Dynamic Learning in Strategic Pricing Games

John R. Birge
Matt Stern
The University of Chicago
Booth School of Business
How to Price and Learn?

• Situation

  – New products (e.g., games, insurance, drugs)
  – Unlimited capacity (e.g., digital goods)
  – Multiple (possibly differentiated) suppliers

• Questions

  – What is optimal (equilibrium) pricing behavior and how is it different from the monopoly setting?
  – What is the impact on consumers, social welfare, and intermediaries?
Outline

• Monopoly pricing
• Collusion, competition, and oligopoly results
• Results from simple policies
• Regret due to learning and influence
• Strategic incomplete learning in simple policies
• “Willful ignorance” in prisoner’s dilemma
• Implications
Monopoly Pricing Results

- Monopolist needs to vary price sufficiently to learn demand
- Monopolist can learn a demand function with:

\[ \text{Regret}(T, \theta, \sigma) = O(\sqrt{T \log T}) \]

where \( \text{Regret}(T, \theta, \sigma) = T\pi^* - E[\sum_{t=1}^{T} \pi_t(\theta, \sigma)] \)

\( \theta \) are parameters, \( \sigma \) is policy, and \( \pi_t \) is revenue with maximum expected value \( \pi^* \).

(Keskin and Zeevi (2013))
Price Optimization in Auto Insurance

- 72% of firms consider competitor’s prices when setting prices
- 26% of firms currently employ price optimization (45% over $1B in revenue)
- An additional 36% plan to start in the near future

Figure 8: Estimating the overall effect of a rate change by company size

© Earnix 2013
Insurance Commissioner v. Engelman

- “The Maryland Insurance Administration (MIA) has determined that the use of price optimization results in rates that are unfairly discriminatory”

- “the use of price optimization may result in two insureds with like risk characteristics being charged different premiums, which is a violation of 27-212(e)(1) of the Insurance Article.”
Results from Limiting Price Variation

• Myopically optimizing prices and estimating demand can lead to incomplete learning and poor prices

• Example (Lai and Robbins 1982):

\[ q_0 \]
\[ q_1 \]
\[ q_2 \]

\[ U=p_1=p_2=\ldots \]
Questions for Competition

• Can sellers in competition learn as efficiently as monopolists?
• Is efficient learning always an optimal strategy?
• How much can optimal equilibrium strategies with competition differ from optimal monopolistic pricing strategies?
New Markets and Price Impact Uncertainty

• New consumer products?

• New drugs?
Relationship to Pricing Collusion

Results

• Stigler (1964) – can maintain collusive prices with mechanism to detect secret price-cutting
• Maskin/Tirole (1988) – can maintain collusion with kinked demand curves
• Cooper et al. (2013) – may obtain collusion by assuming monopoly model
• Quantity/Prisoner’s Dilemma (Kreps et al. (1982), Green/Porter (1984), Abreu/Pierce/Stacchetti (1986), Sannikov/Skrzypacz (2007), Rahmann (2015))
Overview of Results

• General impact of uncertainty in previous models:
  – Makes “cheater” detection harder and is costly for enforcement
  – Effective monitoring necessary to enforce collusion

• Impact of uncertainty with observed actions, unobserved payoffs, and unknown parameters:
  – Uncertain and limited observations may limit learning and lead to collusive outcomes
  – Collusion becomes more likely with limited price ranges and deviations
Setup for Investigation

• Model:

  Two firms face noisy linear demands over $T$ periods

  $d_{it} = \alpha_i + \beta_i p_{it} + \gamma_i p_{-it} + \epsilon_{it}$

  Estimate the parameters via OLS: $\hat{\theta}_{it} = [a_{it}, b_{it}, c_{it}]$

  Forecast the price of their competitor: $p_{-it}^{e}$

  Choose prices: $p_{it} \in [l, u]$, $0 < l < u < \infty$

• Simple candidate policy: best-response plus noise
Classifying Learning: Notions of Regret

\[ \Delta_i(T) = \sum_{t=1}^{T} E[p_i^{NE}(\alpha_i + \beta_i p_i^{NE} + \gamma_i p_{-i}^{NE}) - p_{it}(\alpha_i + \beta_i p_{it} + \gamma_i p_{-it})] \]

Learning: \[ \sum_{t=1}^{T} E[(p_i^{NE} - p_{it})^2] \]

Influence: \[ \sum_{t=1}^{T} E[p_{it} (\mathcal{B}\mathcal{R}_\theta(p_{-i}^{NE}) - \mathcal{B}\mathcal{R}_\theta(p_{-it}))] \]
Myopic (No Noise) Result

Best Response:

$$BR_\theta(p^e_{-it}) = \frac{a_{it}}{-2b_{it}} + \frac{c_{it}}{-2b_{it}} p^e_{-it}$$

Competitive Equilibrium

Figure: Myopic strategies fail to learn demand
Measure Learning

Let $x_t := [1 \ p_{it} \ p_{-it}]^T$ is the public price vector.

- Fisher Information Matrix:

$$
\mathcal{J}_t := \sum_{s=1}^{t} x_s x_s^T = \sum_{s=1}^{t} \begin{bmatrix}
1 & p_{is} & p_{-is} \\
p_{is} & p_{is}^2 & p_{is} p_{-is} \\
p_{-is} & p_{is} p_{-is} & p_{-is}^2
\end{bmatrix}
$$

- Firm $i$'s OLS estimates are

$$
\hat{\theta}_{it} = \mathcal{J}_t^{-1} \left( \sum_{s=1}^{t} d_{is} x_s \right)
$$

- Estimation Error:

$$
\hat{\theta}_{it} - \theta_i = \mathcal{J}_t^{-1} \sum_{s=1}^{t} \varepsilon_{is} x_t
$$
Information and Error Bounds

Lemma (Keskin and Zeevi 2014)

If the demand shocks \( \{\epsilon_{is}\}_{i=\pm 1, s=1, \ldots, t} \) have light-tailed distributions, then there exist finite positive constants \( \rho \) and \( k \) such that,

\[
P \left\{ ||\hat{\theta}_it - \theta_i|| > \delta, \lambda_{\min}(J_t) \geq m \right\} \leq kt \exp \left( -\rho (\delta \wedge \delta^2) m \right),
\]

for all \( \delta, m > 0 \) and all \( t \geq 3 \).

- Since \( J_t \) is public, firms know (with high probability) how close their estimates are to the true value.

- **Question:** How can we grow the minimum eigenvalue of \( J_t \)?
Best Response with Random Dithering:

\[ p_{it} = \mathcal{BR}_{\hat{\theta}}(p_{-it}^e) + \nu_{it} \]

\( \nu_{it} \) conditionally independent, mean zero noise.
Learning Result

Theorem

Under best-response pricing with random dithering, the worst-case regret due to learning is $O(\sqrt{T \log T})$ for both players under the following assumptions:

- Cournot Adjustment: $p_{-i(t-1)}^e = p_{-i(t-1)}$
- $\text{Var}(\nu_{it}) = O\left(\frac{1}{\sqrt{t}}\right)$

• Implication and question:
Can achieve (close to) the monopolist’s learning rate with competitors.
But, is this optimal?
Key Result for Proof

Theorem (Matrix Freedman, Trop 2011)

Consider a finite series $Z = \sum_s Y_s$ of adapted, self-adjoint matrices with dimension $d$ such that

$$\mathbb{E}_{s-1} Y_s = 0 \text{ and } \lambda_{\max}(Y_s) \leq R.$$ 

Let $\nu(Z) := \sum_s \mathbb{E}_{s-1}(Y_s^2)$

Then for all $\delta \geq 0$,

$$\mathbb{P} \left\{ \lambda_{\min}(\sum_s Y_s) \leq -\delta \right\} \leq d \cdot \exp \left( \frac{-\delta^2/2}{\nu(Z) + R\delta/3} \right).$$
Use of Key Result

Applying the matrix Freedman inequality,

$$\mathbb{P} \{ \lambda_{\min}(\mathcal{J}_t) \leq \lambda_{\min}(\mathbb{E}_{s-1}\mathcal{J}_t) - \delta \} \leq 3 \exp \left( \frac{-\delta^2/2}{\nu(\mathcal{J}_t) + R\delta/3} \right)$$

- Show:
  - $\lambda_{\min}(\mathbb{E}_{s-1}\mathcal{J}_t) = O(\sqrt{t})$
  - $\nu(\mathcal{J}_t) = O(\sqrt{t})$

- Choose:
  - $\delta = \sqrt{t}$
Is Learning Optimal?

• With limited noise, may improve outcomes:

Higher revenues for Product 2

Static Nash equilibrium
Is Incomplete Learning Optimal for General Strategies?

• Simplified Case with Two Players:

Consider a Prisoner’s Dilemma game with deterministic payoffs.

\[
\begin{array}{cc}
\rho_0 & \rho_1 \\
p_0 & (R_k, R_k) & (T_k, S_k) \\
p_1 & (S_k, T_k) & (P_k, P_k) \\
\end{array}
\]

\[T_k > R_k > P_k > S_k, \quad k = 1, 2\]

• Common prior:

\[\pi := Pr(\text{Scenario} = 1)\]

• Uncertainty between cooperation or competition:

\[R_1 = P_2\]
Game Setup

\[
\begin{align*}
p_0 & \quad (R_k, R_k) & \quad p_1 & \quad (T_k, S_k) \\
p_0 & \quad (S_k, T_k) & \quad p_1 & \quad (P_k, P_k)
\end{align*}
\]

\[T_k > R_k > P_k > S_k, \quad k = 1, 2\]

- Represents pricing game where \(p_0\) and \(p_1\) are the competitive and cooperative equilibrium prices.
- Assume that the firms choose \((p_1, p_1)\) in the first period and observe \(R_1 = P_2\) payoffs.
- Choice: Continue charging \((p_1, p_1)\) or deviate and learn the game.
Game Reduction

Result:
The $T$ period game can be reduced to a single period game by incorporating the future value into the current beliefs.

Expected Payoff Parameters:
- $x_{\pi}$, the expected payoff to firm $i$ with beliefs $\pi$ of $(p_1, p_1)$
- $y_{\pi}$, the expected payoff to firm $i$ with beliefs $\pi$ of $(p_1, p_0)$.

\[
\begin{align*}
x_{\pi} & := (1 - \pi)(P_2 - S_2) + \pi(R_1 - T_1) \\
y_{\pi} & := (1 - \pi)(T_2 - R_2) + \pi(S_1 - P_1)
\end{align*}
\]
Stable Dynamics of the Game

*Expected Payoff Parameters:*
- \( x_\pi \), the expected payoff to firm \( i \) with beliefs \( \pi \) of \((p_1, p_1)\)
- \( y_\pi \), the expected payoff to firm \( i \) with beliefs \( \pi \) of \((p_1, p_0)\)
Alternate Stable Dynamics

*Expected Payoff Parameters:*

- $x_\pi$, the expected payoff to firm $i$ with beliefs $\pi$ of $(p_1, p_1)$
- $y_\pi$, the expected payoff to firm $i$ with beliefs $\pi$ of $(p_1, p_0)$.

![Diagram](chart.png)
Result of Simple Game

• Willful ignorance is optimal (cooperative equilibrium) if the firms observe the uninformative revenue outcomes
• Firms choose not to experiment and learn the full market information
• Extends to cases with uncertainty in the outcomes
Implications for Dynamic Pricing

• Firms may prefer not to experiment with prices, particularly for new products where products may be complements or substitutes.
• Examples in digital goods and pharmaceuticals.
• Uncertainty makes observation and enforcement difficult but can also ensure collaboration.
Policies for Pricing with Competition

• Use initial priors for optimal prices
• Base decisions to vary prices and learn on support for alternatives
• Issues for government policy:
  Caution for regulations restricting pricing adjustments (e.g., auto insurance in Maryland)
Conclusions

• Simple best-response policies with noisy adjustments can yield efficient learning as in monopolistic pricing

• Learning policies may not be optimal within the class of best-response and adjustment policies

• In general policies, learning may not be optimal for firms

• Policies restricting price adjustments can reduce consumer value
Additional Issues

• How extensive is the set of situations in which firms may choose not to learn the market?
• What should happen in markets with nonstationary behavior?
• What role do intermediaries play in ensuring learning and efficiency?
• How should policy be used to ensure efficiency and maximum social welfare?
Thank you!

Questions?