Quota Distortions in General Equilibrium

David Baqaee and Kunal Sangani

January 2025

Quota Distortions

- Standard approach models misallocation using implicit taxes or "wedges."
  Harberger (1954), Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Baqaee and Farhi (2020).
- Wedges natural in some contexts (ad valorem taxes, markups, sticky prices).
- However, in other contexts, distortions apply directly to quantities.
  - Government quotas (import quotas/licenses, taxicab medallions, land ceilings).
  - Missing markets (land markets, credit markets, insurance markets).
  - Size-based penalties (based on e.g., number of employees).
  - Cournot competition (producers directly implement quantity decisions).
- This paper: A general framework for analyzing economies with quota-like distortions.

Preview of Results

- Result: Any feasible allocation can be implemented as the decentralized equilibrium of an economy with quotas.
- Implication 1: Can study any distorted allocation of resources using (implicit) quotas.
- Implication 2: Can use tools for efficient economies (e.g., Hulten's theorem) to analyze economies with quota distortions.
  - Tractable comparative statics. Low information requirements when quotas are primitives.
  - Can characterize distance to frontier, nonlinearities even far from the frontier.

Preview of Results

- Response of output to quota changes and productivity shocks.
    - Key statistic: Profits earned by quota holders.
    - Don't need elasticities of substitution / production, full input-output structure, etc.
    - Examples: Relaxing H1-B visa cap, zoning restrictions on single-family housing.

Preview of Results

- Response of output to quota changes and productivity shocks.
  - Key statistic: Profits earned by quota holders.
  - Don't need elasticities of substitution / production, full input-output structure, etc.
  - Examples: Relaxing H1-B visa cap, zoning restrictions on single-family housing.
- Distance to the frontier.
  - In terms of profits and size of distortion, $\approx 1/2 \times$ **profits** $\times$ **quantity distortion**.
  - Alternatively, $\approx 1/2 \times$ **profits** $/$ **elasticity of profits to quota changes**.
- Nonlinearities even far from the frontier.
  - How quotas change profits determine whether output is concave / convex in shocks.
  - Example: NYC taxicab medallions.

Table of Contents

General Framework

- Representative household, N goods indexed by i, F factors indexed by f.
- Real output Y maximizes constant-returns aggregator $\mathscr{D}$,

$$Y = \max_{\{c_1,\ldots,c_N\}} \mathscr{D}(c_1,\ldots,c_N),$$

subject to the budget constraint,

$$\sum_i^N p_i c_i = \sum_{f=1}^F w_f L_f + \sum_{i=1}^N \Pi_i,$$

where c_i final demand, p_i prices, L_f factor supplies, w_f wages, and Π_i profits.

General Framework: Quotas

- Each good i produced using constant returns production technology

$$y_i = A_i F_i(x_{i1}, \dots, x_{iN}, L_{i1}, \dots, L_{iF}),$$

where x_{ij} is use of intermediate good j, L_{if} use of factor f, and A_i productivity shifter.

- A quota restricts the production of good i at a quantity y_{i^*},

$$y_i = \min\{y_{i^*}, A_i F_i(x_{i1}, \dots, x_{iN}, L_{i1}, \dots, L_{iF})\}.$$

- Profits for producers of i are revenues less intermediate and factor costs,

$$\Pi_i = p_i y_i - \sum_{i=1}^{N} p_j x_{ij} - \sum_{f=1}^{F} w_f L_{if}.$$

Equilibrium

- Given quotas y_{i^*}, productivities A_i, production functions F_i, and factor supplies L_f,
- An equilibrium consists of prices p_i, wages w_f, outputs y_i, final demands c_i, and intermediate / factor input choices x_{ij} and L_{if} such that:
  - Final demands c_i maximizes real output subject to the budget constraint.
  - Each producer minimizes costs taking prices as given.
  - For all goods with quotas, $y_i \leq y_{i^*}$.
  - Resource constraints satisfied:

$$c_i + \sum_{j=1}^{N} x_{ji} \leq y_i \text{ for all } i \qquad \text{and} \qquad \sum_{i=1}^{N} L_{if} \leq L_f \text{ for all } f.$$

Implementing an Allocation Using Quotas

Definition (Feasible allocation)

An allocation $(\{c_i\}, \{x_{ij}\}, \{L_{if}\})$ is feasible if:

- c_i, x_{ij}, and L_{if} are non-negative for all i, j, and f,
- $y_i \leq A_i F_i(x_{i1}, \ldots, x_{iN}, L_{i1}, \ldots L_{iF})$ for all i,
- Resource constraints are satisfied.

Proposition

Consider some feasible allocation $\mathscr{X}$. Then:

1. *there exists a vector quotas, $\{y_{i^*}\}$, such that the decentralized eqm. has allocation $\mathscr{X}$*
2. *given these quotas, the allocation $\mathscr{X}$ is efficient.*

Implementing an Allocation Using Quotas

Proposition

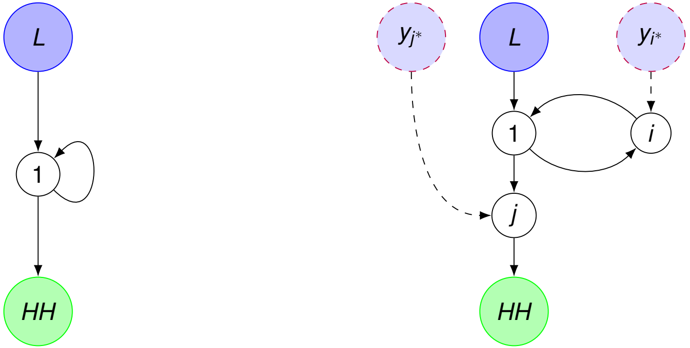
Suppose an allocation $\mathscr{X}$ is feasible. Then:

1. *There is an economy with quotas in which the decentralized eqm. has allocation $\mathscr{X}$.*
2. *Given these quotas, the allocation $\mathscr{X}$ is efficient.*

- Add nodes/quotas to guarantee that competitive eqm. yields desired allocation.
- First Welfare Theorem implies allocation is constrained efficient.
- Implication 1: Recast any distorted allocation as eqm. of economy with implicit quotas.
- Implication 2: Analyze eqm. using tools for efficient economies (e.g., Hulten's Thm).

Implementing an Allocation Using Quotas: Example

- Round-about economy. Feasible allocations: $\{(y_1, c_1, x_{11}) \mid c_1 + x_{11} \leq y_1 = F_1(L, x_{11})\}$.



(a) Original. (b) With quotas.

- In fact, more general than wedges: can implement allocation when L, x_{11} are perfect substitutes / complements.

Table of Contents

Comparative Statics

Proposition

To a first order, the effect of changes in quotas y_{i^*} and productivities A_i on output is

$$d\log Y = \sum_i \Pi_i d\log y_{i^*} + \sum_i (\lambda_i - \Pi_i)\, d\log A_i,$$

where λ_i and Π_i are sales and profits of i divided by GDP.

If all quotas are non-binding, then $d\log Y = \sum_i \lambda_i d\log A_i$.

- Profits of constrained producers are sufficient statistic for effect of quota changes.
- Removing a quota always improves welfare.
  - Conditional on other quotas, no Theory of Second Best.
  - If quotas adjust endogenously, Theory of Second Best returns.
- When all quotas are non-binding, profits are zero $\Rightarrow$ Hulten's Theorem.

Empirical Example 1: H-1B Visa Quota

- H-1B visa for high-skill foreign workers: 85,000 visas per year since mid-2000s.
- To a first order, gains from increasing quota equal to rents earned by visa winners:

$$d \log Y = \Pi_i d \log y_{i^*} \approx \frac{\Pi_i}{y_{i^*}} dy_{i^*}.$$

- Clemens (2013) compares earnings of winners vs. losers of 2007 H-1B lottery.
    - Earnings for workers who won lottery were \$12,641 higher two years after the lottery.
- Doubling number of visas in 2007 would have increased world output \$1.07B.

Empirical Example 2: Zoning Restrictions on Single-Family Housing

- Gains from easing zoning restrictions equal to profits earned by permit holders.
  - Gyourko and Krimmel (2021) isolate permit "rents" by comparing vacant parcels to nearby parcels with existing housing.

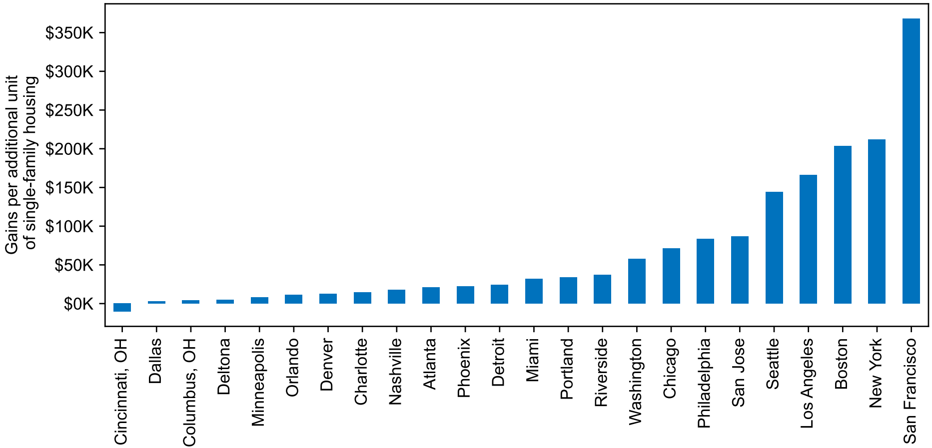


Table of Contents

Distance to the Frontier

Proposition

Let $\Pi_i(\mathbf{y}_*)$ be profits of producer i given the vector of quotas $\mathbf{y}_*$. The distance to the frontier to a second order is

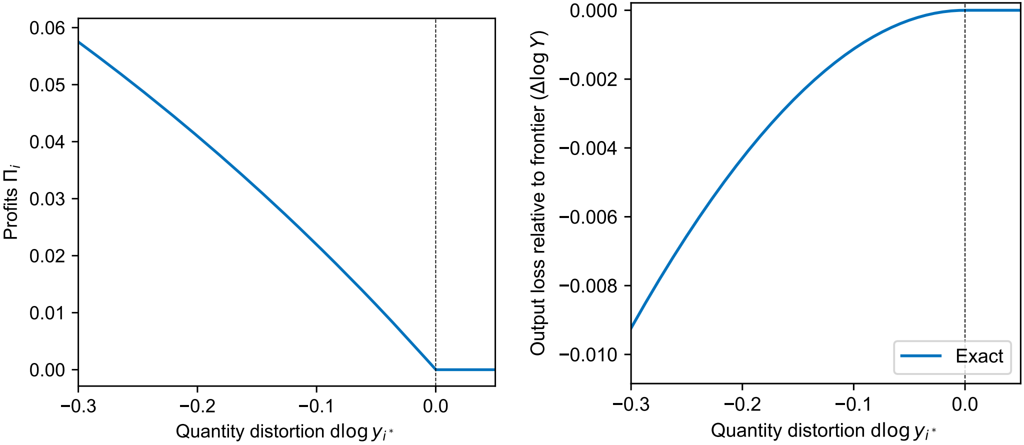
$$\Delta \log Y \approx \frac{1}{2} \sum_i \Pi_i(\mathbf{y}_*) \, d \log y_{i^*} \,,$$

where $d \log y_{i^*} = \log y_{i^*} - \log y_i^{eff}$ is the quantity distortion on producer i relative to its efficient level of production.

- Option 1: Estimate distance using $1/2 \times$ **profits** $\times$ **size of distortion**.
- Intuition: Average of first-order at inefficient point ($\Pi_i \, d \log y_{i^*}$) and at efficient point (0).
- Unlike wedges, note that we don't need to consider "interactions" between quotas.

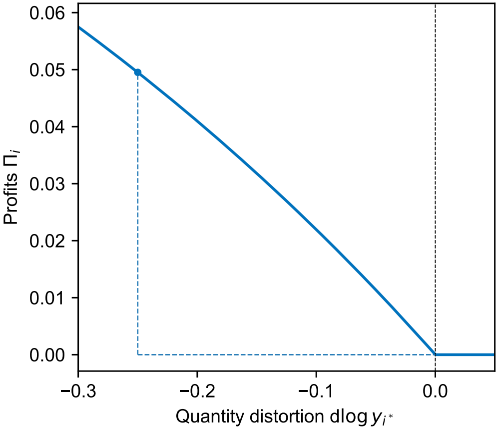
Distance to Frontier Example: Round-about Economy

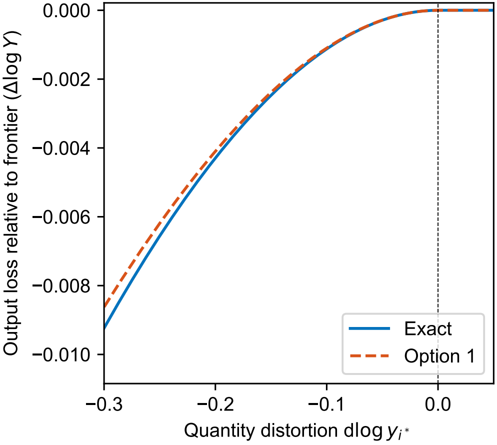
- Round-about economy with quota on the use of input x_{11}.



Distance to Frontier Example: Round-about Economy

- Option 1: $\Delta \log Y \approx 1/2 \times$ **profits** $\times$ **size of distortion** $= 1/2\, \Pi_i d \log y_{i^*}$.





Distance to the Frontier: Option 2

- Option 1 uses profits at distorted allocation: $\Delta \log Y \approx \frac{1}{2} \sum_i \Pi_i(\mathbf{y}_*)\, d \log y_{i^*}$.
- Option 2: Estimate distance to frontier using elasticities of profits to quotas.

Proposition

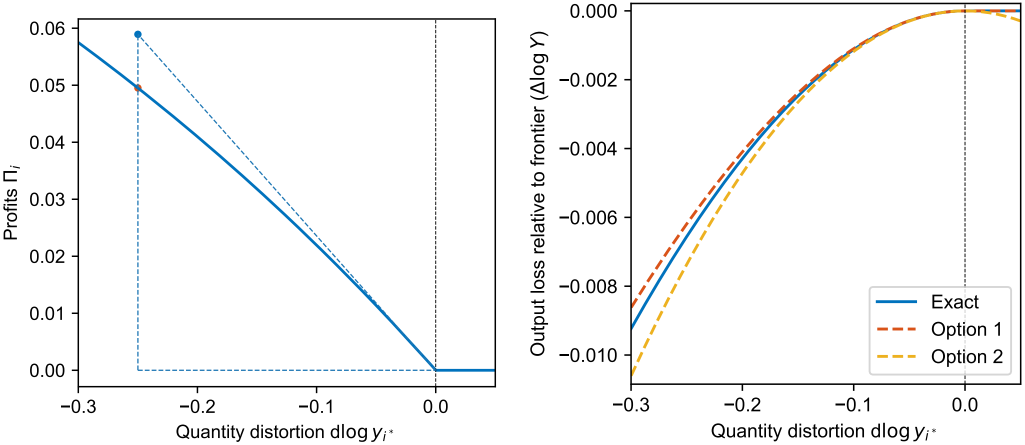
Equivalently, the distance to the frontier to a second order is

$$\Delta \log Y \approx \frac{1}{2} \sum_{i^*} \left[\sum_{k^*} \frac{\partial \Pi_i}{\partial \log y_{k^*}} d \log y_{k^*} \right] d \log y_{i^*}.$$

- Expresses distance in terms of input-output structure and elasticities of substitution.

Distance to Frontier Example: Round-about Economy

- Option 2: $\Delta \log Y \approx 1/2 \, \frac{\partial \Pi_i}{\partial \log y_{i^*}} (d \log y_{i^*})^2 = -\frac{1}{2\theta_1} \frac{\lambda_1 - 1}{\lambda_1} (d \log y_{i^*})^2$.



Distance to the Frontier

- Both formulas require knowing the efficient level of output $d \log y_{i^*} = \log y_{i^*} - \log y_i^{\text{eff}}$.

- Option 3: Estimate distance using **elasticity of profits to quota changes**.

Proposition

To a second order, the output gain from removing the quota y_{i^*} is

$$\Delta \log Y \approx \frac{1}{2} \Pi_i \left[- \frac{d \log \Pi_i}{d \log y_{i^*}} \right]^{-1}.$$

- Intuition: If profits fall quickly with output, close to efficient level $\Rightarrow$ smaller gains.

Distance to Frontier Example: Round-about Economy

- Option 3: Use elasticity of profits to quota: $\Delta \log Y \approx 1/2\, \Pi_i [-\frac{d \log \Pi_i}{d \log y_{i^*}}]^{-1}$.

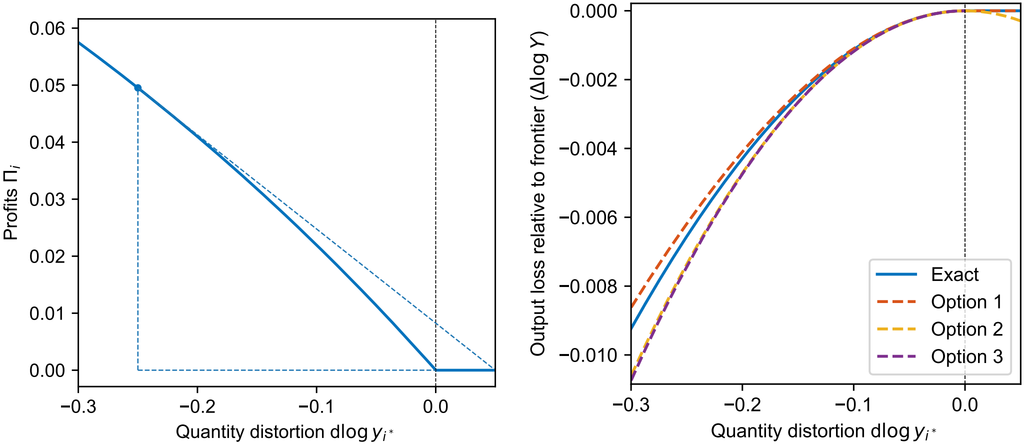


Table of Contents

Nonlinearities Far from the Frontier

Proposition

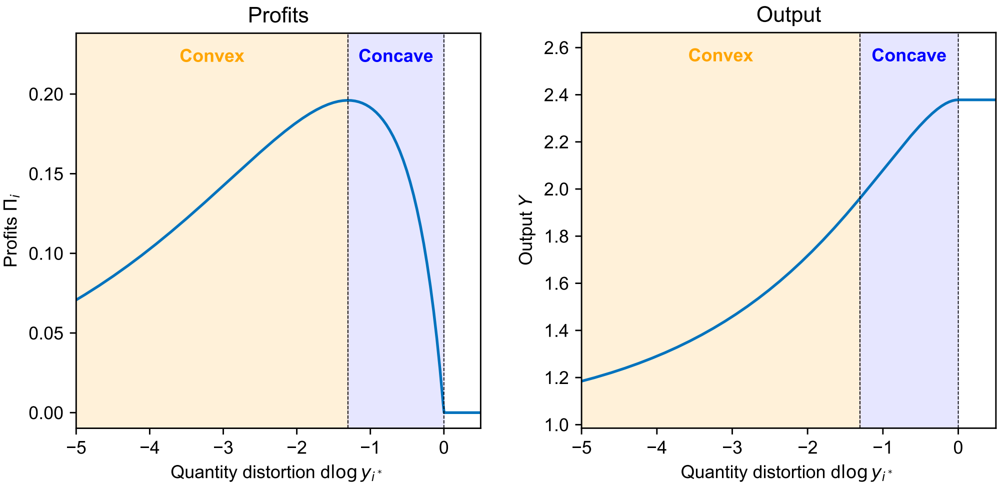
To a second order, the effect of a change in quota y_{i^*} on output is

$$\Delta \log Y \approx \Pi_i d \log y_{i^*} + \frac{1}{2} \frac{d\Pi_i}{d \log y_{i^*}} (d \log y_{i^*})^2.$$

- Profits $=$ income of a "fixed factor," quota changes $=$ shocks to its productivity.
  - Can use existing results to calculate elasticity of profits to quota. (Baqaee and Farhi 2019).
- Elasticity of profits to quota determines concavity / convexity:
  - Output always concave around efficiency ($\Pi_i = 0$).
  - Away from efficiency, may be convex (nonlinearities mitigate costs, amplify benefits).

Illustration: Nonlinearities in Horizontal Economy

- Second-order effect $d\Pi_i / d\log y_{i^*}$ depends on whether profits rising/falling in y_{i^*}.
- Always concave near efficient point. May be convex away from frontier.



Nonlinearities Far from the Frontier: Monopolist

- In general, computing non-linearities requires knowing production network, elasticities.
- Special case where monopolist chooses output quota to maximize real profits.
  - Low info requirements to calculate nonlinear effects of change in monopolist's output!

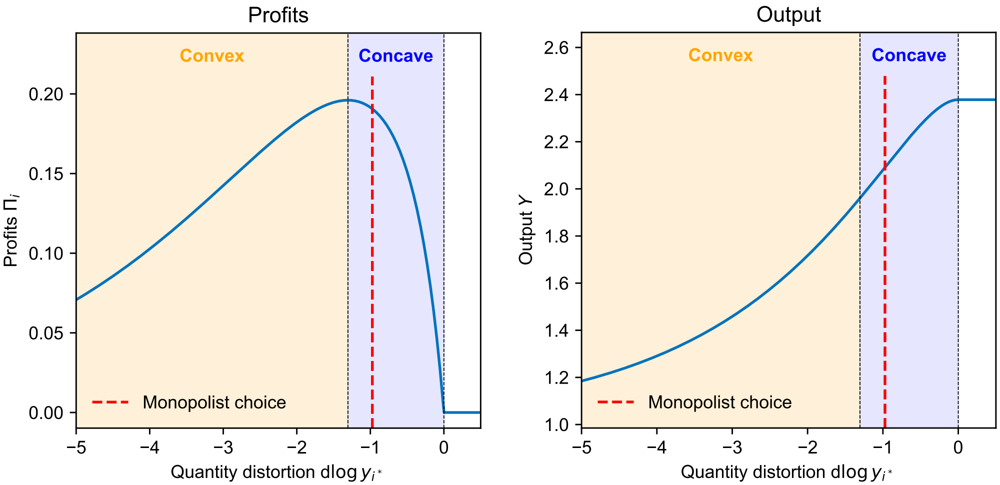
Proposition

Suppose producer i is a monopolist that chooses output quantity y_i to maximize real profits. Then, the effect of changes in the monopolist's quantity on output to a second order are

$$\Delta \log Y \approx \Pi_i d \log y_i - \frac{1}{2}\Pi_i^2 (d \log y_i)^2.$$

Illustration: Monopolist

- Monopolist always chooses quantity in concave region.
- As monopolist becomes infinitesimal, nonlinearities $\to$ zero.

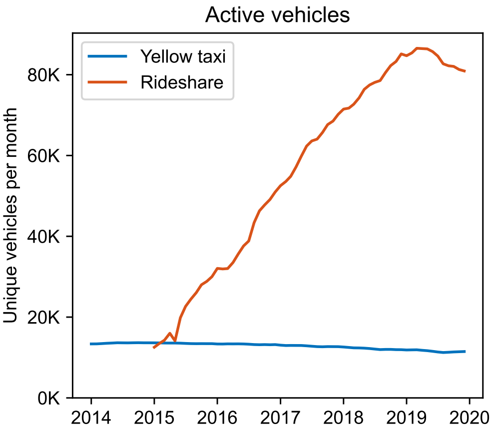


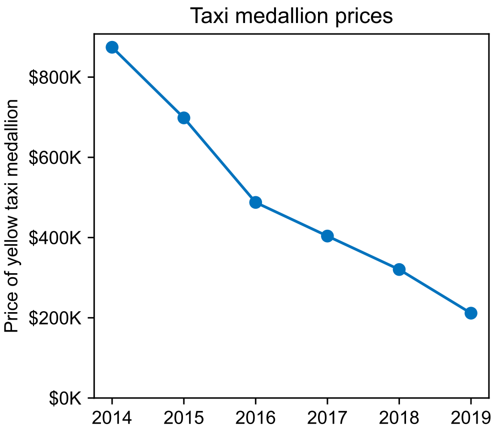
Empirical Example: Taxicab Medallions

- Since 1937, quota on NYC taxicab medallions restricting total supply to $\approx$ 14k.

Empirical Example: Taxicab Medallions

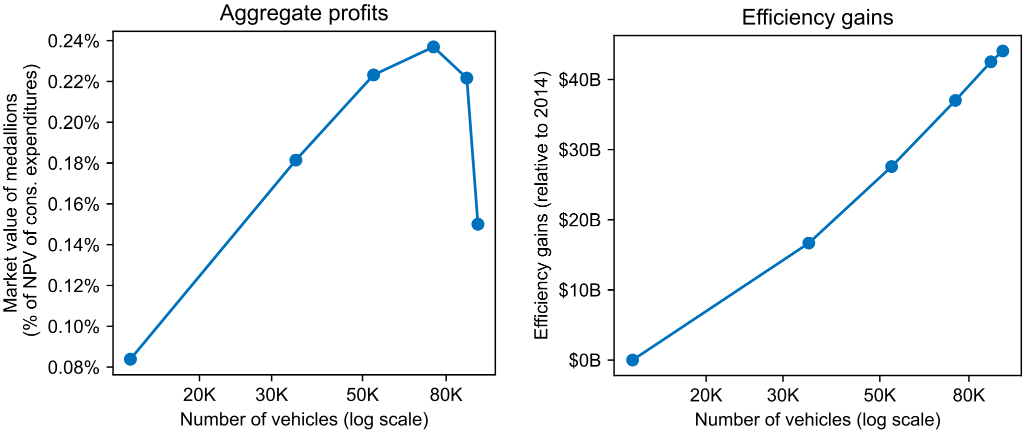
- Since 1937, quota on NYC taxicab medallions restricting total supply to $\approx$ 14k.
- Use arrival of rideshare apps in NYC to quantify gains from relaxing quota on cabs.





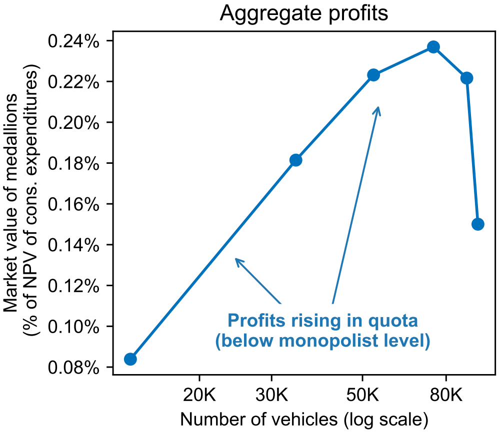
Empirical Example: Taxicab Medallions

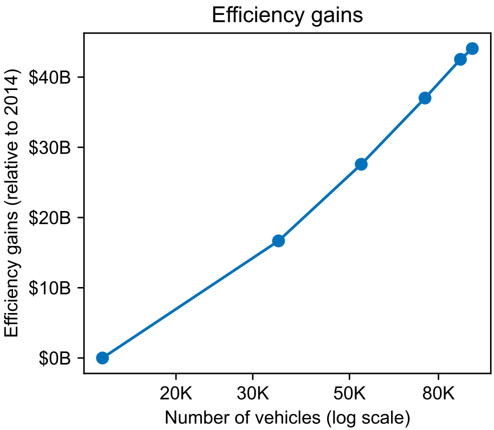
- Assume that medallion transaction prices reflect rents accruing to owners.
- Gains from relaxing taxicab quota are $\Delta \log Y_t \approx \left(\Pi_{it} + \frac{1}{2} d\Pi_{it}\right) d \log y_{i^* t}$.



Empirical Example: Taxicab Medallions

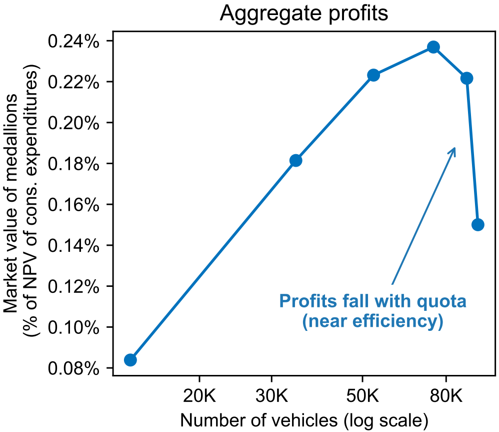
- Assume that medallion transaction prices reflect rents accruing to owners.
- Gains from relaxing taxicab quota are $\Delta \log Y_t \approx \left(\Pi_{it} + \frac{1}{2} d\Pi_{it}\right) d \log y_{i^* t}$.

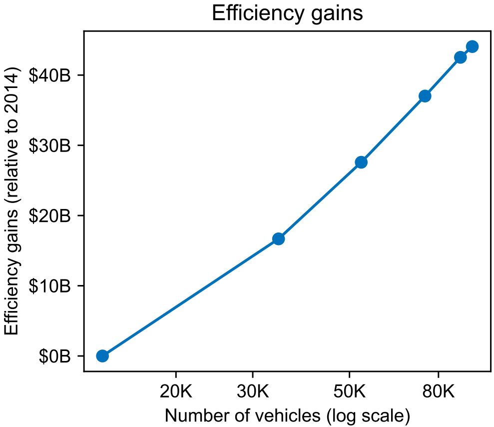




Empirical Example: Taxicab Medallions

- Assume that medallion transaction prices reflect rents accruing to owners.
- Gains from relaxing taxicab quota are $\Delta \log Y_t \approx \left(\Pi_{it} + \frac{1}{2} d\Pi_{it}\right) d \log y_{i^* t}$.





Empirical Example: Taxicab Medallions

- Gains from relaxing quota over 2014–2019.
  - Cumulating gains over each year: $\Delta \log Y \approx \sum_t \left(\Pi_{it} + \frac{1}{2} d\Pi_{it}\right) d \log y_{i^* t}$.
- Not efficient at the end. What is the remaining distance to frontier?
  - Use elasticity of profits to quantity in final year: $\Delta \log Y \approx \frac{1}{2} \Pi_i \left[-\frac{d \log \Pi_i}{d \log y_{i^*}} \right]^{-1}$.

	Change from 2014–2019	Distance to frontier
Output gains	\$44.1B	\$1.8B
Gains per New York MSA household	\$6,029	\$246
% of NPV of transportation expenditures (incl. vehicles/gas)	2.61%	0.11%

Conclusion

- General framework for analyzing economies with quota distortions.
- Two lessons:
  - 1. Any distorted allocation can be recast as equilibrium of an economy with quotas.
  - 2. Economies with quotas are constrained efficient, and thus highly tractable.
- Comparative statics: quota changes, productivity shocks.
- Distance to frontier, nonlinearities even far from the frontier.
- Examples of how to apply results (zoning restrictions, H-1B visas, taxicabs).

Baqaee, D. R. and E. Farhi (2019). The macroeconomic impact of microeconomic shocks: beyond hulten's theorem. *Econometrica 87*(4), 1155–1203.

Baqaee, D. R. and E. Farhi (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics 135*(1), 105–163.

Clemens, M. A. (2013). Why do programmers earn more in houston than hyderabad? evidence from randomized processing of us visas. *American Economic Review 103*(3), 198–202.

Gyourko, J. and J. Krimmel (2021). The impact of local residential land use restrictions on land values across and within single family housing markets. *Journal of Urban Economics 126*, 103374.

Harberger, A. C. (1954). Monopoly and resource allocation. *American Economic Review 44*(2), 77–87.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics 124*(4), 1403–1448.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics 11*(4), 707–720.