Formal Entry, Informal Rents, and Labor Rationing

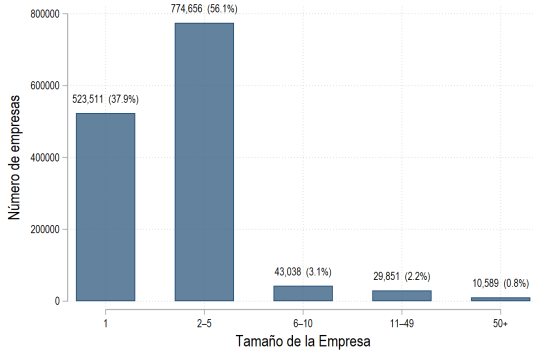
Paúl J. Corcuera
Universidad de Piura

Martín Vargas
Virginia Tech

December 19th, 2025

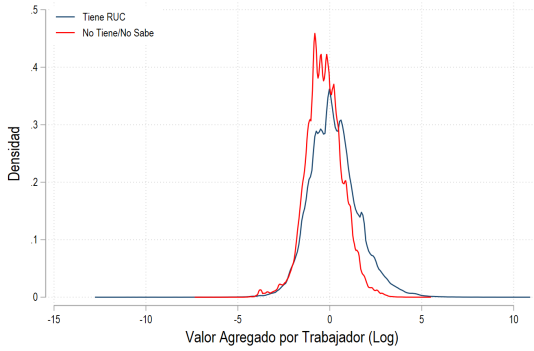
Ubiquity of small firms in developing countries

- Small firms dominate the economic landscape in LMIC.
- Even within urban areas, up to 97% of firms areas have ≤ 10 workers.
- This is true for both the **formal** sector (tax registered) and the **informal** sector.



Do formal and informal firms compete against each other?

- Formal and informal firms are not in separate worlds.
- Consequential overlap in productivity distributions (Ulyssea, 2018).
- If they produce similar goods with similar efficiency, they are likely substitutes.
  - ✓ Formal entry should directly impact informal incumbents.



Two Views on Informality

If we facilitate formal entry (firm entry shock), what happens to the informal sector?

1. **Labor Rationing (Survival) View:**
   - Informality is disguised unemployment due to scarce formal jobs.
   - Formal firm entry creates jobs, absorbs excess labor supply and wages rise.
2. **Rent Erosion (Parasite) View:**
   - Informal firms survive by evading taxes and exercising local market power.
   - Formal firm entry increases competition and erodes informal profits, inducing exit.

▸ Evidence from Brazil

Our Paper

- We study the staggered rollout of Centros de Desarrollo Empresarial (CDE) that induced formal firm entry in Peru.
- **Theory:** Model of informal firms facing formal entry to highlight two channels.
  - **Intensive:** shrink by substituting paid for unpaid labor.
  - **Extensive:** decision to exit due to profit compression.
- **Empirics:** Does the shock displace informal incumbents (rent erosion) or absorb them into better employment (labor rationing)?
  - Survey information of informal workers and their firms' organization.
  - Administrative records of formal firms born from CDEs with their location and 4-digit industry.
  - We evaluate this through a difference-in-differences and an exposure (IV) approach.

(Expected) Contributions

- **Barriers to Entry and Formality** [Bruhn '11; Schiffbauer et al. '25; Kaplan et al. '11; De Mel et al. '13; Mullainathan & Schnabl '10]
  - ✓ We shift focus towards the incumbent informal firms. Even though few informal businesses do register, there might be relevant consequences!
  - ✓ We can track both displacement (exit) and reorganization of family labor.
- **Competition and Spillovers** [Busso et al. '12; Atkin et al. '18; Goldberg & Pavcnik '03]
  - ✓ We study spillover effects from peer-to-peer competition, rather than big vs. small firm dynamics.
  - ✓ Disentangles the falling barriers from "big and productive" firms crushing the small.
- **Structural Views of Informality** [La Porta & Shleifer '14; Maloney '04; Ulyssea '18/'20; Meghir et al. '15; Amodio et al. '25; Breza et al. '21]
  - ✓ Structural models typically assume rationing or sorting.
  - ✓ We provide quasi-experimental evidence of labor rationing.

Roadmap

- Model of informal incumbent firms facing firm entry shocks.
- Data and Empirical Strategy.
- Preliminary results.
- Further steps.

Theoretical Model

Model Setup

- **Consumers**
  - Representative household with CES ($\sigma > 1$) preferences over continuum of varieties $\omega \in \Omega$:

$$U = \left(\int_{w \in \Omega} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}. \tag{1}$$

  - Demand for variety ω given by $q(\omega) = \frac{M}{P} \left(\frac{p(\omega)}{P} \right)^{-\sigma}$, where $P = \left(\int p(\omega)^{1-\sigma} d\omega \right)^{\frac{1}{1-\sigma}}$.

- **Production**
  - Informal firms use imperfectly substitutable inputs (family vs. paid) given elasticity η:

$$y = A \left[\alpha L_f^{\frac{\eta-1}{\eta}} + (1-\alpha) L_p^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}, \quad A \sim G(\cdot). \tag{2}$$

  - They face a market wage (w) and a shadow wage/opportunity cost (λ) to hire each type of worker.

The Intensive Margin: Labor Substitution

- The Cost Minimization Problem is given by

$$\min_{L_p, L_f} \{wL_p + \lambda L_f\} \quad \text{s.t.}$$
$$y = A\left[\alpha L_f^{\frac{\eta-1}{\eta}} + (1-\alpha)L_p^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}$$

- This yields the optimal labor mix condition:

$$\frac{L_{\text{paid}}}{L_{\text{family}}} = \left(\frac{1-\alpha}{\alpha}\right)^{\eta}\left(\frac{\lambda}{w}\right)^{\eta} \tag{3}$$

- Naturally, it implies that if formal entry raises market wages relative to shadow wages ($w/\lambda \uparrow$) , informal firms **substitute away** from paid employees ($L_p/L_f \downarrow$).

The Extensive Margin: Exits

- **Profit Function:** Under monopolistic competition, operating profits are:

$$\pi(A) = \underbrace{\frac{M}{\sigma}\left(\frac{P(\sigma-1)}{\sigma \mathcal{C}(w,\lambda)}\right)^{\sigma-1}}_{\text{Market Competitiveness } (\mathcal{B})} \cdot A^{\sigma-1}, \tag{4}$$

  where $\mathcal{C}(w, \lambda)$ is the unit cost index of the labor bundle.

- **Exit Decision:** Owner exits if $\pi(A) < \bar{u}$ (Outside Option). This defines the cutoff A^*_{exit} given by:

$$A^*_{exit} = \left(\frac{\bar{u}}{\mathcal{B}(P, w, \lambda)}\right)^{\frac{1}{\sigma-1}} \tag{5}$$

The Impact of Competition: Rent Erosion

Proposition (Rent Erosion)

The productivity exit threshold A^*_{exit} is decreasing in the aggregate price index P:

$$\frac{\partial A^*_{exit}}{\partial P} = \bar{u}^{\frac{1}{\sigma-1}} \underbrace{(1-\sigma)}_{(-)} \underbrace{\frac{\partial \mathcal{B}(P,w,\lambda)}{\partial P}}_{(+)} < 0$$

Consequently, a formal entry shock that lowers P (via increased variety) leads to a rise in the exit threshold, forcing low-productivity firms out.

- Stronger effects when owners face better outside options ($\bar{u} \uparrow$).

The Impact of Competition: Labor Rationing

Proposition (Labor Rationing)

The productivity exit threshold A^*_{exit} is increasing in the market wage w:

$$\frac{\partial A^*_{exit}}{\partial w} = \bar{u}^{\frac{1}{\sigma-1}} \underbrace{(1-\sigma)}_{(-)} \underbrace{\frac{\partial \mathcal{B}(P,w,\lambda)}{\partial w}}_{(-)} > 0$$

Consequently, a formal entry shock that increases w due to more labor demand leads to a rise in the exit threshold, forcing low-productivity firms out.

- Stronger effects when owners face better outside options ($\bar{u} \uparrow$).
- If this mechanism dominates firms are facing a “double squeeze” from both revenue and costs.

Taking Stock: Mapping Theory to Empirics

We layout the empirical implications:

Outcome	**Rent Erosion Only**	**Rent Erosion + Labor Rationing**
Exit Rate (A^*_{exit})	$\uparrow$	$\uparrow\uparrow$
Informal Wages (w)	$\approx$ or $\downarrow$	$\uparrow$
Labor Mix (L_p/L_f)	$\approx$	$\downarrow$

- We expect (unambiguously) that the exit rate increases.
- The response on informal wages and labor mix are our key tests between the mechanisms.

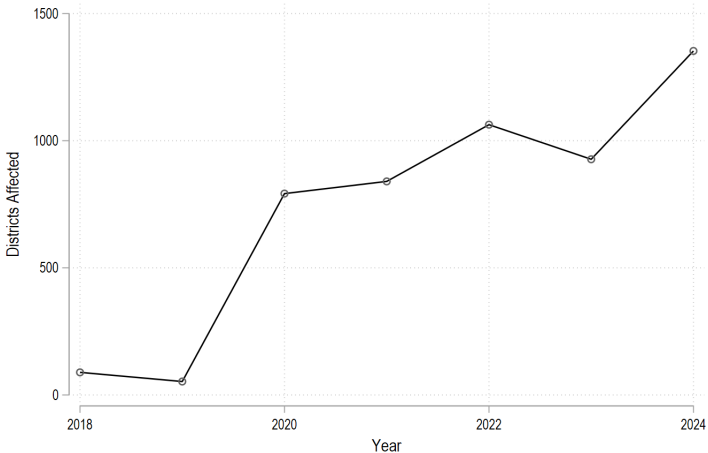
Data and Empirical Strategy

Data Sources

- **National Household Survey (*ENAHO*, 2012-2024):**
  - ✓ Cross-sectional and panel information of informal individuals.
  - ✓ Contains information on sales revenues, costs, and labor organization (paid vs unpaid) of self-employed individuals.
- **Administrative Records of *Centros de Desarrollo Empresarial* (2017-2024)**:
  - ✓ Tax code of every firm created via CDEs.
  - ✓ Contains information of date of entry, geographical location*, and 4-digit industry code.
- **Firm Registry from Peruvian Tax Authority (*Padrón RUC*, 2007-2024)**:
  - ✓ Panel of formal firms with information of their district, date of entry/exit, and 4-digit industry code.
- **National Permanent Employment Survey (*EPEN*, 2022-2024)**:
  - ✓ Quarterly cross-sectional and panel information of workers, representative at quarter-city level.

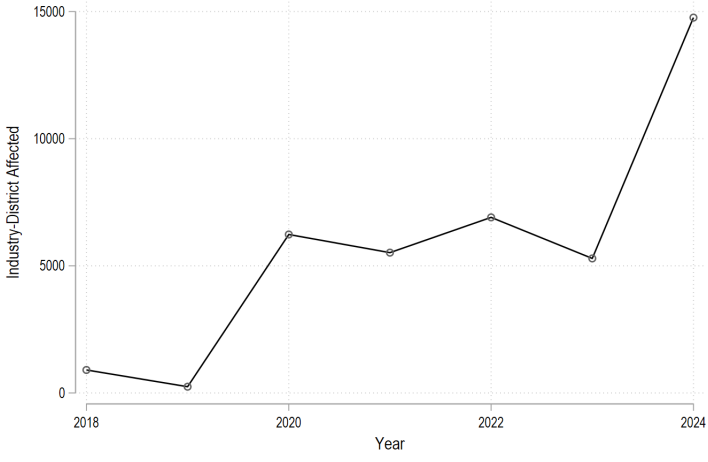
Rollout of Centros de Desarrollo Empresarial

Figure 1: Rollout Across Districts



Rollout of Centros de Desarrollo Empresarial

Figure 2: Rollout Across Local Markets (Industry × District)



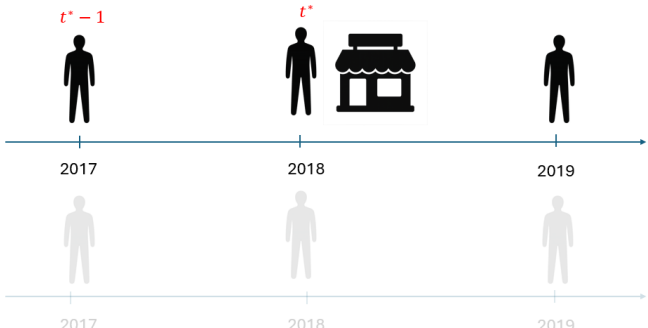
Empirical Strategy I: Differences-in-differences

Matched DiD

- We focus on all individuals in the ENAHO panel and restrict to those between 24 to 75 years old.
- **Treated Group:** Let t^* be the year an individual was exposed for the first time to a CDE born firm at their district and industry.
  - ✓ Reported as self-employed at $t^* - 1$ with non-zero profits.
  - ✓ At least one firm was created via CDE in their local market at t^*.
- **Potential Controls:** those who do not satisfy the previous conditions.
- We construct matched pairs between treated and potential controls based on characteristics at $t^* - 1$ (age, gender, sales revenues).

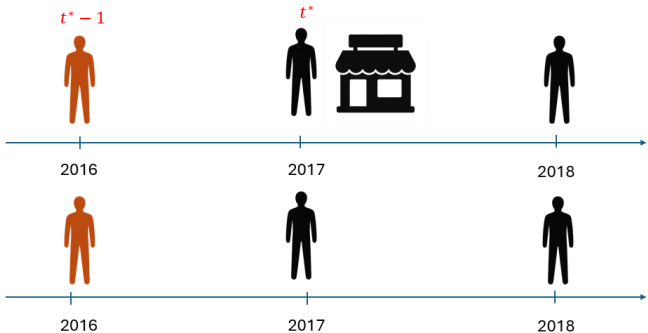
Matched DiD in a Rotating Panel

We do not restrict to a balanced panel, and thus different matched pairs contribute to separate treatment effects.



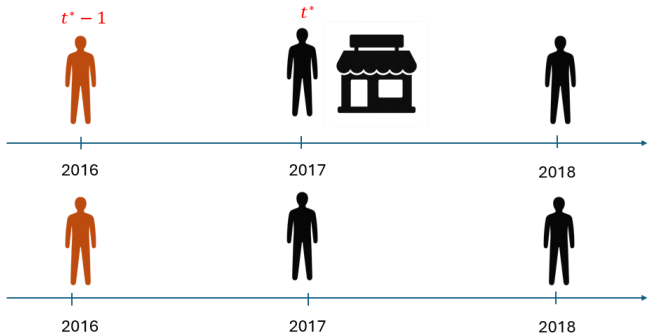
Matched DiD in a Rotating Panel

We do not restrict to a balanced panel, and thus different matched pairs contribute to separate treatment effects.



Matched DiD in a Rotating Panel

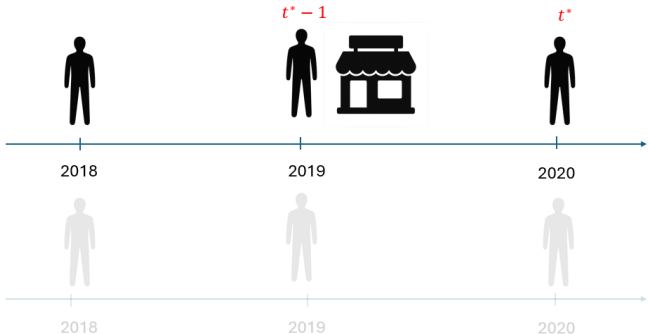
We do not restrict to a balanced panel, and thus different matched pairs contribute to separate treatment effects.



This matched pair contributes to identifying treatment effects at event-time $k = 0$ and $k = 1$.

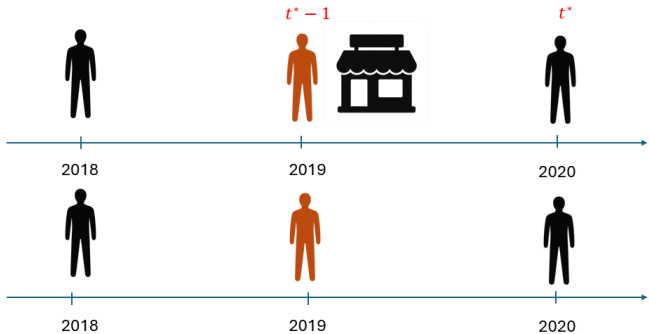
Matched DiD in a Rotating Panel

We do not restrict to a balanced panel, and thus different matched pairs contribute to separate treatment effects.



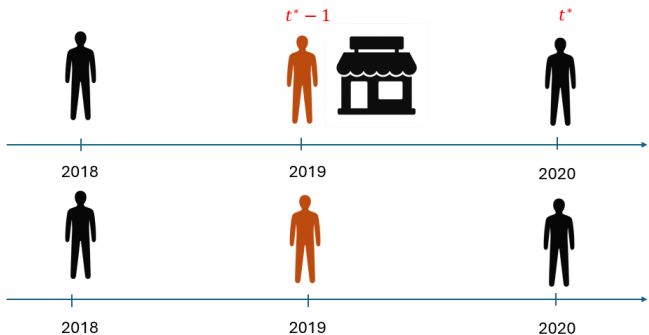
Matched DiD in a Rotating Panel

We do not restrict to a balanced panel, and thus different matched pairs contribute to separate treatment effects.



Matched DiD in a Rotating Panel

We do not restrict to a balanced panel, and thus different matched pairs contribute to separate treatment effects.



This matched pair contributes to identifying treatment effects at event-time $k = -2$ and $k = 0$.

Event-Study Model

Using the matched pairs we estimate the following model:

$$y_{it} = \alpha_i + \lambda_t + \sum_{k=-2}^{2} \gamma_k \mathbf{1}\{t = t_i^* + k\} + \sum_{k=-2}^{2} \theta_k \mathbf{1}\{t = t_i^* + k\} \times \mathsf{T}_i + X_{it}'\beta + \varepsilon_{it}, \tag{6}$$

where y_{it} measures our dependent variables (profits, employment, exit) and T_i takes one when individual belong to treated group.

Identification assumption: In the absence of a formal firm entry shock induced by CDEs, the outcomes of treated and control individuals would've followed parallel trends.

The parameters θ_k trace out the causal effect of the competition shock at event time k.

Event-Study Model

Using the matched pairs we estimate the following model:

$$y_{it} = \alpha_i + \lambda_t + \sum_{k=-2}^{2} \gamma_k \mathbf{1}\{t = t_i^* + k\} + \sum_{k=-2}^{2} \theta_k \mathbf{1}\{t = t_i^* + k\} \times \mathsf{T}_i + X_{it}'\beta + \varepsilon_{it}, \tag{6}$$

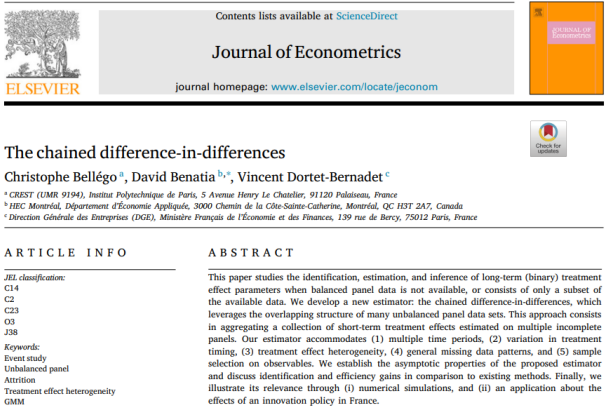
where y_{it} measures our dependent variables (profits, employment, exit) and T_i takes one when individual belong to treated group.

Identification assumption: In the absence of a formal firm entry shock induced by CDEs, the outcomes of treated and control individuals would've followed parallel trends.

The parameters θ_k trace out the causal effect of the competition shock at event time k.

Can we actually use a rotating panel in such a way?

Turns out that yes . . . although the same caveats of any DiD still apply.



Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

The chained difference-in-differences

Christophe Bellégo [a], David Benatia [b,*], Vincent Dortet-Bernadet [c]

[a] CREST (UMR 9194), Institut Polytechnique de Paris, 5 Avenue Henry Le Chatelier, 91120 Palaiseau, France
[b] HEC Montréal, Département d'Économie Appliquée, 3000 Chemin de la Côte-Sainte-Catherine, Montréal, QC H3T 2A7, Canada
[c] Direction Générale des Entreprises (DGE), Ministère Français de l'Économie et des Finances, 139 rue de Bercy, 75012 Paris, France

ARTICLE INFO

JEL classification:
C14
C2
C23
O3
J38

Keywords:
Event study
Unbalanced panel
Attrition
Treatment effect heterogeneity
GMM

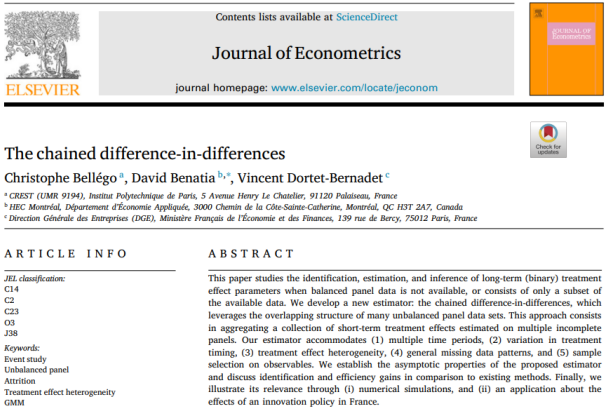
ABSTRACT

This paper studies the identification, estimation, and inference of long-term (binary) treatment effect parameters when balanced panel data is not available, or consists of only a subset of the available data. We develop a new estimator: the chained difference-in-differences, which leverages the overlapping structure of many unbalanced panel data sets. This approach consists in aggregating a collection of short-term treatment effects estimated on multiple incomplete panels. Our estimator accommodates (1) multiple time periods, (2) variation in treatment timing, (3) treatment effect heterogeneity, (4) general missing data patterns, and (5) sample selection on observables. We establish the asymptotic properties of the proposed estimator and discuss identification and efficiency gains in comparison to existing methods. Finally, we illustrate its relevance through (i) numerical simulations, and (ii) an application about the effects of an innovation policy in France.

* **Intuition:** if sampling does not depend on treatment assignment or outcome trends, it won't cause selection and we're good to go.

Can we actually use a rotating panel in such a way?

Turns out that yes . . . although the same caveats of any DiD still apply.



Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

The chained difference-in-differences

Christophe Bellégo [a], David Benatia [b,*], Vincent Dortet-Bernadet [c]

[a] CREST (UMR 9194), Institut Polytechnique de Paris, 5 Avenue Henry Le Chatelier, 91120 Palaiseau, France
[b] HEC Montréal, Département d'Économie Appliquée, 3000 Chemin de la Côte-Sainte-Catherine, Montréal, QC H3T 2A7, Canada
[c] Direction Générale des Entreprises (DGE), Ministère Français de l'Économie et des Finances, 139 rue de Bercy, 75012 Paris, France

ARTICLE INFO

JEL classification:
C14
C2
C23
O3
J38

Keywords:
Event study
Unbalanced panel
Attrition
Treatment effect heterogeneity
GMM

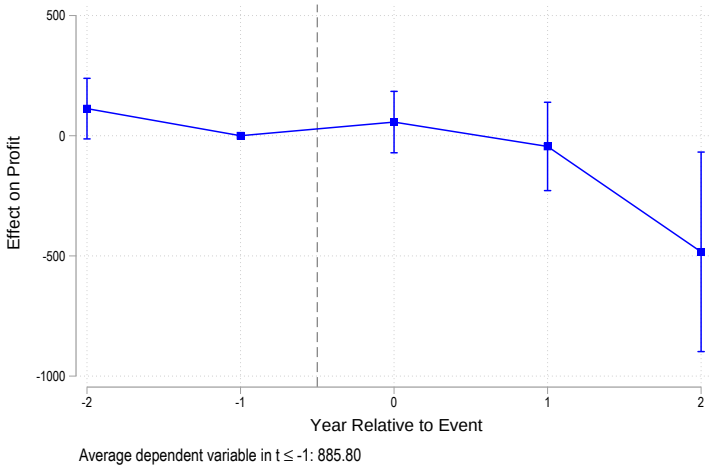
ABSTRACT

This paper studies the identification, estimation, and inference of long-term (binary) treatment effect parameters when balanced panel data is not available, or consists of only a subset of the available data. We develop a new estimator: the chained difference-in-differences, which leverages the overlapping structure of many unbalanced panel data sets. This approach consists in aggregating a collection of short-term treatment effects estimated on multiple incomplete panels. Our estimator accommodates (1) multiple time periods, (2) variation in treatment timing, (3) treatment effect heterogeneity, (4) general missing data patterns, and (5) sample selection on observables. We establish the asymptotic properties of the proposed estimator and discuss identification and efficiency gains in comparison to existing methods. Finally, we illustrate its relevance through (i) numerical simulations, and (ii) an application about the effects of an innovation policy in France.

* **Intuition:** if sampling does not depend on treatment assignment or outcome trends, it won't cause selection and we're good to go.

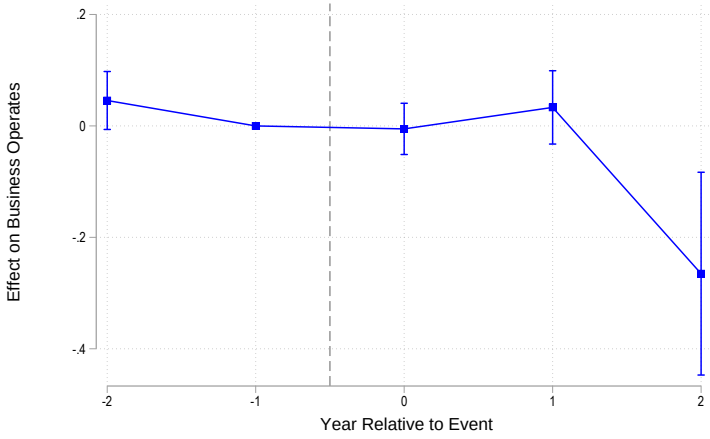
Preliminary Results: Profits

We find a drop in firm profits (ext + int margin).



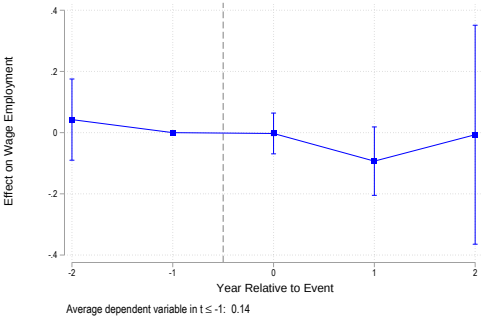
Preliminary Results: Business Operates

We also observe a large decline in the probability that the business is still active.

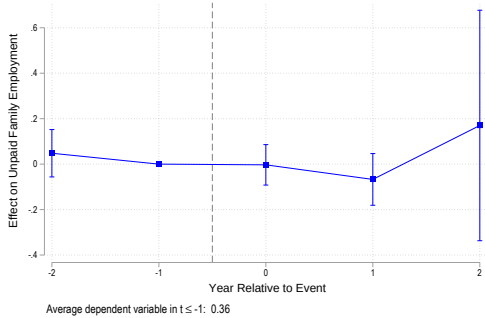


Preliminary Results: Labor Mix

Finally, we find no significant change in the labor mix.



(a) Paid Labor



(b) Unpaid Labor

Taking Stock: Matched DiD

- We find a decline in firm profits at $k = 2$.
- It is not purely an intensive margin effect, since we find an increase in firm exits at $k = 2$.
- We also do not find evidence that there is substitution towards unpaid labor.
- These results seem consistent with an impact driven only through the rent erosion mechanism.

Empirical Strategy II:
Jackknife IV Estimation (JIVE)

The Identification Challenge

To leverage the pooled ENAHO microdata, we turn our focus to the following model:

$$y_{it} = \beta D_{\ell(i,t),t} + \alpha_{d(i,t)} + \lambda_{r(i,t),t} + X_{it}'\gamma + \varepsilon_{it}, \tag{7}$$

where $\ell(i,t)$ and $r(i,t)$ map the local market (industry s and district d) and region of individual i at time t, respectively. Furthermore,

$$D_{\ell,t} = \frac{\text{New Formal Firms in } \ell \text{ at time } t}{\text{Incumbent Formal Firms in } \ell \text{ at time } 2016} \tag{8}$$

Problem: Formal firms might enter local market because local demand at d is booming.

The Identification Challenge

To leverage the pooled ENAHO microdata, we turn our focus to the following model:

$$y_{it} = \beta D_{\ell(i,t),t} + \alpha_{d(i,t)} + \lambda_{r(i,t),t} + X_{it}'\gamma + \varepsilon_{it}, \tag{7}$$

where $\ell(i,t)$ and $r(i,t)$ map the local market (industry s and district d) and region of individual i at time t, respectively. Furthermore,

$$D_{\ell,t} = \frac{\text{New Formal Firms in } \ell \text{ at time } t}{\text{Incumbent Formal Firms in } \ell \text{ at time } 2016} \tag{8}$$

Problem: Formal firms might enter local market because local demand at d is booming.

Constructing the Jackknife Instrument

We construct a **Leave-One-Out Entry Shock** using administrative data from CDEs. In particular, given the district d and sector s of local market ℓ, we compute:

$$Z_{\ell,t} = \frac{\sum_{j \neq d} \text{New CDE Firms in district } j\text{, sector } s \text{ and time } t}{\sum_{j \neq d} \text{Incumbent Formal Firms in district } j\text{, sector } s \text{ and time } 2016} \tag{9}$$

Intuition:

- $Z_{\ell,t}$ captures a national wave of formalization in industry s driven only by the policy (reduced entry cost).
- By excluding district d, we mechanically purge any local demand shock specific to district d.

Econometric Specification (2SLS)

We estimate the impact on individual informal worker i in district d, industry s, at time t:

First Stage (Predicting Local Entry):

$$D_{\ell(i,t),t} = \pi Z_{\ell(i,t),t} + \alpha_{d(i,t)} + \lambda_{r(i,t),t} + X'_{it}\gamma + \varepsilon_{it}$$

Second Stage (Impact on Incumbents):

$$y_{it} = \beta \widehat{D}_{\ell(i,t),t} + \alpha_{d(i,t)} + \lambda_{r(i,t),t} + X'_{it}\gamma + \varepsilon_{it}.$$

- We expect a negative β on profits and the likelihood of still operating the business.
- For the case of the ratio of paid vs. unpaid labor and informal wages, sign(β) will test against one of the mechanisms.

Empirical Strategy III:
Shift Share IV Estimation (SSIV)

Towards high frequency responses

Given that some labor market adjustments (profits, wages and unemployment) happen fast, we opt to use EPEN on a city-quarter panel of outcomes.

- Importantly, EPEN is designed to be representative for **27 major cities**.
- We collapse individual microdata to create the panel of outcomes at this level.
- We restrict to data from 2022-Q1 up to 2024-Q4.

Constructing the Bartik Instrument (B_{ct})

We predict the formal entry shock for city c based on its historical industry s composition:

$$Z_{ct} = \sum_s \underbrace{\left(\frac{\text{Formal Emp}_{c,s,2016}}{\text{Total Emp}_{c,2016}} \right)}_{\text{Share (Local Exposure)}} \times \underbrace{g_{s,t}^{-c}}_{\text{Shift (National Shock)}}$$

- **Share (weights):** Did city c specialize more heavily in industry s prior to the policy?
- **The Shift (shocks):** The national entry rate of CDE firms in industry s, excluding city c.

Intuition: If CDE policy triggers a national firm entry boom in sector s, cities that were already specialized will get a high predicted shock.

Econometric Specification

We focus on the reduced form effect on aggregate city outcomes based on the predicted shock:

$$\overline{y}_{ct} = \alpha + \beta Z_{ct} + \mu_c + \tau_t + \varepsilon_{ct}$$

The sign of β on **Informal Wages** adjudicates the mechanism:

Hypothesis A: Labor Rationing $(\beta > 0)$

- Formal entry increases aggregate labor demand.
- Markets are tighter; firms must pay more to attract workers.
- *"Rising tide lifts all boats."*

Hypothesis B: Rent Erosion $(\beta \leq 0)$

- Formal entry crowds out informal incumbents.
- Displaced workers increase labor supply, dampening wages.
- *"Efficiency gains hurt low-skill wages."*

Discussion and Future Steps

Discussion and Future Steps

- Using ENAHO's rotating panel, we find effects consistent with the rent erosion mechanism.
  - Informal incumbents profits fall and exit rises.
  - *Limitation:* low statistical power to test both channels.
- To rigorously adjudicate between the mechanisms, we will implement two additional designs:
  - **Jackknife IV (JIVE)**
    - ✓ **Data:** Pooled ENAHO cross-sections.
    - ✓ **Goal:** Leverage massive statistical power to detect subtle shifts in **labor composition**.
  - **Shift Share IV (SSIV)**
    - ✓ **Data:** EPEN (high-frequency city-level).
    - ✓ **Goal:** Test the mechanisms based on the general equilibrium response of **informal wages** at the aggregate level.

¡Thank you!

paulj.corcuera@udep.edu.pe

References

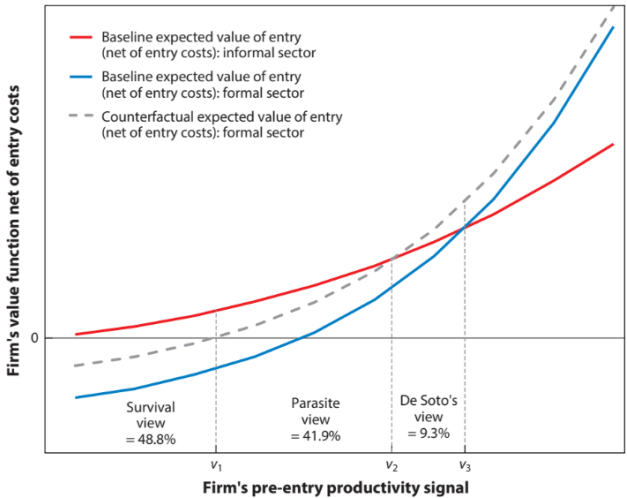
- De Mel, S., McKenzie, C., & Woodruff, C. (2013). *The Demand for, and Consequences of, Formalization among Informal Firms in Sri Lanka*. *American Economic Journal: Applied Economics*, 5(2), 112-150. doi.org/10.1257/app.5.2.122

- Ulyssea, G. (2018). *Firms, Informality and Development: Theory and Evidence from Brazil*. *American Economic Review*, 108(8), 2015–2047. doi.org/10.1257/aer.20141745

- Ulyssea, G. (2020). *Informality: Causes and Consequences for Development*. *Annual Review of Economics*, 12, 525–546. doi.org/10.1146/annurev-economics-082119-121914

Appendix

The Taxonomy of Informality (Ulyssea, 2018)



Volver