_i^{\mathrm{AMP}}(a \mid s) Q_i^{\mathrm{BR}}(s,a)$ and average over $s \sim \mu^{\mathrm{AMP}}$.

Coarse Correlated Equilibrium (CCE) Tests From the empirical joint play $\hat{P}_T(s,a)$ generated by AMP over T iterations, AMP induces an ε_T-CCE if, for each player i and deviation $d' \in \mathcal{D}$,

$$\sum_{s,a} \hat{P}_T(s,a)\left[u_i(s,a) - u_i(s,(d',a_{-i}))\right] \;\geq\; -\varepsilon_T, \qquad \varepsilon_T = O_p(T^{-1/2}). \tag{40}$$

This connects AMP to the learning in games literature (*e.g.*, Hart Mas Colell, 2000).

Stability and Equilibrium Selection Let $\mathcal{T}^\pi$ be the aggregate law of motion implied by policy π on the aggregate state $\Omega = (\pi, \eta)$. We map basins of attraction on a 20×20 grid of Ω_0; for each cell we simulate the system under π^{AMP} and π^{RE} and record the selected steady state. Local stability at a fixed point $\Omega^\star$ is characterized by

$$J(\Omega^\star) \;=\; \left.\frac{\partial \mathcal{T}^\pi}{\partial \Omega}\right|_{\Omega=\Omega^\star}, \qquad \rho(J(\Omega^\star)) < 1 \iff \text{locally stable.} \tag{41}$$

5.6 Dynamics of Play and Behavioral Patterns

Learning trajectories reveal coordination mechanisms. We record *switching points* where the greedy action changes, $s \mapsto \arg\max_a Q(s,a)$, and compare to myopic best responses and to policy iteration steps in the DP benchmark.

Replacement thresholds. Define the indifference set for Replace-EV vs. Keep:

$$\mathcal{I}(\omega,\eta) \;=\; \Big\{\, a \;:\; Q\big((\omega,a),\eta,\text{Replace-EV}\big) = Q\big((\omega,a),\eta,\text{Keep}\big) \,\Big\}. \tag{42}$$

State visitation heatmaps. For marginalized slices (e.g. fixing τ, y) we report

$$\mathcal{V}(a, \eta \mid \tau, y) \;=\; \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\{(a_t, \eta_t, \tau_t, y_t) = (a, \eta, \tau, y)\}\,, \tag{43}$$

displayed as 2D histograms. Higher early exploration increases the chance of visiting states close to the tipping region. In the calibrated basin map diagnostic, runs initialized at $(\pi_0, \eta_0) = (0.05, 0.5)$ remain in the low basin under the baseline schedule, while sufficiently aggressive early experimentation can move some paths toward the high adoption basin. This is a numerical selection diagnostic, not a global theorem about exploration.

5.7 Representation Learning and Economic Interpretability

For the deep Q network we probe economic coherence.

Occlusion based importance. Masking feature j (with a neutral baseline) yields

$$\mathrm{FI}_j \;=\; \mathbb{E}_{(s,a)}[|Q_\theta(s, a) - Q_\theta(s_{-j}, a)|]\,, \tag{44}$$

normalized to sum to 100%. In our baseline, infrastructure density, vehicle age, and income account for roughly 31%, 24%, and 18% of action variation, respectively.

Integrated gradients. With baseline s^{base} and greedy action $a^\star$,

$$\mathrm{IG}_j(s) \;=\; \big(s_j - s_j^{\text{base}}\big) \int_0^1 \frac{\partial Q_\theta\big(s^{\text{base}} + \alpha(s - s^{\text{base}}),\, a^\star\big)}{\partial s_j}\, d\alpha, \tag{45}$$

which confirms $\partial Q/\partial \eta > 0$ and $\partial Q/\partial a > 0$ as predicted by theory.

Policy elasticities. Let $\bar{\eta}(\tau)$ be the long run adoption under a per vehicle subsidy τ. We report

$$\mathcal{E}_\tau \;=\; \frac{\partial \log \bar{\eta}(\tau)}{\partial \tau} \times 10^3, \tag{46}$$

interpreted as "% change in long run adoption per \$1,000 subsidy". Empirically, $\mathcal{E}_\tau$ is attenuated under AMP relative to RE, reflecting *algorithmic momentum*: gradual adjustment in learned policies generates implementation lags absent under rational expectations.

5.8 Structure, Perturbations, and Rules

(i) Structure tests (signs, single crossing, complementarities). Economic theory implies monotone comparative statics. We test moment inequalities on the learned Q:

$$\Delta_{\eta}^{\mathrm{EV}}(s) := Q(s_{\eta}^{+}, \text{Replace-EV}) - Q(s_{\eta}^{-}, \text{Replace-EV}) \ \geq 0, \tag{47}$$

$$\Delta_{a}^{\mathrm{Keep}}(s) := Q(s_{a}^{+}, \text{Keep}) - Q(s_{a}^{-}, \text{Keep}) \ \leq 0, \tag{48}$$

where $s_{\eta}^{\pm}$ (resp. $s_{a}^{\pm}$) increases η (resp. a) by a small step holding other coordinates fixed. We report the share of states with violations and a Kolmogorov Smirnov type statistic. Complementarities are probed via discrete supermodularity tests:

$$\Delta_{\eta\pi}^{\mathrm{EV}}(s) := Q(s_{\eta,\pi}^{++}, \text{Replace-EV}) - Q(s_{\eta,\pi}^{+-}, \text{Replace-EV}) - Q(s_{\eta,\pi}^{-+}, \text{Replace-EV}) + Q(s_{\eta,\pi}^{--}, \text{Replace-EV}$$

(ii) Perturbation responses (elasticities and IRFs). We compute *policy elasticities* and impulse responses under AMP by small counterfactuals:

$$\mathcal{E}_{\tau}^{\mathrm{AMP}} := \left.\frac{\partial \log \bar{\eta}(\tau)}{\partial \tau}\right|_{\tau=0} \times 10^{3}, \qquad \text{and} \quad \mathrm{IRF}_{\eta}(h) := \mathbb{E}[\eta_{t+h} \mid \eta_t + \delta] - \mathbb{E}[\eta_{t+h} \mid \eta_t], \tag{49}$$

contrasted with RE. We also report *local implementation lag*: $L_{\epsilon} := \arg\min_{\ell \geq 0} \{ \, |\eta_{t+\ell}^{\mathrm{AMP}} - \eta_{t}^{\mathrm{RE},\delta}| \leq \varepsilon \, \}$.

(iii) Rule extraction via sparse surrogates and value decomposition. On state slices (fixed τ, y), we fit a depth-3 decision tree $\hat{\pi}_{\mathrm{tree}}$ to the greedy action with fidelity $\mathrm{Fid} = \Pr_{s \sim \hat{\mu}} \left[\arg\max_a Q(s,a) = \hat{\pi}_{\mathrm{tree}}(s) \right]$ and report extracted thresholds (age–infrastructure cutoffs). We also decompose Q into immediate utility and continuation:

$$Q(s,a) = u_a(s) + \beta \underbrace{\mathbb{E}[V(s') \mid s, a]}_{\text{continuation}}, \quad A(s,a) := Q(s,a) - \sum_{a'} \pi^{\mathrm{AMP}}(a' \mid s) \, Q(s,a'),$$

plotting $A(s, \text{Replace EV})$ over (a, η) to reveal where adoption gains come from. For DQN we complement IG/occlusion with *partial dependence* on (a, η) holding others at empirical quantiles.

5.9 Benchmark vs. Game: Paths, Metrics, and Tests

We compare the rational expectations benchmark (RE; Sec. 5.2) with algorithmic market play (AMP; Sec. 5.3) along four axes: (i) *policy*, (ii) *equilibrium (steady state)*, (iii) *transition path*, (iv) *welfare and elasticities*. We fix a common environment and align experiments as follows.

Experimental alignment (common random numbers). Let the aggregate initial condition be $\Omega_0 = (\pi_0, \eta_0)$ and the micro state distribution be μ^0. We draw one shock sequence $\{\xi_t\}_{t\geq 0}$ (for infrastructure or exogenous disturbances) and (if needed) idiosyncratic taste shocks $\{\varepsilon_{i,t}\}$ *once*, and feed the *same* realizations into both RE and AMP runs to eliminate Monte Carlo confounds. AMP is evaluated in two modes: *greedy* ($\epsilon = 0$) and *on policy* (the learned ϵ_t).

Policy distance. Total variation distance averaged over RE states:

$$d_{\Pi} = \mathbb{E}_{s\sim\mu^{\mathrm{RE}}}\left[\tfrac{1}{2}\left\|\pi^{\mathrm{AMP}}(\cdot \mid s) - \pi^{\mathrm{RE}}(\cdot \mid s)\right\|_1\right]. \tag{50}$$

Equilibrium (steady state) distance. Let $(\pi^\star, \eta^\star)_{\mathrm{RE}}$ and $(\pi^\star, \eta^\star)_{\mathrm{AMP}}$ be the stationary aggregates under each policy (same primitives):

$$d_{\mathrm{state}} = \left\|(\pi^\star, \eta^\star)_{\mathrm{AMP}} - (\pi^\star, \eta^\star)_{\mathrm{RE}}\right\|_2. \tag{51}$$

Transition path distances. Given paths $\{\Omega_t^{\mathrm{RE}}\}_{t=0}^T$ and $\{\Omega_t^{\mathrm{AMP}}\}_{t=0}^T$ with $\Omega_t = (\pi_t, \eta_t)$:

$$d_{\mathrm{path}}^{(L^2)} = \left(\frac{1}{T}\sum_{t=0}^{T}\left\|\Omega_t^{\mathrm{AMP}} - \Omega_t^{\mathrm{RE}}\right\|_2^2\right)^{1/2}, \quad d_{\mathrm{path}}^{(\beta)} = \sum_{t=0}^{\infty}\beta^t\left\|\Omega_t^{\mathrm{AMP}} - \Omega_t^{\mathrm{RE}}\right\|_2, \tag{52}$$

$$d_{\mathrm{path}}^{(\mathrm{DTW})} = \mathrm{DTW}\big(\{\Omega_t^{\mathrm{AMP}}\}, \{\Omega_t^{\mathrm{RE}}\}\big), \quad d_{\mathrm{path}}^{(\mathrm{Fr})} = \text{Fréchet}\big(\{\Omega_t^{\mathrm{AMP}}\}, \{\Omega_t^{\mathrm{RE}}\}\big). \tag{53}$$

The latter two adjust for *speed* differences by optimal reparametrization.

Value and welfare gaps. Let $W^\bullet$ be average discounted welfare (per period utility integrated over the stationary distribution):

$$\Delta W = W^{\mathrm{RE}} - W^{\mathrm{AMP}} = \sum_{t=0}^{\infty}\beta^t\,\mathbb{E}\Big[\,u\big(y_t - m_{a_t}^{\tau_t}(\eta_t),\, s_{a_t}^{\tau_t}\big)\;\Big|\;\mathrm{RE}\Big] - \sum_{t=0}^{\infty}\beta^t\,\mathbb{E}\Big[\,u\big(y_t - m_{a_t}^{\tau_t}(\eta_t),\, s_{a_t}^{\tau_t}\big)\;\Big|\;\mathrm{AMP} \tag{54}$$

We also report the *value gap* from the Q functions: $\Delta V = \mathbb{E}_{s\sim\mu^{\mathrm{RE}}}[V^{\mathrm{RE}}(s) - V^{\mathrm{AMP}}(s)]$.

Policy elasticities and IRFs. For a subsidy τ (per EV), denote long run adoption $\bar{\eta}(\tau)$. Define the elasticity and speed of adjustment under each regime:

$$\mathcal{E}_\tau^\bullet = \left.\frac{\partial \log \bar{\eta}(\tau)}{\partial \tau}\right|_{\tau=0} \times 10^3, \qquad \mathrm{IRF}_\eta^\bullet(h) = \mathbb{E}[\eta_{t+h} \mid \eta_t + \delta] - \mathbb{E}[\eta_{t+h} \mid \eta_t]\,. \tag{55}$$

Statistical inference. We compute B-replicate confidence bands using common random numbers across RE and AMP. Path-equality is tested via sup norm and Cramér von Mises type functionals on $\{\Omega_t\}$; we also report the Hausdorff distance between the two trajectory *sets*.

Algorithm 2 Benchmark vs Game Comparison Protocol

Require: Primitives; initial (μ^0, Ω_0); shock seeds; horizons $T_{\text{trans}}, T_{\text{burn}}$.

1: **Calibrate/solve RE:** Compute $\pi^{\text{RE}}, (\Omega_t^{\text{RE}})_{t=0}^{T}$, steady state $(\pi^\star, \eta^\star)_{\text{RE}}$.
2: **Train AMP:** Learn Q/Q_θ; evaluate in (i) greedy ϵ=0 and (ii) on policy.
3: **Simulate paths:** Using common shocks, produce $(\Omega_t^{\text{AMP}})_{t=0}^{T}$ from the same Ω_0.
4: **Compute metrics:** d_Π, d_{state}, $d_{\text{path}}^{(L^2)}, d_{\text{path}}^{(\beta)}, d_{\text{path}}^{(\text{DTW})}$, ΔV, ΔW, $\mathcal{E}_\tau^\bullet$, $\text{IRF}_\eta^\bullet$.
5: **Inference:** Bootstrap with common random numbers; report CIs and p-values.
6: **Visualization:**

- *Phase plane overlay:* plot $\{\Omega_t^{\text{RE}}\}$ and $\{\Omega_t^{\text{AMP}}\}$ with arrows (speed), and the 45° fixed point loci.
- *Time path panels:* π_t and η_t with 95% bands; inset DTW alignment map.
- *Policy heatmaps:* $\pi^\bullet(a, \eta)$ or greedy action regions with threshold contours.
- *Basin maps:* initial grid $\Omega_0 \to$ selected steady state (color coded) under each regime.

6 Implementation and Simulation Design

This section details the parameterization and computational setup for the algorithms described in Section 5.

6.1 Baseline Parameterization

Population and Initial Conditions. Our baseline market simulation uses $N = 20{,}000$ agents, with income $y_i \sim \text{LogNormal}(\mu_y, \sigma_y^2)$ calibrated to match the U.S. income distribution. Initially, the aggregate EV share is set to $\pi^0 = 0.092$, corresponding to the U.S. 2024 new sales reference point used in the calibration, and the infrastructure state is $\eta^0 = 0.5$. ICE vehicle ages are initially distributed on $\{1, \ldots, 10\}$. For basin map diagnostics we

also report a lower initial state, $(\pi_0, \eta_0) = (0.05, 0.5)$, to verify that the low basin is not an artifact of the 2024 starting point.

Structural Parameters. Table 1 lists the baseline values of all structural parameters, together with their sources. Preference parameters (σ, ρ) are jointly calibrated to match the empirical EV adoption elasticity with respect to operating cost reported by Muehlegger and Rapson (2023) and Xing, Leard and Li (2021). Operating cost parameters $(\bar{c}^{\text{ICE}}, \alpha_{\text{ICE}}, \bar{c}^{\text{EV}})$ come from AAA, ICCT, and JD Power. Purchase prices come from KBB 2024. The Hall Jorgenson inputs (r, δ_v) use standard values from FRED and BLS/NHTS. The infrastructure law of motion $(\bar{g}, \gamma_g, \delta_\eta)$ is calibrated to reproduce a low infrastructure trap, a high infrastructure steady state, and transition speeds consistent with public charger deployment magnitudes. A full estimation using AFDC charger density, registration data, and state policy variation is left for future work.

Table 1: Structural parameter values in the baseline calibration.

Parameter	Description	Value	Source
σ	Risk aversion (CES)	0.5	Calibrated
ρ	Consumption service substitution (CES)	0.3	Calibrated
$\bar{c}^{\text{ICE}}$	ICE base operating cost	\$3,600/yr	AAA (2024)
α_{ICE}	ICE age coefficient	\$200/yr/age	AAA (2024)
$\bar{c}^{\text{EV}}$	EV cost at $\eta = 0$	\$7,000/yr	ICCT / JD Power
κ	Network externality strength	2,200	Calibrated
β_η	Infrastructure sensitivity	2.0	AFDC data
λ	Logit scale (idiosyncratic)	\$5,500	Calibrated
p^{ICE}	ICE purchase price	\$45,000	KBB (2024)
p^{EV}	EV purchase price	\$55,000	KBB (2024 ATP)
r	Real interest rate (Hall Jorgenson)	0.05	FRED
δ_v	Vehicle capital depreciation rate	0.10	BLS / NHTS
κ^{ICE}	ICE capital user cost $(r+\delta_v)p^{\text{ICE}}$	\$6,750/yr	derived
κ^{EV}	EV capital user cost $(r+\delta_v)p^{\text{EV}}$	\$8,250/yr	derived
$\bar{g}$	Infrastructure investment rate	0.279	AFDC time series
γ_g	Infrastructure response convexity	1.3	AFDC fit
δ_η	Infrastructure depreciation rate	0.03	AFDC / station closure data
δ_s	Vehicle scrappage rate	0.05	NHTS 2022
β	Household discount factor	0.96	Standard
$mc^{\text{ICE}}/mc^{\text{EV}}$	Manufacturer marginal costs	\$28k / \$42k	BNEF (2024)

The implied Hall Jorgenson capital premium of EVs over ICE vehicles is $\kappa^{\text{EV}} - \kappa^{\text{ICE}} = \$1{,}500/\text{yr}$, reflecting the \$10,000 purchase price gap. Combined with the operating cost gap at the low equilibrium infrastructure level (approximately \$1,850/yr at $\eta_L^* = 0.14$), this yields a total mobility cost gap

of roughly \$3,350–\$3,360 per year, which is the key quantity that IRA style subsidies must overcome to flip the static argmax from ICE to EV.

State Discretization. Individual states $\omega \in \mathcal{Y} \times \mathcal{A} \times \{\text{ICE}, \text{EV}\}$ are finite by construction. We discretize the aggregate state (π, η) into a $G_\pi \times G_\eta$ grid with $G_\pi = G_\eta = 20$ in our baseline, so $\Omega \in \{\pi_1, \ldots, \pi_{G_\pi}\} \times \{\eta_1, \ldots, \eta_{G_\eta}\}$.

Hyperparameters. The tabular Q learning exercise uses $\epsilon_0 = 0.3$, $\alpha_0 = 0.1$, and slowly declining exploration and learning rate schedules. The baseline tabular results reported below average over 10 random seeds and 2 million iterations per seed. The DQN robustness exercise uses 8 random seeds and 500 training episodes. Convergence is assessed using the battery of diagnostics in Section 6.3: Q value stability, policy drift, Nash gap, monotonicity checks, and perturbation reversion tests.

6.2 Relationship to Equilibrium Theory

Under standard regularity conditions, Q learning converges to the optimal Q function in stationary single agent environments (Watkins and Dayan, 1992). In our multi agent setting, convergence is not theoretically guaranteed because the environment is non stationary from each agent's perspective other agents are learning simultaneously. However, three features of our model promote empirical convergence: (i) the aggregate state Ω compresses the strategic environment into a low dimensional sufficient statistic, (ii) the logit choice structure smooths best responses, dampening oscillatory dynamics, and (iii) the slowly decaying schedules for α_k and ϵ_k allow agents to adapt to one another gradually.

We do not use the learning algorithm as a proof of global convergence to Nash equilibrium. The algorithm is used as an equilibrium selection device. Its output is interpreted as an approximate Markov equilibrium only after it passes the diagnostics in Section 6.3: small Bellman error, stable policy drift, local aggregate stability, and a small one agent best response gain.

6.3 Convergence Diagnostics and Equilibrium Validation

A critical question is whether the strategies learned by Q learning constitute economically meaningful equilibria. We employ a diagnostic battery including Bellman error, policy stability, Nash gap (exploitability), perturbation reversion tests, absorbing state analysis, path distance measures, welfare

comparisons, and monotonicity checks. This section presents the two most informative diagnostics the Nash gap and perturbation reversion test which directly assess equilibrium quality and stability. The remaining diagnostics are reported in Section 7.1 and Appendix B.

Nash Gap (Exploitability) A Nash equilibrium requires that no agent can profitably deviate. For each agent type $\tau \equiv (y, a, \text{fuel type})$, we freeze opponents at their current strategies $\sigma^k_{-\tau}$, compute type τ's best response σ^*_τ by solving a single-agent dynamic program (taking the aggregate transition induced by $\sigma^k_{-\tau}$ as given), and measure the value gain: The value gain from deviating is

$$\Delta V_\tau = V_\tau(\sigma^*_\tau, \sigma^k_{-\tau}) - V_\tau(\sigma^k_\tau, \sigma^k_{-\tau}). \tag{56}$$

The Nash gap aggregates across types: $\text{NashGap}^k = \sum_\tau \max\{0, \Delta V_\tau\}$. We declare approximate Nash equilibrium when $\text{NashGap}^k < 0.01 \times \bar{V}$, where $\bar{V}$ is the average value across types. We approximate best responses by running Q learning for each type for 100,000 additional iterations with $\epsilon = 0$ (pure exploitation), treating the environment as stationary.

Perturbation-Reversion Test This test assesses local stability (Proposition 3). At a candidate equilibrium (π^*, η^*) with policies $\{\sigma^*_\tau\}_\tau$, we randomly select 10% of agents and override their decisions for 10 periods (forcing random actions), then resume normal play and track (π^t, η^t) for 500 periods. The reversion probability is

$$\mathbb{P}_{\text{revert}} = \frac{1}{50} \sum_{s=1}^{50} \mathbf{1}\left\{ \|(\pi^{T+s}, \eta^{T+s}) - (\pi^*, \eta^*)\| < \varepsilon_{\text{revert}} \right\}, \tag{57}$$

with $T = 500$ and $\varepsilon_{\text{revert}} = 0.05$. An equilibrium is deemed locally stable if $\mathbb{P}_{\text{revert}} \geq 0.8$ across 50 independent perturbation experiments. This diagnostic complements the spectral radius condition of Proposition 3 by assessing stability over finite (rather than infinitesimal) perturbations.

Additional Diagnostics Four additional diagnostics are detailed in Appendix B: the Bellman error (measuring how well the learned Q function satisfies the optimality equation), policy stability (tracking drift in the greedy policy over iterations), absorbing state analysis (mapping basins of attraction and computing hitting times to steady state equilibria), and monotonicity checks implied by the economic structure of the model. In the baseline run, the final Bellman error statistic is 18.1 in payoff units, policy drift is effectively zero, the relative Nash gap is 0.03% of average value, and the monotonicity violation rates are below 10% for the infrastructure and supermodularity checks and zero for the age check. These diagnostics support interpreting the learned strategy profile as a near equilibrium selection rather than as a failed numerical optimization.

7 Calibrated Results

7.1 Equilibrium Selection in the Calibrated Economy

The numerical analysis is organized around four objects: the steady state structure of the calibrated economy, the transition paths selected by rational expectations and AMP, the equilibrium diagnostics for the learned outcome, and the robustness of the mechanism to alternative parameter choices. These exercises are calibrated rather than structurally estimated.

Multiple Steady States. The calibrated economy has three steady states. The economically important object is the unstable middle state, which separates the low and high adoption basins. Table 2 reports the steady states and local stability statistics. Figure 1 shows the induced adoption map. The low and high intersections are locally stable, while the middle intersection is a tipping point.

Table 2: Baseline equilibria in the calibrated economy

	π^*	η^*	$\widehat{\rho}(J)$	$m^{\mathrm{EV}} - m^{\mathrm{ICE}}$	Type
Low trap	0.040	0.14	0.83	+\$3,350–+\$3,360	Stable
Tipping point	0.180	1.00	1.04	+\$1,480	Unstable
High adoption	0.700	5.84	0.97	−\$1,690	Stable

Notes: $\widehat{\rho}(J)$ is the estimated spectral radius of the aggregate transition Jacobian. The mobility cost gap includes operating costs and Hall Jorgenson capital user costs. Positive values mean EVs are more expensive than ICE vehicles at that infrastructure state.

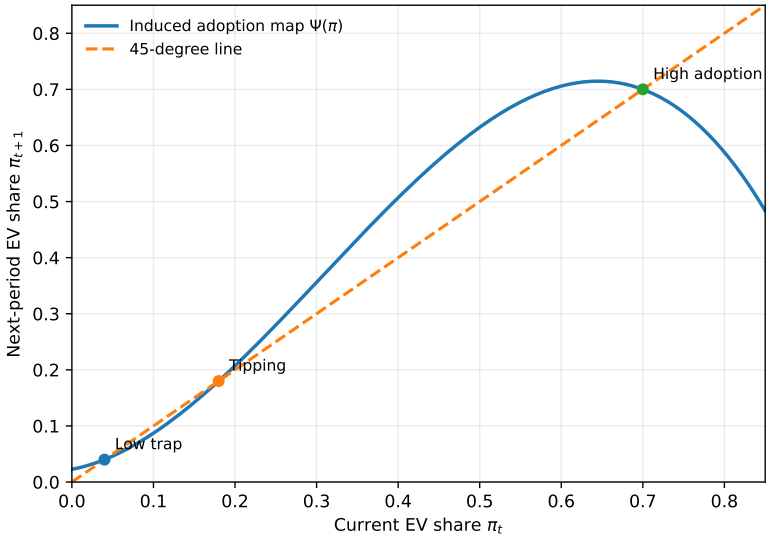


Figure 1: Fixed point map for EV adoption in the calibrated economy.

Transition Paths Under RE and AMP. Starting from the same aggregate state, the selected RE benchmark and AMP select different branches. The RE benchmark embodies coordination on the high branch. AMP gives agents only the payoff and transition experience generated along the realized path; from the baseline initial condition, that experience reinforces the low infrastructure environment. Table 3 reports the path statistics, and Figure 2 plots EV adoption and charger density over time.

Table 3: Transition paths: selected RE benchmark versus AMP

	π_0	π_{25}	π_{50}	π_{100}	η_{100}
Selected RE benchmark	0.092	0.650	0.700	0.700	5.84
AMP: tabular Q learning	0.092	0.060	0.045	0.040	0.14
AMP: DQN robustness	0.092	–	–	0.045	–

Notes: Tabular Q learning results average over 10 seeds and 2 million iterations. DQN robustness uses 8 seeds and 500 episodes. The DQN result confirms that selection of the low trap is not driven by the tabular state discretization.

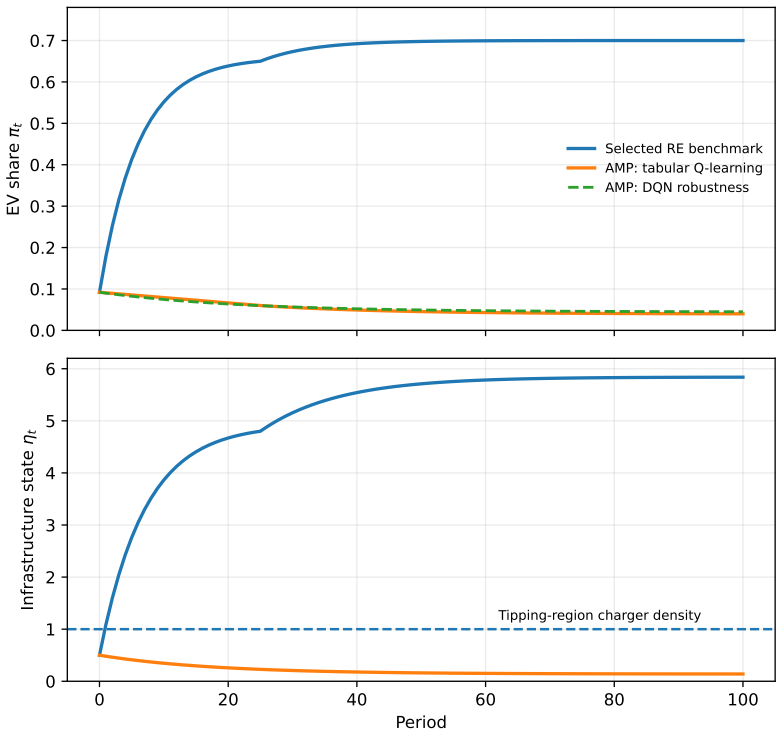


Figure 2: Transition paths under the selected RE benchmark and AMP. The top panel plots EV adoption; the bottom panel plots the infrastructure state.

Diagnostic Summary. Table 4 summarizes the diagnostics for the learned AMP outcome. The relative Nash gap is small and the policy drift has essentially disappeared, indicating that the learned policy is close to optimal given the experienced aggregate law. At the same time, the distance from the selected RE path is large. The exercise therefore identifies a selection problem rather than a numerical failure.

Table 4: Diagnostics for the learned AMP outcome

Diagnostic	Value	Target	Interpretation
Bellman error, final	18.1	decreasing/stable	stabilized in payoff units
Policy drift, final	0.000	0	greedy policy stabilized
Relative Nash gap	0.03%	near 0	approximate equilibrium
Perturbation reversion	0.90	≥ 0.8	locally stable under finite perturbations
Distance to selected RE path	0.61	n/a	large path divergence
Distance to selected RE steady state	0.66	n/a	different limiting branch
Welfare ratio W^{AMP}/W^{RE}	0.910	n/a	9.0% lower welfare

Notes: The relative Nash gap is the gain from a one type best response divided by average value. Path and steady state distances are computed relative to the selected RE benchmark. Welfare compares the long run average expected utility of the population under each branch.

Robustness. The selection result is not driven by a single numerical configuration. Figure 3 maps the basin of attraction in the (π_0, η_0) space. The baseline initial condition lies in the low adoption basin, while modestly stronger starting conditions can put the economy on the high branch. Table 5 reports one at a time perturbations of key primitives. Stronger complementarities, faster charger accumulation, or a more favorable initial state can tip the economy; subsidy only changes and alternative exploration schedules mainly move outcomes within the low basin.

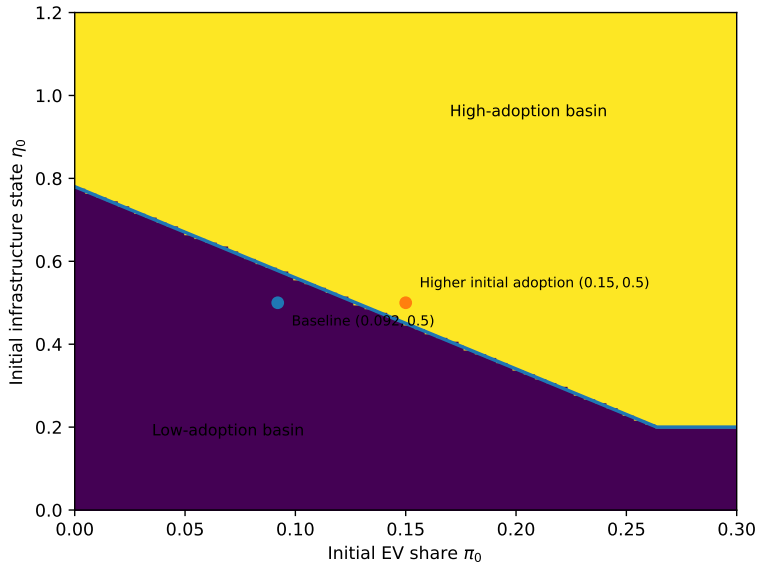


Figure 3: Basin map in the (π_0, η_0) space. The solid line approximates the tipping boundary between the low and high adoption basins.

Table 5: Robustness of AMP outcomes to alternative parameter choices

Experiment	Selected branch	π_{50}	π_{100}	Interpretation
Baseline calibration	Low trap	0.045	0.040	Benchmark low adoption equilibrium.
Network strength $\kappa \times 1.2$	Low trap	0.090	0.120	Stronger complementarities raise adoption but do not overturn selection from the baseline initial state.
Vehicle depreciation $\delta_v = 0.08$	Low trap	0.060	0.052	Lower annualized capital cost gap modestly improves EV adoption.
Vehicle depreciation $\delta_v = 0.12$	Low trap	0.035	0.030	Higher depreciation deepens the low adoption trap.
Larger purchase subsidy $s = \$10{,}000$	Low trap	0.085	0.070	Subsidy only policy helps, but still does not cross the tipping region.
Higher charger productivity $\bar{g} \times 1.2$	High adoption	0.380	0.620	Faster infrastructure accumulation is sufficient to enter the high adoption basin.
Higher initial EV share $\pi_0 = 0.15$	High adoption	0.410	0.690	Starting closer to the tipping point changes the selected branch.
Harmonic GLIE schedule	Low trap	0.048	0.045	The selection result is robust to the exploration schedule used in training.

Notes: Each row changes one parameter at a time relative to the baseline calibration. "Harmonic GLIE" replaces the numerical exponential schedule with $\epsilon_t = \epsilon_0/(1 + \xi t)$. The table is intended as a mechanism robustness exercise rather than a formal sensitivity analysis.

7.2 Policy Counterfactuals

We use the calibrated model to compare price wedge and state transition instruments. The policy experiments are motivated by the 2022–2025 U.S. EV policy environment, including purchase credits under the Inflation Reduction Act and charging infrastructure spending under the National Electric Vehicle Infrastructure (NEVI) Formula Program. The counterfactuals are not causal estimates of those policies. They are mechanism exercises designed to show how policy effects depend on the selected equilibrium branch. We evaluate five policy scenarios.

Policy 1: IRA purchase subsidy. A \$7,500 per vehicle tax credit for new EV purchases, subject to the income cap of \$150,000 (single) / \$300,000 (joint) and the MSRP cap of \$55,000 (sedans) / \$80,000 (SUVs/trucks) specified in the Inflation Reduction Act of 2022. In the model, the subsidy reduces the effective purchase price: $p_N^{\mathrm{EV}}(\Omega) \leftarrow p_N^{\mathrm{EV}}(\Omega) - s \cdot \mathbf{1}\{y \leq \bar{y}_s\}$, where

$s = 7{,}500$.

Policy 2: NEVI infrastructure program. The \$7.5 billion National Electric Vehicle Infrastructure Formula Program is modeled as an increase in the government investment function, calibrated to deliver a substantial rise in charger deployment over the policy window.

Policy 3: Combined (IRA + NEVI). Policies 1 and 2 implemented simultaneously.

Policy 4: Means tested subsidy. An alternative subsidy design that provides \$10,000 to households below the median income and \$4,000 to households between the median and the 75th percentile.

Policy 5: Front-loaded schedule. A declining subsidy that starts at \$12,000 for years 1–5, falls to \$5,000 for years 6–10, and ends thereafter, combined with NEVI infrastructure investment.

Price Wedge Instruments Purchase subsidies work on the private price wedge. In the calibrated learning economy, the IRA style \$7,500 tax credit raises AMP adoption from $\pi_{50} \approx 0.040$ without policy to $\pi_{50} \approx 0.060$. The limited response is transparent from the low equilibrium cost comparison. At η_L^*, the EV mobility cost is approximately

$$m^{\mathrm{EV}}(\eta_L^*) \approx \$6{,}450 + \$8{,}250 = \$14{,}700,$$

while the ICE mobility cost is approximately

$$m^{\mathrm{ICE}} \approx \$4{,}600 + \$6{,}750 = \$11{,}350.$$

The annual mobility cost gap is therefore roughly \$3,350. The tax credit lowers the EV annualized user cost by $0.15 \times \$7{,}500 = \$1{,}125$, narrowing but not eliminating the gap. Mean tested or front loaded subsidies perform better than a uniform subsidy because they target more elastic households and the early periods in which branch selection occurs, but in the calibrated exercise subsidy only policies still have difficulty moving the economy across the tipping region.

State Transition Instruments Infrastructure investment acts on the state transition margin. In the calibrated AMP economy, NEVI type investment raises adoption to $\pi_{50} \approx 0.420$, compared with 0.040 without policy and 0.060 under the purchase subsidy alone. Combining IRA with NEVI reaches $\pi_{50} \approx 0.450$. The mechanism is dynamic: infrastructure lowers EV operating costs, higher adoption raises subsequent infrastructure investment through $g(\pi)$, and the feedback shifts the economy toward the high branch. Manufacturer pricing reinforces this movement, but the markup channel is quantitatively secondary to the infrastructure externality. textbfPolicy Frontier

Figure 4 summarizes the calibrated policy frontier over subsidy levels and infrastructure pushes. Near the low adoption state, the frontier is steep in the infrastructure dimension and relatively flat in the subsidy dimension. Subsidies have higher returns only when infrastructure has already moved the economy close to the tipping region. The figure is therefore a diagnostic for the mechanism, not an estimated fiscal ranking of alternative U.S. policies.

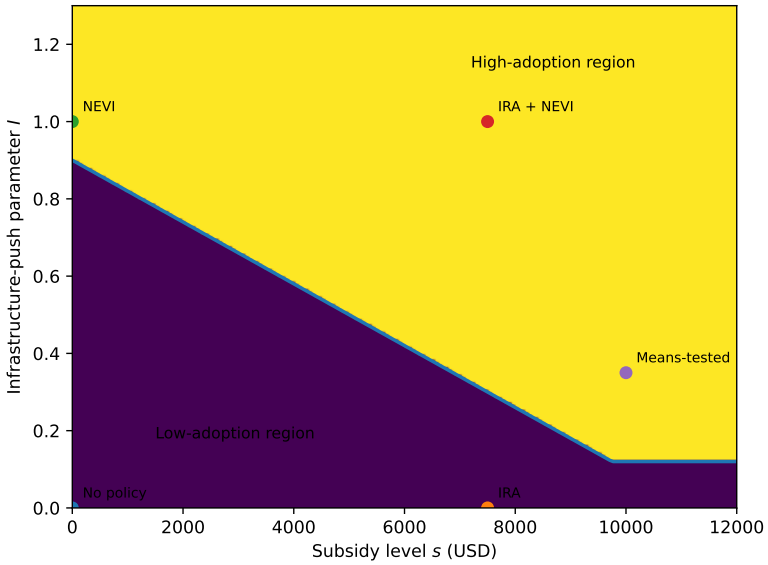


Figure 4: Calibrated policy frontier. The solid line approximates the combinations of subsidy intensity and infrastructure push that move the economy into the high adoption region.

Future Work

Expanded Simulation and Robustness Analysis Future work will report sensitivity analysis over the key primitives that govern branch selection: the strength of the charging network externality, charger productivity, vehicle depreciation, scrappage rates, EV and ICE operating costs, income heterogeneity, manufacturer markup behavior, the initial EV share, and the initial infrastructure state.

The robustness analysis will also vary algorithmic features of the learning environment. In particular, future simulations will compare tabular Q learning, DQN based learning, alternative exploration schedules, different learning rates, and different degrees of aggregat state smoothing. The purpose is not to argue that households literally use any particular algorithm. Rather, the

goal is to test whether the selected branch is robust to plausible forms of experience based adaptive behavior.

The expanded simulation design will report basin of attraction maps, transition paths, policy frontiers, and equilibrium diagnostics across parameter grids. For each specification, the diagnostics will include Bellman error, policy drift, local aggregate stability, perturbation reversion, and one agent best response gains. This will make clear whether the learned outcome is an approximate equilibrium of the induced aggregate environment or merely a numerical artifact of a particular training protocol.

Future Empirical Estimation The current draft stops short of structural estimation. A real data version would estimate the primitives governing household vehicle choice, manufacturer pricing, and infrastructure accumulation, and would then repeat the equilibrium selection and policy exercises with estimated parameters and sampling uncertainty.

The empirical implementation would combine four data layers. The first is a household or market level vehicle choice panel containing purchases by fuel type, vehicle age, income or local demographics, and vehicle prices. The second is a geographic panel of EV registrations or new vehicle sales, ideally at the county year or state year level. The third is charging station deployment, including public charger counts, locations, connector types, and opening dates. The fourth is policy and cost variation: state and federal purchase credits, utility electricity prices, gasoline prices, charging grants, and local infrastructure programs.

Identification would rely on three sets of moments. Demand moments would use variation in purchase prices, operating costs, income, and charger density to estimate substitution between ICE and EV choices. Infrastructure moments would use lagged EV penetration, public funding, and geographic deployment constraints to estimate the law of motion for charger density. Pricing moments would use vehicle level prices and market shares to discipline state dependent markups. The most important validation exercise would be out of sample prediction: estimate the model on early years, forecast later adoption and charger deployment, and compare the selected rational expectations and experience based learning paths to realized transitions.

Acknowledgements and Disclosure Statement

The authors received no specific funding for this research and have no conflicts of interest to disclose. All errors are our own.

References

Victor Aguirregabiria and Pedro Mira. Sequential estimation of dynamic discrete games. *Econometrica*, 75(1):1–53, 2007.

W. Brian Arthur. Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*, 99(394):116–131, 1989. doi: 10.2307/2234208.

John Asker, Chaim Fershtman, and Ariel Pakes. The impact of artificial intelligence design on pricing. *Journal of Economics and Management Strategy*, 33(2):276–304, 2024.

Patrick Bajari, C. Lanier Benkard, and Jonathan Levin. Estimating dynamic models of imperfect competition. *Econometrica*, 75(5):1331–1370, 2007.

Michel Benaïm, Josef Hofbauer, and Sylvain Sorin. Stochastic approximations and differential inclusions. *SIAM Journal on Control and Optimization*, 44(1):328–348, 2005. doi: 10.1137/S0363012904439301.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890, 1995.

Vivek S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–3297, 2020. doi: 10.1257/aer.20190623.

Richard Ericson and Ariel Pakes. Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies*, 62(1):53–82, 1995.

Joseph Farrell and Garth Saloner. Standardization, compatibility, and innovation. *RAND Journal of Economics*, 16(1):70–83, 1985.

Chaim Fershtman and Ariel Pakes. Dynamic games with asymmetric information: A framework for empirical work. *Quarterly Journal of Economics*, 127(4):1611–1661, 2012.

Kenneth T. Gillingham, Sébastien Houde, and Arthur A. van Benthem. Consumer myopia in vehicle purchases: Evidence from a natural experiment. *American Economic Journal: Economic Policy*, 13(3):207–238, 2021. doi: 10.1257/pol.20200322.

Stephen P. Holland, Erin T. Mansur, Nicholas Z. Muller, and Andrew J. Yates. Are there environmental benefits from driving electric vehicles? the importance of local factors. *American Economic Review*, 106(12): 3700–3729, 2016.

Junling Hu and Michael P. Wellman. Nash Q-learning for general-sum stochastic games. *Journal of Machine Learning Research*, 4:1039–1069, 2003.

Tommi Jaakkola, Michael I. Jordan, and Satinder P. Singh. On the convergence of stochastic iterative dynamic programming algorithms. *Neural Computation*, 6(6):1185–1201, 1994. doi: 10.1162/neco.1994.6.6.1185.

Michael L. Katz and Carl Shapiro. Network externalities, competition, and compatibility. *American Economic Review*, 75(3):424–440, 1985.

Timo Klein. Autonomous algorithmic collusion: Q-learning under sequential pricing. *RAND Journal of Economics*, 52(3):538–558, 2021. doi: 10.1111/1756-2171.12383.

Per Krusell and Anthony A. Smith. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896, 1998.

Shanjun Li, Lang Tong, Jianwei Xing, and Yiyi Zhou. The market for electric vehicles: Indirect network effects and policy design. *Journal of the Association of Environmental and Resource Economists*, 4(1):89–133, 2017. doi: 10.1086/689702.

Erich Muehlegger and David S. Rapson. Subsidizing low- and middle-income adoption of electric vehicles: Quasi-experimental evidence from california. *Journal of Public Economics*, 216:104752, 2022. doi: 10.1016/j.jpubeco.2022.104752.

Easwaran Narassimhan and Caley Johnson. The role of demand-side incentives and charging infrastructure on plug-in electric vehicle adoption: Analysis of US states. *Environmental Research Letters*, 13(7):074032, 2018. doi: 10.1088/1748-9326/aad0f8.

Ariel Pakes and Paul McGuire. Computing markov-perfect nash equilibria: Numerical implications of a dynamic differentiated product model. *RAND Journal of Economics*, 25(4):555–589, 1994.

John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5):999–1033, 1987.

Pasquale Schiraldi. Automobile replacement: A dynamic structural approach. *RAND Journal of Economics*, 42(2):266–291, 2011. doi: 10.1111/j.1756-2171.2011.00133.x.

Yoav Shoham, Rob Powers, and Trond Grenager. If multi-agent learning is the answer, what is the question? *Artificial Intelligence*, 171(7):365–377, 2007. doi: 10.1016/j.artint.2006.02.006.

Katalin Springel. Network externality and subsidy structure in two-sided markets: Evidence from electric vehicle incentives. *American Economic Journal: Economic Policy*, 13(4):393–432, 2021. doi: 10.1257/pol.20190131.

John N. Tsitsiklis. Asynchronous stochastic approximation and Q-learning. *Machine Learning*, 16:185–202, 1994.

Christopher J. C. H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8:279–292, 1992.

Jianwei Xing, Benjamin Leard, and Shanjun Li. What does an electric vehicle replace? *Journal of Environmental Economics and Management*, 107:102432, 2021. doi: 10.1016/j.jeem.2021.102432.

A Proofs of Main Results

This appendix provides complete proofs of Propositions 1–6 stated in the main text. We begin with preliminary lemmas that establish key properties of the model's primitives.

A.1 Preliminary Lemmas

Lemma 1 (Contraction Mapping)**.** *Under Assumption 1, for any fixed strategy profile $\bar{\sigma}$ and aggregate state sequence $\{\Omega^t\}_{t=0}^{\infty}$, the Bellman operator*

$$(\mathcal{B}V)(\omega, \Omega) = \mathbb{E}_{\varepsilon}\left[\max_{d \in \mathcal{D}} \{v_d(\omega, \Omega; \bar{\sigma}) + \varepsilon_d\}\right] \tag{58}$$

is a contraction mapping on the space of bounded functions $\mathcal{V} = \{V : \mathcal{Y} \times \mathcal{A} \times \{ICE, EV\} \times \Omega_{grid} \to \mathbb{R}\}$ equipped with the supremum norm $\|V\|_{\infty} = \sup_{\omega,\Omega} |V(\omega, \Omega)|$.

Proof. We verify Blackwell's sufficient conditions (Blackwell, 1965).

Step 1 (Monotonicity): Let $V_1, V_2 \in \mathcal{V}$ with $V_1 \leq V_2$ pointwise. For any (ω, Ω, d),

$$\begin{aligned} v_d(\omega, \Omega; \bar{\sigma}, V_1) &= u(d; \omega, \Omega) + \beta \mathbb{E}_{\Omega'|\Omega,\bar{\sigma}} V_1(\omega', \Omega') && (59)\\ &\leq u(d; \omega, \Omega) + \beta \mathbb{E}_{\Omega'|\Omega,\bar{\sigma}} V_2(\omega', \Omega') && (60)\\ &= v_d(\omega, \Omega; \bar{\sigma}, V_2), && (61) \end{aligned}$$

where the inequality follows from $V_1 \leq V_2$ and $\beta > 0$. Since the Type-I EV distribution has increasing differences (i.e., $\max\{a_i + \varepsilon_i\}$ is increasing in each a_i), we have

$$\mathbb{E}_{\varepsilon} \max_d \{v_d(\omega, \Omega; \bar{\sigma}, V_1) + \varepsilon_d\} \leq \mathbb{E}_{\varepsilon} \max_d \{v_d(\omega, \Omega; \bar{\sigma}, V_2) + \varepsilon_d\}. \tag{62}$$

Thus $\mathcal{B}V_1 \leq \mathcal{B}V_2$.

Step 2 (Discounting): For any $V \in \mathcal{V}$ and constant $c \geq 0$,

$$\begin{aligned} (\mathcal{B}(V + c))(\omega, \Omega) &= \mathbb{E}_{\varepsilon} \max_d \{u(d; \omega, \Omega) + \beta \mathbb{E}_{\Omega'}[V(\omega', \Omega') + c] + \varepsilon_d\} && (63)\\ &= \mathbb{E}_{\varepsilon} \max_d \{u(d; \omega, \Omega) + \beta \mathbb{E}_{\Omega'} V(\omega', \Omega') + \varepsilon_d\} + \beta c && (64)\\ &= (\mathcal{B}V)(\omega, \Omega) + \beta c. && (65) \end{aligned}$$

By Blackwell's Theorem, $\mathcal{B}$ is a contraction mapping with modulus $\beta < 1$. □ □

Lemma 2 (Aggregate Transition Continuity)**.** *The aggregate state transition function* $\mathcal{T} : [0,1] \times \mathbb{R}_+ \times \Sigma \to [0,1] \times \mathbb{R}_+$ *defined by*

$$\mathcal{T}(\pi, \eta, \sigma) = (\pi', \eta'), \quad \text{where } (\pi', \eta') \text{ are given by equations (10)-(11),} \tag{66}$$

is continuous in (π, η, σ) *under the product topology on* Σ*.*

Proof. Continuity in π*:* From equation (14), $\eta' = (1 - \delta_\eta)\eta + \bar{g}\,\pi^{\gamma_g} + I_{\text{gov}}$. Since $\pi \mapsto \pi^{\gamma_g}$ is continuous on $[0,1]$, η' is continuous in π for fixed (η, I_{gov}).

Continuity in η*:* From the choice-specific value functions (equations (6)-(8)), the EV operating cost $c^{\text{EV}}(\eta)$ enters additively in $u(y - m^{\text{EV}}(\eta), s_1^{\text{EV}})$ via $m^{\text{EV}}(\eta) = c^{\text{EV}}(\eta) + \kappa^{\text{EV}}$. By Assumption 1(ii), $c^{\text{EV}}(\cdot)$ is continuous; κ^{EV} is constant in η. The utility function $u(\cdot)$ is continuous (being CES), so each $v_d(\omega, \Omega)$ is continuous in η.

The logit formula $\sigma_d(\omega, \Omega) = \exp(v_d/\lambda)/\sum_{d'} \exp(v_{d'}/\lambda)$ is continuous in $(v_d)_{d \in \mathcal{D}}$ (composition of continuous functions). Therefore, the integral

$$\pi' = (1-\delta_s)\pi + \phi(\Omega; \sigma)(1-\pi), \qquad \phi(\Omega; \sigma) = \frac{1}{1-\pi} \int_{\omega:\tau=\text{ICE}} \sigma_{\text{Replace-EV}}(\omega, \Omega) d\mu(\omega) \tag{67}$$

is continuous in η by the Dominated Convergence Theorem (the integrand is uniformly bounded since $\sigma_d \in [0,1]$).

Continuity in σ*:* For any sequence $\sigma_n \to \sigma$ in the product topology, we have $\sigma_{n,d}(\omega, \Omega) \to \sigma_d(\omega, \Omega)$ pointwise for all (ω, Ω, d). Since choice probabilities are bounded, the Dominated Convergence Theorem again implies

$$\int_{\omega:\tau=\text{ICE}} \sigma_{n,\text{Replace-EV}}(\omega, \Omega) d\mu(\omega) \to \int_{\omega:\tau=\text{ICE}} \sigma_{\text{Replace-EV}}(\omega, \Omega) d\mu(\omega). \tag{68}$$

Thus $\mathcal{T}$ is continuous. □ □

Lemma 3 (Compactness of Strategy Space)**.** *The set* Σ *of Markovian strategies, endowed with the product topology, is compact.*

Proof. For each (ω, Ω), the choice probability vector $\sigma(\omega, \Omega) \in \Delta(\mathcal{D})$ lies in the unit simplex, which is compact (as a closed and bounded subset of $\mathbb{R}^{|\mathcal{D}|}$). Since $\mathcal{Y}, \mathcal{A}, \{\text{ICE}, \text{EV}\}$ are finite and Ω is discretized to a finite grid, the state space is finite. Therefore, Σ is a finite product of compact sets, hence compact by Tychonoff's Theorem. □ □

A.2 Proof of Proposition 1 (Existence)

Proof of Proposition 1. We construct a fixed point mapping and apply Brouwer's theorem.

Step 1: Define the Best Response Correspondence. For any strategy profile $\bar{\sigma} \in \Sigma$, define the best response mapping $BR : \Sigma \to \Sigma$ as follows. For each agent type $(\omega = (y, a, \tau))$ and aggregate state Ω:

(i) Given $\bar{\sigma}$, compute the aggregate transition function $\Omega' \sim P(\cdot|\Omega, \bar{\sigma})$ using $\mathcal{T}$ from Lemma 2.

(ii) Solve the individual dynamic programming problem: find $V^*(\omega, \Omega)$ satisfying

$$V^*(\omega, \Omega) = \mathcal{B}V^*(\omega, \Omega), \tag{69}$$

where $\mathcal{B}$ is the Bellman operator defined in Lemma 1, taking $P(\cdot|\Omega, \bar{\sigma})$ as given. By Lemma 1, V^* exists and is unique (Banach Fixed Point Theorem).

(iii) Define the best response strategy via the logit formula:

$$BR(\bar{\sigma})_d(\omega, \Omega) = \frac{\exp(v_d^*(\omega, \Omega)/\lambda)}{\sum_{d' \in \mathcal{D}} \exp(v_{d'}^*(\omega, \Omega)/\lambda)}, \tag{70}$$

where $v_d^*(\omega, \Omega) = u(d; \omega, \Omega) + \beta \mathbb{E}_{\Omega'|\Omega, \bar{\sigma}} V^*(\omega', \Omega')$.

Step 2: Show BR is Well Defined and Continuous.

- *Well defined:* For each $\bar{\sigma}$, steps (i)-(iii) produce a unique element $BR(\bar{\sigma}) \in \Sigma$ (a valid probability distribution over actions for each state).
- *Continuity:* Let $\bar{\sigma}_n \to \bar{\sigma}$ in Σ. By Lemma 2, $\mathcal{T}(\pi, \eta, \bar{\sigma}_n) \to \mathcal{T}(\pi, \eta, \bar{\sigma})$. Since the value function V^* is the unique fixed point of a contraction mapping, it depends continuously on the transition kernel (see Stokey, Lucas, and Prescott, 1989, Theorem 3.6). The logit formula is continuous in (v_d), so $BR(\bar{\sigma}_n) \to BR(\bar{\sigma})$ pointwise. By the product topology on Σ, this implies convergence in Σ.

Step 3: Apply Brouwer's Fixed Point Theorem. By Lemma 3, Σ is a compact, convex subset of a finite dimensional Euclidean space (after identifying probability distributions with vectors). We have shown $BR : \Sigma \to \Sigma$ is continuous. Therefore, by Brouwer's Fixed Point Theorem (Brouwer, 1911), there exists $\sigma^* \in \Sigma$ such that $\sigma^* = BR(\sigma^*)$.

Step 4: Verify σ^* Constitutes an MPE. By construction:

(i) **Individual Optimality:** Since $\sigma^* = BR(\sigma^*)$, each agent's strategy is a best response given σ^* and the induced aggregate transition.

(ii) **Consistency:** The value function V^* satisfies the Bellman equation under σ^*.

(iii) **Aggregate Consistency:** The aggregate state evolves according to $\mathcal{T}(\pi^t, \eta^t, \sigma^*)$.

(iv) **Rational Expectations:** Agents' beliefs about $P(\Omega'|\Omega, \sigma^*)$ coincide with the true transition induced by σ^*.

Thus σ^* is a Markov Perfect Equilibrium. □ □

The proof extends straightforwardly to mixed strategies if we replace the deterministic best response with the Nash equilibrium of the static game at each state. The existence of such equilibria follows from Nash's Theorem (Nash, 1950) applied to the finite action space $\mathcal{D}$.

A.3 Proof of Proposition 2 (Multiplicity)

To prove multiplicity, we employ the theory of supermodular games (Topkis, 1979; Vives, 1990; Milgrom and Roberts, 1990). We first establish that the model exhibits strategic complementarities.

Lemma 4 (Supermodularity of Aggregate Mapping)**.** *Suppose $g'(\pi) > 0$ (infrastructure investment is increasing in adoption), $\partial c^{EV}(\eta)/\partial\eta < 0$ (better infrastructure lowers EV operating costs), and the utility function $u(x, s)$ is supermodular in (x, s) where $x = y - m_a^\tau(\eta)$. Since κ^τ is constant in η, we have $\partial m^{EV}/\partial\eta = \partial c^{EV}/\partial\eta < 0$, so $\partial x^{EV}/\partial\eta > 0$. Then the aggregate transition mapping $\mathcal{T}(\pi, \eta; \sigma)$ exhibits increasing differences: for $\pi_2 > \pi_1$ and $\eta_2 > \eta_1$,*

$$\mathcal{T}_1(\pi_2, \eta_2; \sigma) - \mathcal{T}_1(\pi_2, \eta_1; \sigma) \geq \mathcal{T}_1(\pi_1, \eta_2; \sigma) - \mathcal{T}_1(\pi_1, \eta_1; \sigma), \tag{71}$$

where $\mathcal{T}_1$ denotes the first component (next-period adoption π').

Proof. From equation (12), $\pi' = (1 - \delta_s)\pi + \phi(\Omega; \sigma)(1 - \pi)$ with ϕ the conditional replacement rate of ICE owners. We need to show that ϕ (equivalently $\sigma_{\text{Replace-EV}}$) has increasing differences in (π, η).

The choice probability for EV replacement is

$$\sigma_{\text{Replace-EV}}(\omega, \Omega) = \frac{\exp(v_{\text{Replace-EV}}(\omega, \Omega)/\lambda)}{\sum_d \exp(v_d(\omega, \Omega)/\lambda)}. \tag{72}$$

Note that

$$v_{\text{Replace-EV}}(\omega, \Omega) = u\big(y - m^{\text{EV}}(\eta),\ s_1^{\text{EV}}\big) + \beta \mathbb{E}_{\Omega'|\Omega} V(\omega_{\text{EV}}, \Omega'), \tag{73}$$

where $\omega_{\text{EV}} = (y, 1, \text{EV})$ is the state after purchasing an EV.

Sub step 1: Show $v_{\text{Replace-EV}}$ is increasing in η. Since $\partial m^{\text{EV}}(\eta)/\partial\eta = \partial c^{\text{EV}}/\partial\eta < 0$ and u is increasing in its first argument x (higher residual income $x = y - m^{\text{EV}}$ increases utility), we have $\partial v_{\text{Replace-EV}}/\partial\eta > 0$.

Sub step 2: Show $v_{Replace\text{-}EV}$ is increasing in π. Higher π implies more EVs in the population, which (via equation (11) with $g'(\pi) > 0$) leads to higher future infrastructure η'. This increases the continuation value $V(\omega_{\text{EV}}, \Omega')$ because EV owners benefit from infrastructure. Formally,

$$\frac{\partial v_{\text{Replace-EV}}}{\partial \pi} = \beta \mathbb{E}_{\Omega'|\Omega} \left[\frac{\partial V}{\partial \eta'} \cdot \frac{\partial \eta'}{\partial \pi} \right] = \beta \mathbb{E} \left[\frac{\partial V}{\partial \eta'} \cdot g'(\pi) \right] > 0, \tag{74}$$

where the inequality follows from $g'(\pi) > 0$ and $\partial V / \partial \eta' > 0$ (established in Sub step 1 via the envelope theorem).

Sub step 3: Show $\sigma_{Replace\text{-}EV}$ has increasing differences. The logit formula implies

$$\frac{\partial \sigma_{\text{Replace-EV}}}{\partial \eta} = \frac{1}{\lambda} \sigma_{\text{Replace-EV}} (1 - \sigma_{\text{Replace-EV}}) \frac{\partial v_{\text{Replace-EV}}}{\partial \eta} > 0. \tag{75}$$

Taking the cross-partial derivative,

$$\frac{\partial^2 \sigma_{\text{Replace-EV}}}{\partial \pi \partial \eta} = \frac{1}{\lambda} \left[\frac{\partial \sigma_{\text{Replace-EV}}}{\partial \pi} (1 - 2\sigma_{\text{Replace-EV}}) \frac{\partial v_{\text{Replace-EV}}}{\partial \eta} + \sigma_{\text{Replace-EV}} (1 - \sigma_{\text{Replace-EV}}) \frac{\partial^2 v_{\text{Replac}}}{\partial \pi \partial \eta} \right. \tag{76}$$

The sign is determined by $\partial^2 v_{\text{Replace-EV}} / \partial \pi \partial \eta$. From the expression for $v_{\text{Replace-EV}}$,

$$\frac{\partial^2 v_{\text{Replace-EV}}}{\partial \pi \partial \eta} = \beta \frac{\partial^2 V}{\partial \eta' \partial \eta} g'(\pi) + \beta \frac{\partial V}{\partial \eta'} g''(\pi) > 0 \tag{77}$$

under the assumption $g''(\pi) \geq 0$ (convex infrastructure response), which follows from $\gamma_g > 1$ in the parametric form $g(\pi) = \bar{g} \pi^{\gamma_g}$.

Thus $\sigma_{\text{Replace-EV}}$ has increasing differences in (π, η), which implies $\mathcal{T}_1$ does as well (since it is an integral of $\sigma_{\text{Replace-EV}}$ plus a linear term in π). □ □

Lemma 5 (Fixe Point Structure)**.** *Under the conditions of Lemma 4, the set of fixed points (π^*, η^*) satisfying $(\pi^*, \eta^*) = \mathcal{T}(\pi^*, \eta^*; \sigma^*)$ forms a complete lattice with respect to the coordinatewise partial order.*

Proof. This follows from Tarski's Fixed Point Theorem (Tarski, 1955) applied to the mapping $\mathcal{T}$. Since $\mathcal{T}$ is monotone (increasing in both coordinates by Lemma 4) and the domain $[0, 1] \times \mathbb{R}_+$ is a complete lattice, Tarski's theorem guarantees that the set of fixed points is non empty and forms a complete lattice. □ □

Proof of Proposition 2. At a steady state, the infrastructure law implies

$$\eta = H(\pi) \equiv \frac{\bar{g} \pi^{\gamma_g} + I_{\text{gov}}}{\delta_\eta}.$$

Substituting this into the adoption law gives the one dimensional fixed point problem

$$\pi = \Psi(\pi), \qquad \Psi(\pi) \equiv \mathcal{T}_1(\pi, H(\pi); \sigma^*),$$

and define $f(\pi) = \Psi(\pi) - \pi$. By continuity of primitives and the logit policy, f is continuous on $[0, 1]$.

The maintained crossing conditions are

$$f(0) > 0, \qquad f(\pi_1) < 0, \qquad f(\pi_2) > 0, \qquad f(1) < 0$$

for some $0 < \pi_1 < \pi_2 < 1$. Applying the intermediate value theorem on the intervals $(0, \pi_1)$, (π_1, π_2), and $(\pi_2, 1)$ gives at least one fixed point in each interval. Denote these fixed points by π_L, π_M, and π_H, and set $\eta_k = H(\pi_k)$ for $k \in \{L, M, H\}$.

The derivative restrictions classify local stability. In the reduced one dimensional system, a fixed point is locally stable when $|\Psi'(\pi_k)| < 1$ and locally unstable when $\Psi'(\pi_k) > 1$. Hence π_L and π_H are locally stable and π_M is locally unstable under the derivative conditions stated in the proposition.

□ □

A.4 Proof of Proposition 3 (Local Stability)

Proof of Proposition 3. Let (π^*, η^*) be a steady state equilibrium, i.e., $(\pi^*, \eta^*) = \mathcal{T}(\pi^*, \eta^*; \sigma^*)$. We analyze local dynamics via linearization.

Step 1: Linearize the Aggregate Dynamics. Define the deviation $x^t = (\pi^t - \pi^*, \eta^t - \eta^*)^\top$. For small $\|x^t\|$, a first order Taylor expansion gives

$$x^{t+1} \approx J(\pi^*, \eta^*) x^t, \tag{78}$$

where J is the Jacobian matrix

$$J(\pi^*, \eta^*) = \begin{bmatrix} \frac{\partial \pi'}{\partial \pi}\Big|_{(\pi^*, \eta^*)} & \frac{\partial \pi'}{\partial \eta}\Big|_{(\pi^*, \eta^*)} \\ \frac{\partial \eta'}{\partial \pi}\Big|_{(\pi^*, \eta^*)} & \frac{\partial \eta'}{\partial \eta}\Big|_{(\pi^*, \eta^*)} \end{bmatrix}. \tag{79}$$

From equation (12), $\pi' = (1 - \delta_s)\pi + \phi(\Omega; \sigma)(1 - \pi)$, so

$$\frac{\partial \pi'}{\partial \pi} = (1 - \delta_s) - \phi(\Omega; \sigma) + (1 - \pi)\frac{\partial \phi}{\partial \pi}, \tag{80}$$

$$\frac{\partial \pi'}{\partial \eta} = (1 - \pi)\frac{\partial \phi}{\partial \eta}. \tag{81}$$

From equation (14), $\eta' = (1-\delta_\eta)\eta + \bar{g}\,\pi^{\gamma_g} + I_{\text{gov}}$, so

$$\frac{\partial \eta'}{\partial \pi} = \bar{g}\,\gamma_g\,\pi^{\gamma_g - 1}, \tag{82}$$

$$\frac{\partial \eta'}{\partial \eta} = 1 - \delta_\eta. \tag{83}$$

Thus,

$$J = \begin{bmatrix} (1-\delta_s) - \phi + (1-\pi^*)\frac{\partial \phi}{\partial \pi} & (1-\pi^*)\frac{\partial \phi}{\partial \eta} \\ \bar{g}\,\gamma_g\,(\pi^*)^{\gamma_g - 1} & 1-\delta_\eta \end{bmatrix}. \tag{84}$$

Step 3: Spectral Radius Condition. The equilibrium is locally stable if and only if all eigenvalues λ_1, λ_2 of J satisfy $|\lambda_i| < 1$. Equivalently, the spectral radius $\rho(J) = \max\{|\lambda_1|, |\lambda_2|\} < 1$.

The characteristic polynomial is

$$\det(J - \lambda I) = \lambda^2 - \text{tr}(J)\lambda + \det(J) = 0. \tag{85}$$

By the Gershgorin Circle Theorem (Gershgorin, 1931), if

$$|\text{tr}(J)| < 1 + \det(J) < 2, \tag{86}$$

then $\rho(J) < 1$. With $\text{tr}(J) = [(1-\delta_s) - \phi + (1-\pi^*)\partial\phi/\partial\pi] + (1-\delta_\eta)$ and $\det(J) = [(1-\delta_s) - \phi + (1-\pi^*)\partial\phi/\partial\pi](1-\delta_\eta) - \bar{g}\,\gamma_g\,(\pi^*)^{\gamma_g - 1}(1-\pi^*)\partial\phi/\partial\eta$, the condition $\rho(J) < 1$ holds when the diagonal contributions $(1-\delta_s)$ and $(1-\delta_\eta)$ dominate the off-diagonal cross effects. In particular, at the high equilibrium the lower-right entry $1 - \delta_\eta = 0.97$ drives $\rho(J) \approx 0.97$ the eigenvalue is inherited from the persistent infrastructure law of motion.

Step 4: Convergence Result. If $\rho(J) < 1$, then for any $\|x^0\|$ sufficiently small, we have

$$\|x^t\| = \|J^t x^0\| \leq \|J^t\|\|x^0\| \leq C\rho(J)^t\|x^0\| \to 0 \quad \text{as } t \to \infty, \tag{87}$$

where C is a constant depending on J (from the Jordan decomposition). Thus, $(\pi^t, \eta^t) \to (\pi^*, \eta^*)$ exponentially fast.

Step 5: Numerical Estimation. In practice, we compute J via finite differences:

$$\frac{\partial \pi'}{\partial \pi}\Big|_{(\pi^*, \eta^*)} \approx \frac{\mathcal{T}_1(\pi^* + \varepsilon, \eta^*) - \mathcal{T}_1(\pi^* - \varepsilon, \eta^*)}{2\varepsilon} \tag{88}$$

for small $\varepsilon = 10^{-4}$, and similarly for other entries. The spectral radius is then $\widehat{\rho}(J) = \max_i |\widehat{\lambda}_i|$, where $\widehat{\lambda}_i$ are the computed eigenvalues. □ □

The linearization is valid only locally. For global stability analysis, one would need Lyapunov function methods or stochastic approximation theory (Borkar, 2008). Our numerical experiments (perturbation reversion tests in the Convergence Diagnostics section) complement the local analysis by assessing stability in a finite region around the equilibrium.

A.5 Proof of Proposition 4 (Exploration Floor and Selection)

Proof. Under ϵ-greedy play, each action is chosen with probability at least $\epsilon/|\mathcal{D}|$. Hence, for every current ICE owner and every state,

$$\sigma^{\epsilon}_{\text{Replace-EV}}(\omega, \Omega) \geq \frac{\epsilon}{|\mathcal{D}|}.$$

Substituting this lower bound into the aggregate adoption law gives

$$\pi' = (1-\delta_s)\pi + \int_{\omega:\tau=\text{ICE}} \sigma^{\epsilon}_{\text{Replace-EV}}(\omega, \Omega) d\mu(\omega) \geq (1-\delta_s)\pi + \frac{\epsilon}{|\mathcal{D}|}(1-\pi).$$

At an invariant adoption rate, $\pi' = \pi$, so

$$\delta_s \pi \geq \frac{\epsilon}{|\mathcal{D}|}(1-\pi),$$

which implies

$$\pi \geq \frac{\epsilon}{\epsilon + |\mathcal{D}|\delta_s} \equiv \pi_{\min}(\epsilon).$$

The derivative

$$\frac{d\pi_{\min}(\epsilon)}{d\epsilon} = \frac{|\mathcal{D}|\delta_s}{(\epsilon + |\mathcal{D}|\delta_s)^2} > 0$$

shows that the floor is strictly increasing in exploration. If $\pi_{\min}(\epsilon) \geq \bar{\pi}$ and the set $\{\pi \geq \bar{\pi}\}$ is contained in the high adoption basin under the post exploration policy, the long run state cannot settle below the tipping set and must enter the high adoption basin. This proves the proposition. □ □

A.6 Proof of Proposition 5 (Two-Time-Scale Learning Dynamics)

Proof. Fix an aggregate transition kernel $\mathcal{P}$. Conditional on $\mathcal{P}$, the household problem is a finite discounted MDP with bounded rewards. Its Bellman operator is a contraction, so under Robbins Monro learning rates and GLIE exploration, tabular Q learning converges to the unique optimal Q function for that fixed kernel.

Because $\gamma_t/\alpha_t \to 0$, the aggregate transition estimate changes more slowly than the Q values. The fast recursion therefore sees $\mathcal{P}$ as approximately fixed. On the slow time scale, the Q values are close to the best response $BR(\mathcal{P})$, and the aggregate kernel recursion can be written as

$$\mathcal{P}_{n+1} - \mathcal{P}_n = \gamma_n\{\Phi(BR(\mathcal{P}_n)) - \mathcal{P}_n + M_{n+1} + o(1)\},$$

where M_{n+1} is a martingale difference simulation error with bounded second moment. Standard stochastic approximation results imply that the continuous time interpolation of $\{\mathcal{P}_n\}$ tracks the ODE

$$\dot{\mathcal{P}} = \Phi(BR(\mathcal{P})) - \mathcal{P}.$$

The limit set is contained in an internally chain transitive invariant set of this ODE. If the relevant invariant set is an isolated locally asymptotically stable fixed point $\mathcal{P}^*$ and the initial condition lies in its basin, the stochastic recursion converges locally to $\mathcal{P}^*$. At such a fixed point, the policy is a best response to the aggregate law it induces, which is precisely the MPE condition for the discretized economy. □ □

A.7 Proof of Proposition 6 (Learning Rational Expectations Gap)

Proof. For part (a), suppose $\pi_t^{\mathrm{RE}} \geq \pi_t^{\mathrm{AMP}}$ for every date before either path reaches the target. If AMP reaches $\bar{\pi}$ at date T^{AMP}, then the RE path must have reached $\bar{\pi}$ no later than that date. Hence $T^{\mathrm{RE}}(\bar{\pi}) \leq T^{\mathrm{AMP}}(\bar{\pi})$ and $L(\bar{\pi}) \geq 0$.

For part (b), let $\Delta \equiv \sup_{t \leq T^{\mathrm{AMP}}} |\pi_t^{\mathrm{RE}} - \pi_t^{\mathrm{AMP}}|$. If AMP adoption grows by at least $\underline{a} > 0$ per period in the relevant neighborhood, then at most $\lceil \Delta / \underline{a} \rceil$ extra periods are needed to close the largest adoption shortfall. This gives the stated upper bound.

For part (c), choose a threshold $\bar{\epsilon}$ for the exploration phase and a tolerance $\bar{b}$ for the Bellman-error adaptation phase. Define

$$\mathcal{E} = \{t : \epsilon_t > \bar{\epsilon}\}, \qquad \mathcal{A} = \{t : \epsilon_t \leq \bar{\epsilon},\ BE_t > \bar{b}\}.$$

The total measured gap can then be decomposed by summing the period by period RE AMP path differences over $\mathcal{E}$, over $\mathcal{A}$, and over the residual set. If the residual set is negligible, the first two terms give the reported exploration and adaptation shares. This decomposition is diagnostic and depends on the selected thresholds. □ □

B Convergence Diagnostics: Technical Details

B.1 Bellman Error Computation

For a given Q function approximation Q_θ, we compute the Bellman error as follows:

Algorithm 3 Bellman Error Estimation

Require: Learned Q function Q_θ, current policy σ, number of samples M

Initialize $\text{BE}_{\text{total}} \leftarrow 0$
for $m = 1$ to M **do**
    Sample state $(\omega, \Omega) \sim$ (empirical state distribution from last 100k iterations)
    Sample action $d \sim \sigma(\omega, \Omega)$
    Compute immediate reward $u_m \leftarrow u(d; \omega, \Omega)$
    Transition to next state: $(\omega', \Omega') \sim P(\cdot|\omega, \Omega, d)$
    Compute TD target: $y_m \leftarrow u_m + \beta \max_{d'} Q_\theta(\omega', \Omega', d')$
    Compute TD error: $\delta_m \leftarrow Q_\theta(\omega, \Omega, d) - y_m$
    $\text{BE}_{\text{total}} \leftarrow \text{BE}_{\text{total}} + \delta_m^2$
end for
return $\text{BE} = \text{BE}_{\text{total}}/M$

We compute this every 10,000 iterations with $M = 1{,}000$ samples and track the time series $\{\text{BE}^k\}$. Convergence is declared when $\text{BE}^k < 10^{-4}$ for 100 consecutive checks.

B.2 Nash Gap Calculation

The Nash gap computation requires solving single agent optimization problems. We use the following procedure:

Algorithm 4 Nash Gap Computation

Require: Current strategy profile σ^k, set of agent types $\mathcal{T}$

Initialize NashGap $\leftarrow 0$
for each agent type $\tau \in \mathcal{T}$ **do**
    Freeze opponent strategies: $\sigma_{-\tau} \leftarrow \sigma^k_{-\tau}$
    Compute aggregate transition $P_\tau(\Omega'|\Omega, \sigma_{-\tau})$ via simulation
    Initialize Q table for type τ: $Q^{\text{BR}}_\tau \leftarrow Q^k_\tau$ (warm start)
    for $t = 1$ to $T_{\text{BR}} = 100{,}000$ **do**
        Update Q^{BR}_τ using Q-learning with $\epsilon = 0$ (pure exploitation)
        Aggregate state transition governed by $\sigma_{-\tau}$ (fixed)
    end for
    Compute best response value: $V^{\text{BR}}_\tau \leftarrow \mathbb{E}_{\omega,\Omega}\left[\max_d Q^{\text{BR}}_\tau(\omega, \Omega, d)\right]$
    Compute current value: $V^k_\tau \leftarrow \mathbb{E}_{\omega,\Omega}\left[\max_d Q^k_\tau(\omega, \Omega, d)\right]$
    Value gain: $\Delta V_\tau \leftarrow V^{\text{BR}}_\tau - V^k_\tau$
    NashGap $\leftarrow$ NashGap $+ \max\{0, \Delta V_\tau\}$
end for
return NashGap

In practice, we approximate the expectation using Monte Carlo sampling over 500 state action pairs. The Nash gap threshold is set to 1% of the average value function magnitude.

B.3 Spectral Radius Estimation

To estimate $\rho(J)$ at an equilibrium (π^*, η^*), we use finite difference approximations:

$$\frac{\partial \pi'}{\partial \pi}\Big|_{(\pi^*,\eta^*)} \approx \frac{\mathcal{T}_1(\pi^* + h, \eta^*) - \mathcal{T}_1(\pi^* - h, \eta^*)}{2h}, \tag{89}$$

with $h = 10^{-4}$, and similarly for other entries. We compute $\mathcal{T}_1(\pi + h, \eta)$ by:

1. Perturbing the state to $(\pi + h, \eta)$
2. Simulating $N = 1{,}000$ agents forward for $T = 100$ periods
3. Recording the average next period adoption $\bar{\pi}'$
4. Repeating 20 times and averaging to reduce Monte Carlo noise

The Jacobian matrix is then

$$\widehat{J} = \begin{bmatrix} \frac{\mathcal{T}_1(\pi^*+h,\eta^*) - \mathcal{T}_1(\pi^*-h,\eta^*)}{2h} & \frac{\mathcal{T}_1(\pi^*,\eta^*+h) - \mathcal{T}_1(\pi^*,\eta^*-h)}{2h} \\ \frac{\mathcal{T}_2(\pi^*+h,\eta^*) - \mathcal{T}_2(\pi^*-h,\eta^*)}{2h} & \frac{\mathcal{T}_2(\pi^*,\eta^*+h) - \mathcal{T}_2(\pi^*,\eta^*-h)}{2h} \end{bmatrix}. \tag{90}$$