Quotas in General Equilibrium

David Baqaee
UCLA

Kunal Sangani
Northwestern

June 2025

Quota Distortions

- Many policies / frictions directly constrain quantities without regard to prices.
  - E.g., import quotas, visa caps, zoning restrictions, emissions limits, local content requirements, land use ceilings, taxicab medallions.
  - Missing markets (land markets, credit markets, insurance markets).
- The classic approach to analyzing distortions is to recast them as implicit taxes.
- But mapping quotas to implicit taxes requires detailed info about economy.
- This paper: A general framework for analyzing economies with quota-like distortions.

Preview of Results

- Much like implicit taxes/wedges, quotas can decentralize any feasible allocation.
- But, economies with quotas are constrained eff. and obey macro-envelope conditions.
  - Comparative statics disciplined by simple sufficient statistics.
  - Not subject to Theory of Second Best.

Preview of Results

- Much like implicit taxes/wedges, quotas can decentralize any feasible allocation.
- But, economies with quotas are constrained eff. and obey macro-envelope conditions.
  - Comparative statics disciplined by simple sufficient statistics.
  - Not subject to Theory of Second Best.
- How small quota changes and productivity shocks affect output.
- How large quota changes affect output (i.e., nonlinearities).
- Distance to the efficient frontier (misallocation cost of quotas).

Environment

- F factors in fixed supply, N goods produced with arbitrary neoclassical technologies.
- Representative consumer with homothetic preferences.
- Exogenous quota y_i^* on good i: $y_i \leq y_i^*$.
- Perfect competition given quotas, general equilibrium.
- Denote real GDP by Y.
- Much like wedges, quotas can decentralize any feasible, inefficient allocation.

Comparative Statics

- Unlike equilibria with wedges, equilibrium with quotas is constrained efficient.
- Comparative statics governed by simple sufficient statistics:

$$\frac{d\log Y}{d\log y_i^*} = \frac{rents_i}{GDP} = \Pi_i, \qquad \frac{d\log Y}{d\log TFP_i} = \frac{sales_i - rents_i}{GDP} = \lambda_i - \Pi_i.$$

where $rents_i$ are excess profits earned by producers that hold quota rights.

Comparative Statics

- Unlike equilibria with wedges, equilibrium with quotas is constrained efficient.
- Comparative statics governed by simple sufficient statistics:

$$\frac{d\log Y}{d\log y_i^*} = \frac{rents_i}{GDP} = \Pi_i, \qquad \frac{d\log Y}{d\log TFP_i} = \frac{sales_i - rents_i}{GDP} = \lambda_i - \Pi_i.$$

  where $rents_i$ are excess profits earned by producers that hold quota rights.
- If equilibrium efficient, quotas non-binding ($\mathbf{\Pi} = 0$) and we recover Hulten (1978):

$$\frac{d\log Y}{d\log y_i^*} = 0, \qquad \frac{d\log Y}{d\log TFP_i} = \frac{sales_i}{GDP} = \lambda_i.$$

- Holding other quotas fixed, removing a quota always raises output.
  - Holding other wedges fixed, removing a wedge can lower output (Theory of 2nd Best).

Empirical Example: Zoning Restrictions on Single-Family Housing

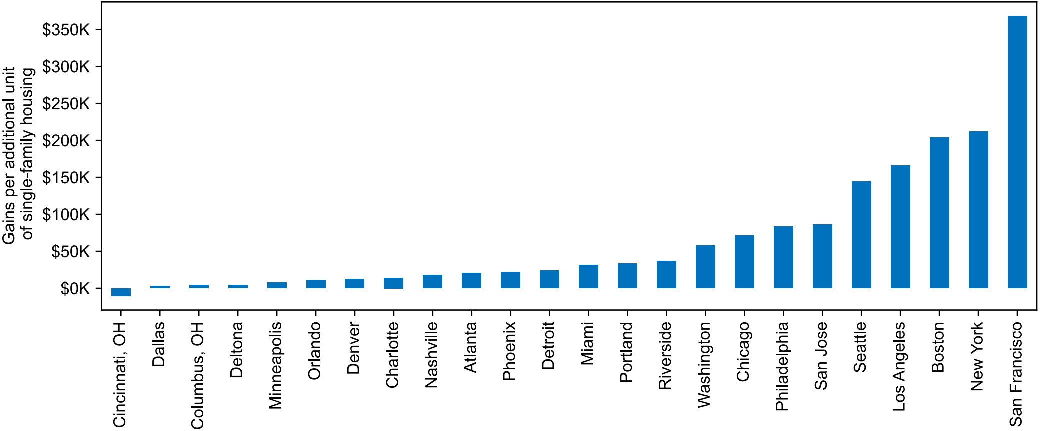
- What are the gains from loosening zoning restrictions on single-family housing?

- To a first order, given by value of rights to build new single-family housing.
  - Gyourko and Krimmel (2021) isolate “zoning taxes” by comparing land value for parcels with rights to build new single-family housing to value of land with existing housing.

- Note: Efficiency gains expressed directly in terms of new units permitted.
  - Wedge approach would require mapping quantities into changes in effective zoning tax.

Empirical Example: Zoning Restrictions on Single-Family Housing

- What are the gains from loosening zoning restrictions on single-family housing?



Nonlinearities

- What about the effects of a large liberalization?
- Since first-order effect depends on rents, nonlinearities depend on **change** in rents:

$$\Delta \log Y \approx \Pi_i \Delta \log y_i^* + \frac{1}{2} \underbrace{\frac{d\Pi_i}{d \log y_i^*}}_{\Delta \text{ rents}} (\Delta \log y_i^*)^2.$$

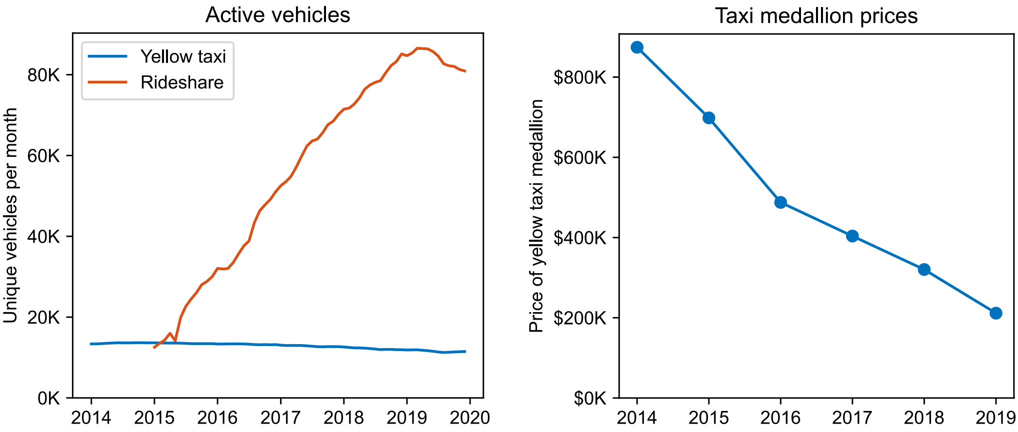
- If rents rise with quota, first-order approx. understates gains from large liberalization.
- Can solve for Δ rents using input-output network & elasticities. (à la Baqaee and Farhi 2019).
- Or obtain Δ rents from ex-post variation: Taxicab medallions in New York.

Empirical Example: Taxicab Medallions

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

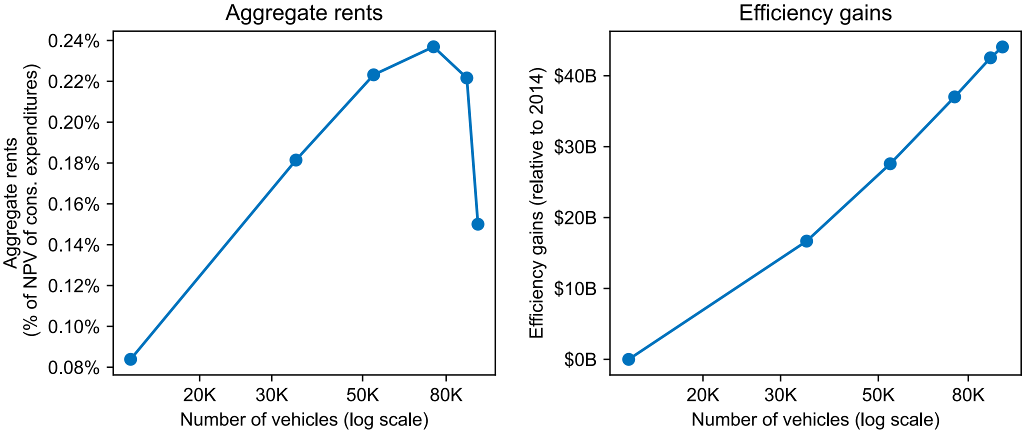
Empirical Example: Taxicab Medallions

- 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.



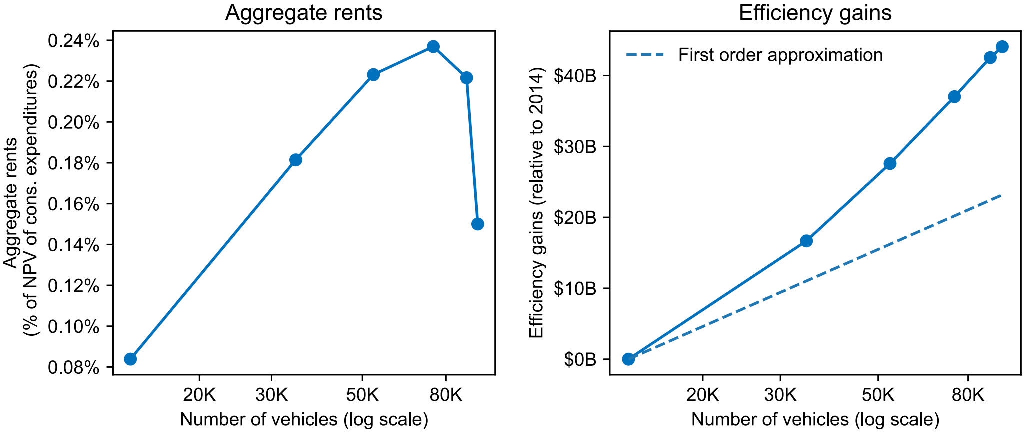
Nonlinearities: Taxicab Medallions

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



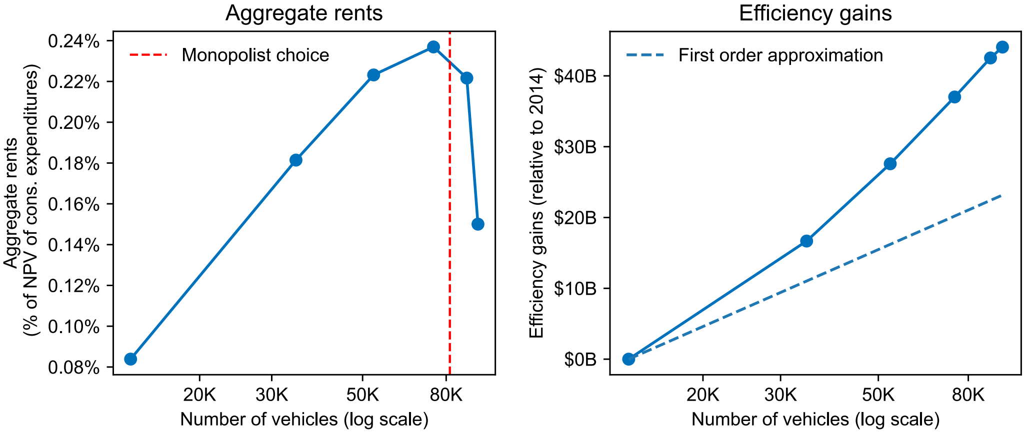
Nonlinearities: Taxicab Medallions

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



Nonlinearities: Taxicab Medallions

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



Empirical Example: Taxicab Medallions

- Gains from relaxing quota over 2014–2019.
  - Cumulating gains over each year: $\Delta \log Y \approx \sum_t \left(\Pi_{it} + \frac{1}{2} d\Pi_{it}\right) \Delta \log y^*_{it}$.

	Change from 2014–2019
Output gains	\$44.1B
Gains per New York MSA household	\$6,029
% of NPV of transportation expenditures	2.61%

Distance to the Efficient Frontier

- What are gains from eliminating a quota altogether?
- To a second-order, gains are average of first-order effect at distorted pt and efficient pt:

$$\Delta \log Y \approx \frac{1}{2}\Pi_i(\Delta \log y_i^*) + \frac{1}{2}(0),$$

where $\Delta \log y_i^* = \log y_i^* - \log y_i^{\text{eff}}$ is gap between quota and undistorted level.

Distance to the Efficient Frontier

- What are gains from eliminating a quota altogether?
- To a second-order, gains are average of first-order effect at distorted pt and efficient pt:

$$\Delta \log Y \approx \frac{1}{2}\Pi_i(\Delta \log y_i^*) + \frac{1}{2}(0),$$

where $\Delta \log y_i^* = \log y_i^* - \log y_i^{\text{eff}}$ is gap between quota and undistorted level.

Distance to the Efficient Frontier

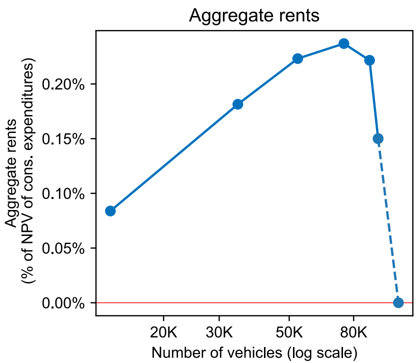
- What are gains from eliminating a quota altogether?
- To a second-order, gains are average of first-order effect at distorted pt and efficient pt:

$$\Delta \log Y \approx \frac{1}{2}\Pi_i(\Delta \log y_i^*) + \frac{1}{2}(0),$$

where $\Delta \log y_i^* = \log y_i^* - \log y_i^{\text{eff}}$ is gap between quota and undistorted level.

- Estimate increase in quota necessary to decrease rents to zero.

Distance to the Efficient Frontier



- Estimate increase in quota necessary to decrease rents to zero.

Distance to the Efficient Frontier

- What are gains from eliminating a quota altogether?
- To a second-order, gains are average of first-order effect at distorted pt and efficient pt:

$$\Delta \log Y \approx \frac{1}{2}\Pi_i(\Delta \log y_i^*) + \frac{1}{2}(0),$$

where $\Delta \log y_i^* = \log y_i^* - \log y_i^{\text{eff}}$ is gap between quota and undistorted level.

- Estimate increase in quota necessary to decrease rents to zero.

Distance to the Efficient Frontier

- What are gains from eliminating a quota altogether?
- To a second-order, gains are average of first-order effect at distorted pt and efficient pt:

$$\Delta \log Y \approx \frac{1}{2}\Pi_i \underbrace{\left[-\frac{d\log \Pi_i}{d\log y_i^*}\right]^{-1}}_{\substack{\text{Inverse elasticity} \\ \text{of rents to quota}}}$$

  where $\Delta \log y_i^* = \log y_i^* - \log y_i^{\text{eff}}$ is gap between quota and undistorted level.

- Estimate increase in quota necessary to decrease rents to zero.

Distance to the Efficient Frontier

- What are gains from eliminating a quota altogether?
- To a second-order, gains are average of first-order effect at distorted pt and efficient pt:

$$\Delta \log Y \approx \frac{1}{2}\Pi_i \underbrace{\left[-\frac{d\log \Pi_i}{d\log y_i^*}\right]^{-1}}_{\substack{\text{Inverse elasticity} \\ \text{of rents to quota}}}$$

  where $\Delta \log y_i^* = \log y_i^* - \log y_i^{\text{eff}}$ is gap between quota and undistorted level.

- Estimate increase in quota necessary to decrease rents to zero.
- If rents fall quickly when quota relaxed, close to efficiency $\Rightarrow$ smaller gains.

Empirical Example: Taxicab Medallions

- Gains from relaxing quota over 2014–2019.
  - Cumulating gains over each year: $\Delta \log Y \approx \sum_t \left(\Pi_{it} + \frac{1}{2} d\Pi_{it}\right) \Delta \log y^*_{it}$.
- Not efficient at the end. What is the remaining distance to frontier?

	Change from 2014–2019
Output gains	\$44.1B
Gains per New York MSA household	\$6,029
% of NPV of transportation expenditures	2.61%

Empirical Example: Taxicab Medallions

- Gains from relaxing quota over 2014–2019.
  - Cumulating gains over each year: $\Delta \log Y \approx \sum_t \left(\Pi_{it} + \frac{1}{2} d\Pi_{it}\right) \Delta \log y_{it}^*$.
- Not efficient at the end. What is the remaining distance to frontier?
  - Use elasticity of rents to quota 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	2.61%	0.11%

Nonlinearities: Multiple Quotas

- Method scales up to multiple interacting quotas

$$\Delta \log Y \approx \mathbf{\Pi}' d \log \mathbf{y}^* + \frac{1}{2} (d \log \mathbf{y}^*)' \frac{d\mathbf{\Pi}}{d \log \mathbf{y}^*} (d \log \mathbf{y}^*),$$

- Quota demand system $\frac{d\mathbf{\Pi}}{d \log \mathbf{y}^*}$ summarizes responses of rents to quotas.

Nonlinearities: Multiple Quotas

- Method scales up to multiple interacting quotas

$$\Delta \log Y \approx \mathbf{\Pi}' d \log \mathbf{y}^* + \frac{1}{2} (d \log \mathbf{y}^*)' \frac{d\mathbf{\Pi}}{d \log \mathbf{y}^*} (d \log \mathbf{y}^*),$$

- Quota demand system $\frac{d\mathbf{\Pi}}{d \log \mathbf{y}^*}$ summarizes responses of rents to quotas.

- Similarly, gains from eliminating quotas simultaneously given by:

$$\Delta \log Y \approx -\frac{1}{2} \mathbf{\Pi}' \left[\frac{d\mathbf{\Pi}}{d \log \boldsymbol{y}^*} \right]^{-1} \mathbf{\Pi}.$$

- If *i*'s rents fall when *j*'s quota relaxed, then

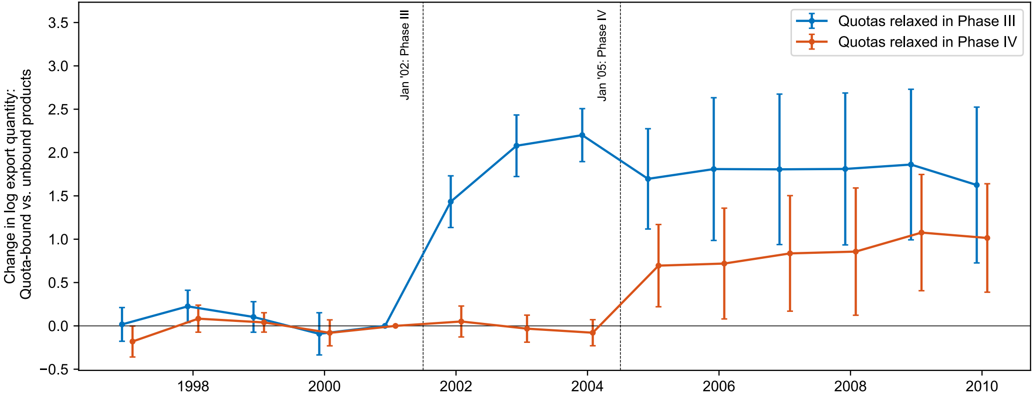
gains from relaxing both quotas $<$ sum of gains from relaxing each.

Empirical Example: China’s Textile & Clothing Exports

- 1975–1994 Multi-Fiber Agreement capped China’s textile & clothing exports to US, EU.
- Staged phase-out:
  - Jan 2002 (Phase III): Knit fabrics, gloves, dressing gowns, brassieres, etc.
  - Jan 2005 (Phase IV): Silk, wool, and cotton textiles, other apparel categories, etc.
- Obtain quota demand system using initial rents & response of exports to liberalization.
- Use quota auction prices for initial rents: $\Pi_{\text{Phase III}} = \$38\text{B}$, $\Pi_{\text{Phase IV}} = \$394\text{B}$.

Empirical Example: China's Textile & Clothing Exports

- Reaction of export quantities as quotas are removed.
- As second group liberalized, quantity of first group falls. (Nonlinear interaction.)



Empirical Example: China's Textile & Clothing Exports

- Estimated quota demand system:

$$\Pi = \begin{bmatrix} \Pi_{\text{Phase III}} \\ \Pi_{\text{Phase IV}} \end{bmatrix} = \begin{bmatrix} \$38B \\ \$394B \end{bmatrix}, \qquad \frac{d \log \boldsymbol{\Pi}}{d \log \mathbf{y}^*} = \begin{bmatrix} -0.472 & -0.200 \\ -0.019 & -1.258 \end{bmatrix}.$$

Intervention	Efficiency gains (2001 USD $B)
(A) Relaxing Phase III quotas only	$40

Empirical Example: China's Textile & Clothing Exports

- Estimated quota demand system:

$$\Pi = \begin{bmatrix} \Pi_{\text{Phase III}} \\ \Pi_{\text{Phase IV}} \end{bmatrix} = \begin{bmatrix} \$38\text{B} \\ \$394\text{B} \end{bmatrix}, \qquad \frac{d \log \boldsymbol{\Pi}}{d \log \mathbf{y}^*} = \begin{bmatrix} -0.472 & -0.200 \\ -0.019 & -1.258 \end{bmatrix}.$$

Intervention	Efficiency gains (2001 USD \$B)
(A) Relaxing Phase III quotas only	\$40
(B) Relaxing Phase IV quotas only	\$158

Empirical Example: China's Textile & Clothing Exports

- Estimated quota demand system:

$$\Pi = \begin{bmatrix} \Pi_{\text{Phase III}} \\ \Pi_{\text{Phase IV}} \end{bmatrix} = \begin{bmatrix} \$38\text{B} \\ \$394\text{B} \end{bmatrix}, \qquad \frac{d \log \mathbf{\Pi}}{d \log \mathbf{y}^*} = \begin{bmatrix} -0.472 & -0.200 \\ -0.019 & -1.258 \end{bmatrix}.$$

Intervention	Efficiency gains (2001 USD \$B)
(A) Relaxing Phase III quotas only	\$40
(B) Relaxing Phase IV quotas only	\$158
(C) Relaxing both Phase III and IV quotas	\$185

Empirical Example: China's Textile & Clothing Exports

- Estimated quota demand system:

$$\Pi = \begin{bmatrix} \Pi_{\text{Phase III}} \\ \Pi_{\text{Phase IV}} \end{bmatrix} = \begin{bmatrix} \$38\text{B} \\ \$394\text{B} \end{bmatrix}, \qquad \frac{d \log \boldsymbol{\Pi}}{d \log \mathbf{y}^*} = \begin{bmatrix} -0.472 & -0.200 \\ -0.019 & -1.258 \end{bmatrix}.$$

Intervention	Efficiency gains (2001 USD $B)
(A) Relaxing Phase III quotas only	$40
(B) Relaxing Phase IV quotas only	$158
(C) Relaxing both Phase III and IV quotas	$185
Difference: $\text{C} - (\text{A} + \text{B})$	$13

- Gains from relaxing both quotas $<$ sum of estimated gains from relaxing each.

Conclusion

- General framework for analyzing economies with quota distortions.
- Comparative statics simple because of constrained efficiency.
- Nonlinearities, distance to efficient frontier using quota demand system.
- Can be identified with local variation, e.g., response of rents to quota changes.
- Other applications in paper: H-1B visa cap, Argentina's capital controls.