A Revaled-Preference Approach to Measuring Information Frictions in Migration Decisions

Thomas Fujiwara
Eduardo Morales
Charly Porcher

Princeton University

October 11, 2019

Importance of Migration

- Labor demand shocks differ widely across regions within countries.
- Possible sources of region-specific labor demand shocks:
  - Sector-demand shocks: Bartik (1991), Blanchard and Katz (1992);
  - Establishment of new plants: Greenstone, Hornbeck, and Moretti (2010);
  - International competition: Topalova (2010), Autor, Dorn, and Hanson (2013).
- Standard spatial equilibrium models predict that, by migrating or commuting, workers can avoid negative labor demand shocks in their region of residence and benefit from positive shocks in other regions.
  - Monte, Redding, Rossi-Hansberg (2018), Bryan and Morten (2018).
- Migration patterns often do not respond to regional labor demand shocks.
  - Autor, Dorn, Hanson (2013), Kovak (2013), Dix-Carneiro and Kovak (2017).

In this Paper

- Are workers' limited migration responses to local labor demand shocks due to lack of information about the potential net gains from regional migration?
- To answer this question, we use *RAIS* data:
  - Sample period: 2001 to 2014.
  - Sample: all formally employed Brazilian workers.
- Two approaches:
  - Reduced-form: heterogeneous delay in reaction to local labor demand shocks.
  - Structural: model-based moment conditions plus tests of overidentifying restrictions as tool to rule out content of potential migrants' information sets.

Reduced-form Approach

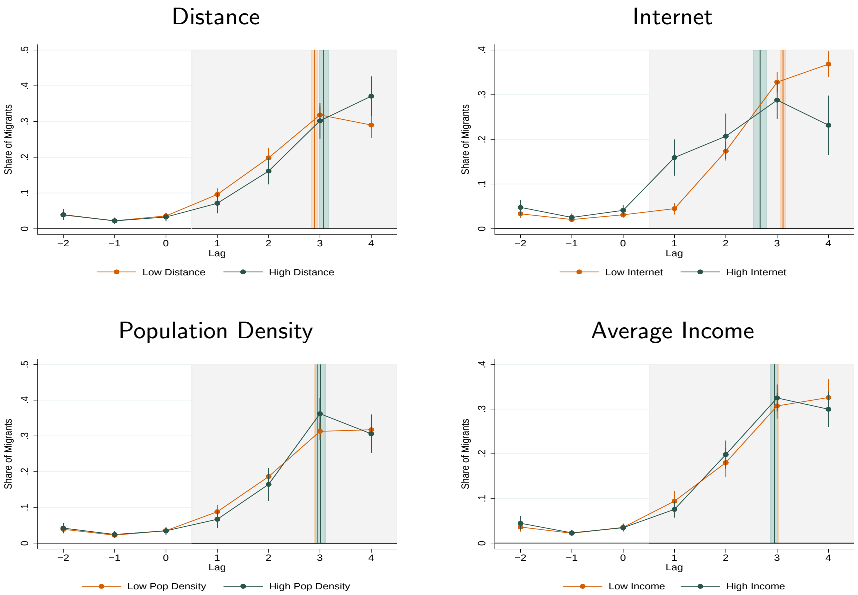
- Identify a microregion l as having experienced a positive labor demand shock at year t if three criteria are met:
  1. Average yearly immigration rate between t and $t+4$ is at least 50% larger than average yearly immigration rate between $t-2$ and t.
  2. Average yearly population between $t-2$ and t is at least 20,000.
  3. Total number of immigrants between $t-2$ and $t+4$ is 20,000.

Reduced-form Approach

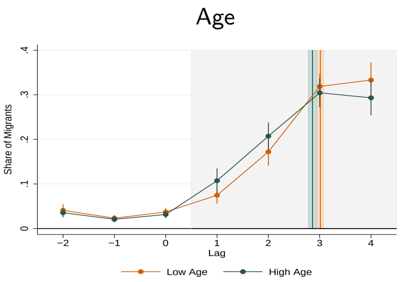
- List of microregions:

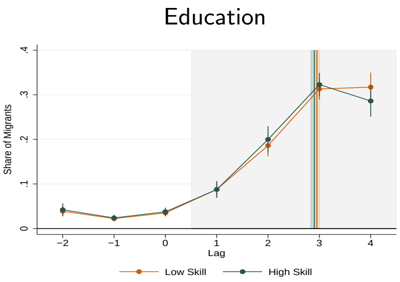
Municipality	Year Shock	Immigrants	Source Labor Shock
Ipojuca	2009	89,067	Refinery Construction
Natal	2008	74,294	Oil Boom
Santo Antonio	2008	65,746	Dam Construction
Maceio	2008	49,943	Tourism Boom
Belo Monte	2010	40,644	Dam Construction
Uberaba	2008	40,301	Sugar Cane Energy
Araucaria	2006	37,684	Tourism Boom
Cidelandia	2010	37,216	Palm Oil Boom
Suape	2008	32,752	Refinery Construction
Itabira	2010	21,115	Mining Boom

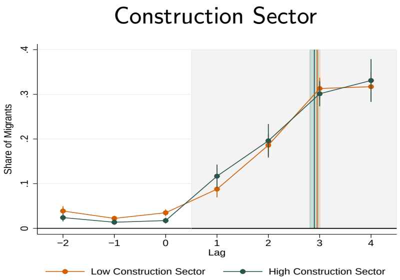
Reduced-form Approach: Results for Belo Monte

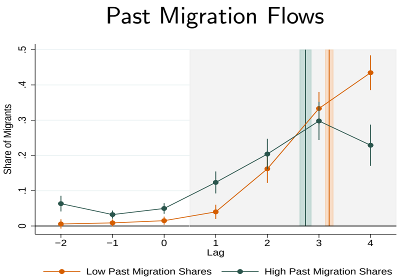


Reduced-form Approach: Results for Belo Monte



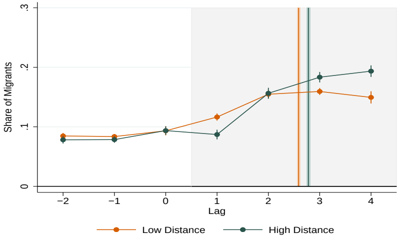




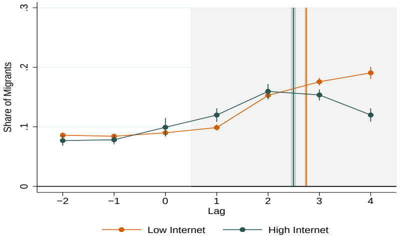


Reduced-form Approach: Results from Pooled Regression

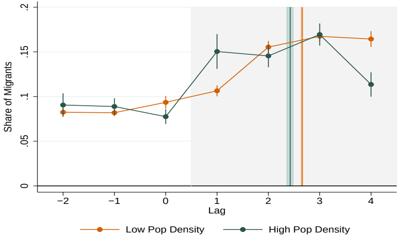
Distance



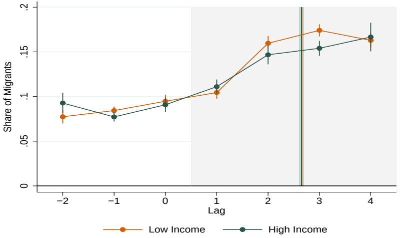
Internet



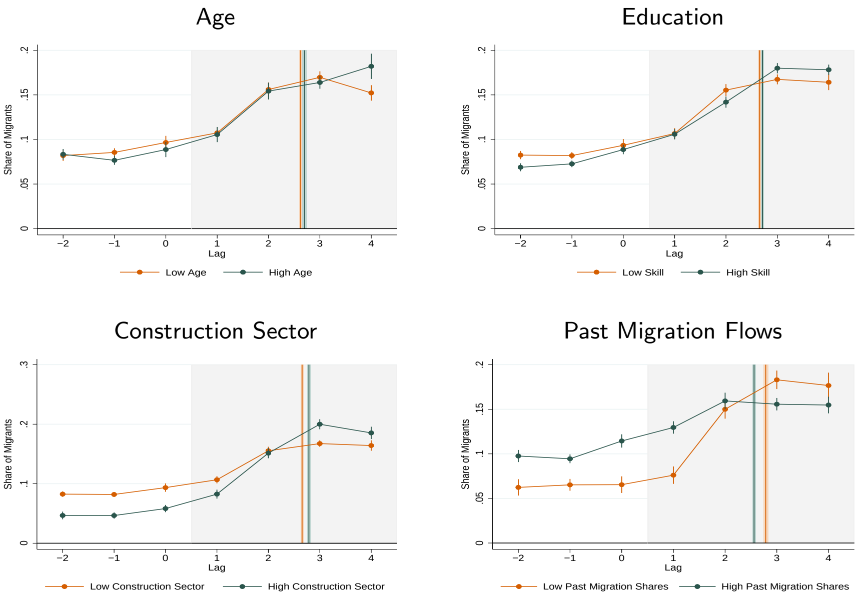
Population Density



Average Income



Reduced-form Approach: Results from Pooled Regression



Structural Approach: Model

- In any year t, every worker i decides the market $m(i,t)$ in which to supply labor. Each **market** m is the combination of a **sector** n and a **region** l.
- Workers are partitioned into **demographic types**. Workers belong to the same demographic type if they share education level, gender, and age. We denote the demographic type of worker i as $d(i)$.
- Workers are also partitioned into **groups**. Workers belong to the same group if they belong to the same demographic type and were employed in the same market in the previous period:

$$g(i,t) = g(i',t') \quad \text{iff} \quad d(i) = d(i') \text{ and } m(i,t-1) = m(i',t'-1).$$

- Every worker i decides in any year t the market $m(i,t)$ in which to supply labor as the outcome of a dynamic optimization problem.

Structural Approach: Static Utility

- Static utility depends on a group-year term and a idiosyncratic shock:

$$U_m(i,t) = u_m(g(i,t),t) + \nu_m(i,t).$$

- Group-year term depends on wages, and on regional and sectoral terms:

$$u_m(g(i,t),t) = \alpha w_m(d(i),t) + u_{\check{l}(m)}(g(i,t),t) + u_{\check{n}(m)}(g(i,t),t).$$

- Regional term depends on regional amenities, migration costs, and prices:

$$u_{\check{l}(m)}(g(i,t),t) = a_{\check{l}(m)}(d(i),t) - \Upsilon_{\check{l}(m)}(d(i),l(i,t-1),t) - \alpha p_{\check{l}(m)}(d(i),t).$$

- Sectoral term depends on sectoral amenities and sectoral reallocation costs:

$$u_{\check{n}(m)}(g(i,t),t) = b_{\check{n}(m)}(d(i),t) - \Gamma_{\check{n}(m)}(d(i),n(i,t-1),t).$$

- Idiosyncratic shock is independent across i and t, Type I Extreme Value distributed, and independent of every other variable entering the model.

Structural Approach: Information Set

- For every worker i and period t,

$$\mathcal{V}(i,t) = (\mathcal{W}(i,t), \nu(i,t)),$$

with $\nu(i,t)$ defined as

$$\nu(i,t) \equiv (\nu_m(i,t); m = 1, \ldots, M)$$

and $\mathcal{W}(i,t)$ satisfying a limited-heterogeneity constraint

$$\mathcal{W}(i,t) = \mathcal{W}(i',t') \quad \text{if and only if} \quad g(i,t) = g(i',t') \text{ and } t = t',$$

and a minimum-content constraint

$$\mathcal{Z}(g(i,t),t) \subseteq \mathcal{W}(g(i,t),t),$$

where $\mathcal{Z}(g(i,t),t)$ is a vector observed by the researcher.

Structural Approach: Optimal Choice

- Optimal choice of worker i at period t is:

$$m(i,t) = \underset{m \in M}{\operatorname{argmax}} \, \mathbb{E}[V_m(i,t) | \mathcal{V}(i,t)]$$

with the m-specific discounted sum of future payoffs defined as

$$V_m(i,t) \equiv v_m(g(i,t),t) + \nu_m(i,t),$$

with

$$v_m(g(i,t),t) \equiv u_m(g(i,t),t) + \delta \mathsf{V}(\mathfrak{g}(g,m),t+1),$$

δ is the discount factor, $\mathfrak{g}(g,m)$ denotes the group to which worker i belongs at $t+1$ if i belonged to group g at t and chose market m at t, and

$$\mathsf{V}(\mathfrak{g}(g,m),t+1) \equiv \max_{m \in M} \mathbb{E}[V_m(i,t+1) | \mathcal{W}(\mathfrak{g}(g,m),t+1)].$$

Structural Approach: Choice Probabilities

- The probability that a worker i of type g chooses market m at period t is

$$P(m = m(i,t)|g(i,t) = g, t) = \frac{\exp\left(\mathbb{E}[v_m(g,t)|\mathcal{W}(g,t)]\right)}{\sum_{m' \in M} \exp\left(\mathbb{E}[v_{m'}(g,t)|\mathcal{W}(g,t)]\right)}.$$

- By transforming these probabilities adequately, we obtain

$$\ln\left(\left(\frac{P(m = m(i,t)|g(i,t) = g, t)}{P(m' = m(i,t)|g(i,t) = g, t)}\right) \times \left(\frac{P(m' = m(i,t+1)|g(i,t+1) = \mathfrak{g}(g,m), t+1)}{P(m' = m(i,t+1)|g(i,t+1) = \mathfrak{g}(g,m), t+1)}\right)^{\delta}\right)$$

$$=$$

$$\chi(d(i), \check{l}(m), l(i,t-1), t) + \alpha \mathbb{E}\left[w_m(d(i),t) - w_{m'}(d(i),t) \middle| \mathcal{W}(g(i,t),t)\right],$$

where $m = (n(i,t-1), l)$ for some region l, $m' = (n(i,t-1), l(i,t-1))$, and $\chi(\cdot)$ is an unobserved effect that varies by demographic type, regions l and $l(i,t-1)$, and period t.

Structural Approach: Moment Conditions

- Given that agents' expectations are rational and given that

$$\mathcal{Z}(g(i,t),t) \subseteq \mathcal{W}(g(i,t),t),$$

the following moment conditions hold

$$\mathbb{E}\Bigg[\ln\Bigg(\left(\frac{P(m=m(i,t)|g(i,t)=g,t)}{P(m'=m(i,t)|g(i,t)=g,t)}\right)\times$$

$$\left(\frac{P(m'=m(i,t+1)|g(i,t+1)=\mathfrak{g}(g,m),t+1)}{P(m'=m(i,t+1)|g(i,t+1)=\mathfrak{g}(g,m),t+1)}\right)^{\delta}\Bigg)$$

$$-$$

$$\Big(\chi(d(i),\check{l}(m),l(i,t-1),t)+\alpha\big(w_m(d(i),t)-w_{m'}(d(i),t)\big)\Big)\Big|\mathcal{Z}(g(i,t),t)\Bigg]=0,$$

with $m=(n(i,t-1),l)$ for some region l, and $m'=(n(i,t-1),l(i,t-1))$.

Structural Approach: Properties of Moment Conditions

- If we build a moment condition using a vector $\mathcal{Z}(g(i,t),t)$ such that

$$\mathcal{Z}(g(i,t),t) \not\subseteq \mathcal{W}(g(i,t),t),$$

then the moment conditions are invalid and the corresponding Generalized Method of Moments (GMM) estimator of α will be biased towards zero:

$$plim(\hat{\alpha}) < \alpha.$$

- Assume that $\mathcal{Z}^*(g(i,t),t) \subseteq \mathcal{W}(g(i,t),t)$. If moment conditions built using $\mathcal{Z}^*(g(i,t),t)$ as instruments are sufficient to identify α, then we can use a J-test of overidentifying restrictions to test the null hypothesis that some alternative vector of instruments $\mathcal{Z}'(g(i,t),t)$ verifies

$$\mathcal{Z}'(g(i,t),t) \subseteq \mathcal{W}(g(i,t),t).$$

Structural Approach: Possible Instruments

$$\mathcal{Z}_1'(g(i,t),t) = \left\{ \begin{array}{c} \textbf{exact } t-1 \text{ difference in average wages} \\ \text{between any two } \textbf{microregions} \text{ in sector } n(i,t-1) \end{array} \right\}$$

$$\mathcal{Z}_2'(g(i,t),t) = \left\{ \begin{array}{c} \textbf{sign} \text{ of } t-1 \text{ difference in average wages} \\ \text{between any two } \textbf{microregions} \text{ in sector } n(i,t-1) \end{array} \right\}$$

$$\mathcal{Z}_3'(g(i,t),t) = \left\{ \begin{array}{c} \textbf{exact } t-1 \text{ difference in average wages} \\ \text{between any two } \textbf{mesoregions} \text{ in sector } n(i,t-1) \end{array} \right\}$$

$$\mathcal{Z}^*(g(i,t),t) = \left\{ \begin{array}{c} \textbf{sign} \text{ of } t-1 \text{ difference in average wages} \\ \text{between any two } \textbf{mesoregions} \text{ in sector } n(i,t-1) \end{array} \right\}$$

- Note that

$$\begin{aligned} \mathcal{Z}^*(g(i,t),t) \subset \mathcal{Z}_3'(g(i,t),t) &\subset \mathcal{Z}_1'(g(i,t),t), \\ \mathcal{Z}_2'(g(i,t),t) &\subset \mathcal{Z}_1'(g(i,t),t). \end{aligned}$$

Structural Approach: Estimates of Migration Elasticity

	OLS	TSLS			
Instrumental Variable:		$\mathcal{Z}_1'$	$\mathcal{Z}_2'$	$\mathcal{Z}_3'$	$\mathcal{Z}^*$
Forward-looking agents - Large set of fixed effects					
α	0.130^a	0.168^a	0.177^a	0.318^a	0.303^a
	(0.0116)	(0.0185)	(0.0183)	(0.0329)	(0.0327)
Observations			536,876		
First-stage IV estimate		0.690^a	0.159^a	0.542^a	0.0981^a
		(0.00719)	(0.00165)	(0.00972)	(0.00160)

Structural Approach: Testing Content of Information Set

Regions tested	Variable tested	J-test statistic	J-test p-value
All	$\mathcal{Z}_1'$	32.88	0.00
	$\mathcal{Z}_2'$	8.14	0.01
	$\mathcal{Z}_3'$	3.85	0.04
Geographically far away	$\mathcal{Z}_1'$	25.22	0.00
	$\mathcal{Z}_2'$	8.24	0.01
	$\mathcal{Z}_3'$	5.25	0.03
Geographically close	$\mathcal{Z}_1'$	9.16	0.01
	$\mathcal{Z}_2'$	3.38	0.06
	$\mathcal{Z}_3'$	1.87	0.17
Low internet penetration	$\mathcal{Z}_1'$	13.12	0.00
	$\mathcal{Z}_2'$	5.05	0.03
	$\mathcal{Z}_3'$	3.13	0.07
High internet penetration	$\mathcal{Z}_1'$	6.94	0.02
	$\mathcal{Z}_2'$	2.98	0.11
	$\mathcal{Z}_3'$	1.16	0.28

Conclusions

- We provide an approach to identify the content of agents' information sets.
- Our approach requires specifying how workers' migration decisions depend on their expectations of the current and future real wages and amenities in any location they may consider migrating to.
- We show that estimated migration elasticities are sensitive to the assumptions that the researcher imposes on potential migrants' information sets.
- We also show evidence that:
  1. workers living in regions with higher internet penetration rates have more information about other labor markets, no matter whether these are located;
  2. workers have more information about geographically close labor markets than about labor markets that are geographically farther away.
- Future steps:
  - Measure economic impact of information frictions for workers.
  - Relax assumption that information sets are constant within a group $\rightarrow$ moment inequality approach.

Thank you!