

# Game-Theoretic Simulations Meet AI: Fast Policy Recommendations Under Data Scarcity (Student Abstract)

Phuong-Anh Duong<sup>1</sup>, Luong-Ngoc-Anh Vu<sup>1</sup>, Binh Nguyen Thi<sup>1</sup>

<sup>1</sup>Foreign Trade University, Hanoi City, Vietnam  
k61.2212140005@ftu.edu.vn, k61.2212150028@ftu.edu.vn, ntbinh@ftu.edu.vn

## Abstract

We tackle the challenge of setting smart prices and advertising budgets for dairy products sold through three retail channels—General Trade (GT), Modern Trade (MT), and E-commerce (EC)—in markets where data is scarce. Traditional economic models can capture the complex relationships between price, trust, and advertising, but solving these models every time a manager wants to ask a question is slow and impractical. Our solution is to turn a game-theoretic market simulator into training data: we generate 10,000 market scenarios, solve for the best pricing and ad strategies, and train an AI to imitate those decisions. On unseen scenarios, the AI remains accurate (price RMSE  $\leq 0.047$ , ad RMSE  $\leq 0.031$ ) and economically sound (Ratio of Means = 1.0010, Regret = 0.40%). To make it easy to use, we add a simple natural-language interface: users can say things like “trust is low, ad cost is high,” and the system returns price and ad suggestions along with confidence ranges. This creates a practical bridge between AI and economic rigor—delivering defensible decisions in seconds, even when data is limited.

## Introduction

In emerging markets, the retail milk sector is evolving through three competing sales channels: General Trade (GT), Modern Trade (MT), and E-commerce (EC). GT offers broad reach but weak controls, MT provides quality at a premium, and EC enables personalized marketing with limited physical presence (NIQ 2025). The competitive dynamics is necessitates comprehensive research.

Milk is a trust-sensitive product—consumer demand hinges on perceived authenticity and safety (Hao et al. 2022). Cross-channel spillovers and substitution effects make this a textbook case of strategic interdependence (Tsay and Agrawal 2004). Game-theoretic models are well-suited to capture these dynamics, but solving them repeatedly is computationally intensive. Data scarcity, especially in fragmented or low-tech markets, adds another layer of difficulty.

We propose a hybrid framework that combines structural economic modeling with lightweight AI. A market simulator captures trust, substitution, and ad spillovers dynamics. It generates synthetic equilibria to train a policy model that recommends near-optimal pricing and advertising—without

re-solving the game. A natural language interface maps qualitative cues (e.g., “low trust”) to parameter, enabling real-time decision support even in data-poor settings.

## Methodology

We address data-scarce multi-channel pricing-advertising by converting a structural game-theoretic model, grounded in prior literature with context-specific trust, substitution, and ad spillover effects, into a simulator-driven learning pipeline (Tsay and Agrawal 2004; Miller, Remer, and Sheu 2013). By uniformly sampling interpretable market states and solving for near-Nash ( $P^*, a^*$ ), we generate (*state, action, outcome*) triplets to train a lightweight *constrained* policy. Predicted prices are clipped to cost-relative bounds and ads to  $[0, 20]$ , allowing instant recommendations without re-solving the game. This method has been used in many economics papers (Bernasconi et al. 2023; Zheng et al. 2022).

To support real-world use, we add a natural-language interface: qualitative queries (e.g., “trust = low, ad cost = high”) are mapped to parameter intervals, from which we draw  $M$  i.i.d. states by sampling each parameter independently from a uniform distribution over its interval (simple random sampling; no MCMC). The policy outputs *median* and 5–95% action ranges, offering robust guidance under soft constraints even without precise inputs.

Economic quality is evaluated by re-scoring predicted actions in the structural model and comparing to simulator optima using robust metrics: ratio of means, trimmed ratio, and mean regret over high-profit rows.

## Structural Model (From Prior Literature with Context-Specific Modifications)

For  $i \in \{GT, MT, EC\}$ , demand is isoelastic with cross-price effects, trust, and ad spillovers:

$$Q_i = \alpha_i P_i^{-\beta_i} \prod_{j \neq i} \left( \frac{P_j}{P_i} \right)^{\gamma_{ij}} \times (1 + \kappa_i a_i)^{\psi_i} \left( 1 + \sum_{j \neq i} \sigma_{ij} a_j \right) \tau_i(T)$$

where

$$\tau_i(T) = \frac{T^{p_i}}{T^{p_i} + \theta_i^{p_i}}$$

The terms are chosen to reflect milk retail realities:  $\tau_i(T)$  is an S-shaped *trust* multiplier with channel-specific  $(p_i, \theta_i)$  because demand for dairy in online (EC) tends to be more trust-elastic than offline (GT/MT); the multiplicative advertising block captures diminishing returns to own ads and spillovers across channels; cross-price terms encode observed substitution between GT, MT, and EC when relative prices change; and costs are asymmetric across channels— $c_i = w_i + l_i$  for GT/MT (wholesale plus logistics) versus  $c_{EC} = w_{EC}/(1 - \mu_{EC})$  for EC (platform commissions).

**Per-Channel Profit:**  $\Pi_i(P_i, a_i) = (P_i - c_i)Q_i - \frac{a_i^2}{2\chi_i}$

**Structured Costs:** GT/MT:  $c_i = w_i + l_i$ ; EC:  $c_{EC} = \frac{w_{EC}}{1 - \mu_{EC}}$ .

**Price Best Response:**

$$P_i^* = \frac{\eta_i}{\eta_i - 1} c_i \quad \text{for } \eta_i = \beta_i + \sum_{j \neq i} \gamma_{ij} > 1$$

Prices are clipped to  $P_i \in [m_{lo}c_i, m_{hi}c_i] = [1.05, 6]c_i$ .

**Advertising Best Response:**

$$A_i \psi_i \kappa_i (1 + \kappa_i a_i)^{\psi_i - 1} = \frac{a_i}{\chi_i}$$

where

$$A_i = (P_i - c_i) \alpha_i P_i^{-\beta_i} \prod_{j \neq i} \left( \frac{P_j}{P_i} \right)^{\gamma_{ij}} \times \left( 1 + \sum_{j \neq i} \sigma_{ij} a_j \right) \tau_i(T)$$

We iterate a small number of passes across channels to reach a near-Nash fixed point.

### Synthetic Data Generation

We simulate 10,000 independent market states by uniformly sampling parameters: trust level  $T \in [0, 1]$ ; baseline demand scale  $\alpha_i \in [0.8, 1.2]$ , own-price elasticity  $\beta_i \in [0.8, 1.2]$ , advertising effectiveness  $\kappa_i \in [0.8, 1.2]$ ; diminishing-returns exponent for own advertising  $\psi_i \in [0.3, 0.7]$ , ad cost-efficiency parameter  $\chi_i \in [0.8, 1.5]$ ; trust-curve steepness  $p_i \in [2, 6]$ , trust inflection point  $\theta_i \in [0.3, 0.7]$ ; cross-price elasticity (from  $j$  to  $i$ )  $\gamma_{ij} \in [0.3, 0.7]$ , advertising spillover (from  $j$  to  $i$ )  $\sigma_{ij} \in [0.3, 0.7]$ ; wholesale cost  $w_i \in [0.6, 1.2]$ ; logistics/retailing cost  $l_{GT}, l_{MT} \in [0.1, 0.6]$ ,  $l_{EC} = 0$ ; platform commission rate (EC)  $\mu_{EC} \in [0.05, 0.25]$ ,  $\mu_{GT} = \mu_{MT} = 0$ ; and advertising effort  $a_i \in [0, 20]$ .

Cost structure follows the ‘‘Structural Model’’ specification. For each sampled state, we compute optimal decisions  $(P_i^*, a_i^*)$ , resulting demands  $Q_i$ , and profits  $\Pi_i$ , then log the complete state–action–outcome tuples for training.

### Policy Model Training

We use an 80/10/10 train/validation/test split on the simulated dataset. Features are standardized (targets left unscaled). We fit a lightweight MLP with ReLU activations, two hidden layers of sizes (256, 128), and the Adam optimizer (learning rate  $3 \times 10^{-4}$ ,  $\text{max\_iter} = 1000$ ). Early

stopping is applied using `sklearn`’s internal 10% validation and patience 10. At inference, the trained model produces instant, feasible recommendations without re-solving the game-theoretic equilibrium.

### Natural-Language Interface

To support practitioner inputs, we add a rule-based language-to-state encoder: qualitative levels (e.g., trust = low, substitutability = high, ad\_cost = high) are mapped to numeric *intervals* consistent with our simulation priors. At query time, we draw  $M$  i.i.d. states by sampling each constrained parameter independently from a uniform distribution over its specified interval (simple random sampling; no MCMC, seed = 2025), then apply the scaler and the constrained policy to return median and 5–95% bands by channel. This preserves real-time operation and communicates uncertainty when only coarse inputs are available. Optionally, a learned text-to-state encoder can replace the rules using synthetic descriptions, but the pipeline remains unchanged.

## Results and Discussion

**Action Accuracy** On the test set, RMSEs for predicted prices are: GT **0.0472**, MT **0.0471**, EC **0.0421**; and for ads: GT **0.0214**, MT **0.0245**, EC **0.0309**. These errors are small relative to the feasible action ranges (prices tied to cost-relative bounds; ads in  $[0, 20]$ ), indicating strong fidelity to near-Nash equilibria.

**Economic Quality** Evaluating predicted actions in the structural simulator yields: *Ratio of Means* = 1.0010, *Trimmed Ratio* (5–95%) = 0.9992, *Mean Regret* = 0.40%, and *Average Profit Gap* = 0.0095. The plain mean of per-row profit ratios is unstable and therefore not used. Negative-profit shares are 0.0% for optimal actions vs. 1.6% for predicted ones, which can be mitigated via post-hoc penalties or quantile-aware loss.

**Practical Insight** The learned policy delivers *instant*, constraint-aware pricing and advertising recommendations that closely approximate simulator-optimal behavior. A lightweight natural-language interface maps qualitative inputs to parameter intervals, draws  $M$  i.i.d. states by sampling each parameter independently from a uniform distribution over its interval (simple random sampling; no MCMC), and returns median and 5–95% bands for real-time decisions under uncertainty.

## Conclusion

We bridge structural economics and AI: a transparent game-theoretic simulator provides supervision, and a lightweight constrained policy imitates pricing–advertising equilibria to deliver fast, feasible, interpretable recommendations under data scarcity. A natural-language interface maps qualitative inputs to parameter ranges for intuitive what-if analysis without re-solving. The approach suits settings where trust, spillovers, and substitution interact but logs are limited. Future work can focus on dynamic/Stackelberg variants, conservative offline RL, and calibration with partial real data.

## References

- Bernasconi, M.; Vittori, E.; Trovò, F.; and Restelli, M. 2023. Dealer markets: A reinforcement learning mean field game approach. *The North American Journal of Economics and Finance*, 68: 101974–101974.
- Hao, N.; Zhang, Y.; Zheng, Q.; and Wetzstein, M. 2022. Spillover effects of food recalls: A milk recall scenario experiment in China. *npj Science of Food*, 6.
- Miller, N. H.; Remer, M.; and Sheu, G. 2013. Using cost pass-through to calibrate demand. *Economics Letters*, 118: 451–454.
- NIQ. 2025. Winning Vietnam’s Tet holiday: Optimizing FMCG growth for the next season.
- Tsay, A. A.; and Agrawal, N. 2004. Channel Conflict and Coordination in the E-Commerce Age. *Production and Operations Management*, 13: 93–110.
- Zheng, S.; Trott, A.; Srinivasa, S.; Parkes, D. C.; and Socher, R. 2022. The AI Economist: Taxation policy design via two-level deep multiagent reinforcement learning. *Science Advances*, 8.