

Monopsony Makes It Big

IAB Conference on Imperfect Competition in the Labor Market

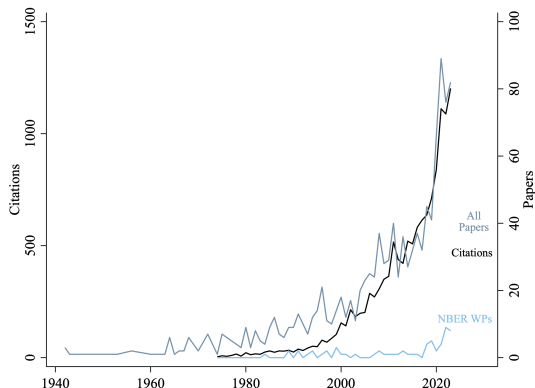
Keynote, July 2026

Sydnee Caldwell

UC Berkeley & NBER

Based on “*Monopsony Makes It Big*,”
joint work with Arindrajit Dube & Suresh Naidu

The rise of monopsony



A field of **booms and busts**:

- **1933**: Robinson coins “monopsony.”
- **Mid-century**: treated as a curiosity.
- **1990s**: revival, after Card & Krueger (1995).
- **2010s →**: an outright explosion — fueled by matched worker–firm data and cheaper experiments.

This talk

1. What is monopsony power — and how much do firms have?

- One measurable object: the firm-specific labor supply elasticity η .
- 91 studies — experiments, quasi-experiments, structural models — say median $\eta \approx 2$: a markdown of $\sim 33\%$.

2. Where does it come from — why don't workers move?

- A unifying framework: limited **consideration sets** + idiosyncratic tastes.
- Survey experiments: we asked 10,000 German workers: **preferences**, not information, keep workers put.

3. What do firms do with that power?

- If the wage also buys **effort**, the firm may not even be on its supply curve — efficiency wages meet monopsony: “*monergony*.”

Defining and Sourcing Monopsony Power

Two common approaches to studying labor market power

1. Measure the markdown: $(MRPL - w)/w$ — power as exercised **over wages**

- Estimate marginal products from production/revenue functions; compare to wages.
- Requires estimating MRPL *and* assumptions on firm conduct.
- Ideal if you are interested in the markdown itself.

2. Measure the labor supply elasticity: $\eta = \partial \log L / \partial \log w$ — *latent capacity*

- Causal response of employment (or quits, applications) to the firm's *own* wage.
- A feature of the firm's environment, not of its conduct.

Under the textbook Lerner rule the two coincide: $\text{markdown} = 1/\eta$.

What is market power?

The U.S. Department of Justice:

“Market Power” is the ability of a firm . . . to profitably ~~charge prices above~~ pay wages below the competitive level for a sustained period of time.

Robinson (1933): exploitation arises because the supply of labor is
“imperfectly elastic to the unit of control.”

Manning (2003):

“What happens if an employer cuts the wage it pays its workers by one cent?”

Same answer: power is a **finite labor supply elasticity**.

What this definition does — and doesn't — imply

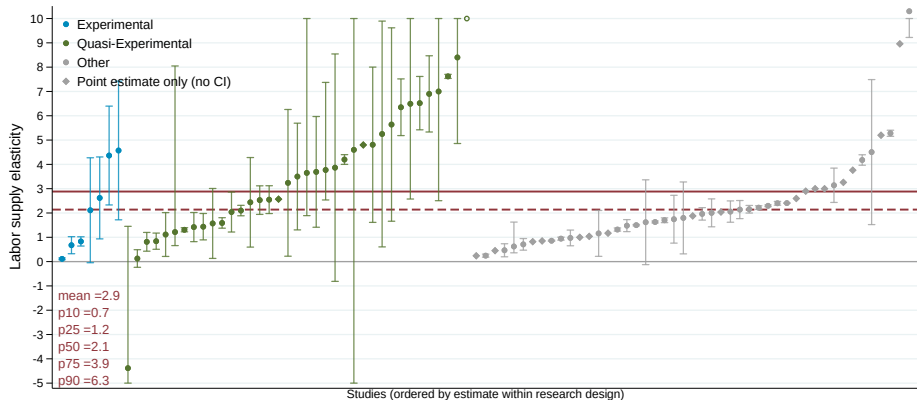
Power is about the environment, not the number of employers.

- Firms can have power over their workers even with *many* other employers in the market.
- Conversely, a very large employer may have *little* power — if workers view other jobs as close substitutes.

Latent power may not be exercised power.

- Textbook benchmark: labor supply the *only* constraint, the wage the *only* instrument $\Rightarrow$ markdown $= 1/\eta$.
- A firm can face finite η yet *not* mark down wages:
 - Binding minimum wage $\Rightarrow$ pays full MRPL, but a wage cut would still not lose all workers.
 - Fairness norms, efficiency wages, bargaining ...

How much monopsony?



Mean ≈ 2.9 , **median ≈ 2.1** (IQR 1.2–3.9). $\eta \approx 2 \Rightarrow$ markdown of $\sim 33\%$.

Varieties of methodologies

Estimation Strategy	Average	Std. Dev.	Median	N
Calibrated model	3.24	2.73	2.60	7
Experimental	2.18	1.78	2.11	7
Non-experimental	2.16	1.95	1.72	30
Production function	2.06	1.41	1.96	10
Quasi-experimental demand-shock	3.65	2.87	3.58	20
Quasi-experimental wage-shock	3.88	4.08	2.55	17

- **True heterogeneity** across firms, markets, and horizons (short-run $\ll$ long-run: $\eta \approx 1-2$ vs. 4-6).
- **Identification**: wage-shock & experimental designs cluster at $\eta \approx 2$; demand-shock designs run ~ 3 ; differences in complier characteristics.

A unifying framework: consideration sets

Herbert Simon (1955):

“The organism may make its choice within a set of alternatives more limited than the whole range objectively available to it.”

Two ingredients:

1. Workers choose the best option within a **limited** consideration set S_i (which always includes nonemployment).
 - Search (random or directed), recruitment/marketing, networks, discrimination.
2. **Random utility** over wages and amenities *within* the set.

The worker's problem

- Worker i gets $V_i^j(w_j) = U(w_j) + \epsilon_{ij}$ from firm j ; ϵ_{ij} i.i.d. tastes.
- Consideration set S_i (random, includes nonemployment $j = 0$).

$$P[i \text{ chooses } j \mid w_j, \mathbf{w}_{-j}, S_i, b_i] = P\left(j = \arg \max_{k \in S_i} V_i^k(w_k)\right)$$

Under independence, this factors into three pieces:

$$\underbrace{P(j \in S_i \mid w_j)}_{\text{in consideration set}} \times \underbrace{\prod_{k \in S_i} P(V_j > V_k)}_{\text{beats other offers}} \times \underbrace{P(V_j > b_i)}_{\text{beats outside option}}$$

Under some assumptions, η decomposes

$$\eta_j = \underbrace{\eta_j^{\text{Consideration}}}_{\substack{\text{directed search} \\ (=0 \text{ if random})}} + \underbrace{\sum_{k \neq j, k > 0} \eta_{jk}}_{\text{cross-firm}} + \underbrace{\eta_{j0}}_{\text{market}}$$

More elastic to the *firm* than the *market*; more firms in $S_i \Rightarrow$ more elastic.

The **consideration set** S_i is itself shaped by many forces:

- **Networks** (Caldwell & Harmon); **Discrimination** (Bertrand & Mullainathan; Kline et al.);
- **Cognitive frictions** (Matejka & McKay); **Platforms** that measure — or *manipulate* — sets.

So the nature of search — a long-standing macro-labor question — affects the elasticity.

See paper for a more careful treatment of the decomposition and its assumptions.

One framework nests the three workhorse models

Size / Concentration

No tastes, no frictions. One firm $\Rightarrow$ extensive-margin supply $F_b(w)$. Cournot:
 $\eta_j = \eta_0 / s_j$.

Preference heterogeneity

Many firms, full consideration, T1EV tastes
 $\Rightarrow$ CES / logit: $P(j) \propto w_j^\eta$. Berger et al. (2022) add concentration.

Search frictions

No tastes, random offers $\Rightarrow$ static search (Mortensen 2003):

$$P(j) = e^{-\lambda(1-F(w_j))}.$$

Dynamic limit nests Burdett–Mortensen (1998) and Bloesch–Larsen (2026).

Separation elasticity $\approx \frac{1}{2}\eta$ in steady state.

Two questions this raises

1. The nature of search. The consideration term is nonzero only if workers **direct** their search.

- Is directed search meaningful?
- Is limited mobility about *search frictions* or *preferences*?

→ CHH surveys (next)

2. Are we measuring labor supply? η assumes employment moves *along* the labor supply curve.

- What if the wage also buys **effort** — so the firm sits *off* its supply curve, and headcount $\neq$ labor supply?

→ monergony (later)

What Do Workers Know — and Why Do They Stay?

Can we just ask workers?

The ideal experiment: randomly vary the wage a worker faces at each firm and watch where they go — tracing labor supply directly.

Feasible alternative: **survey experiments**

- Validated against the field: survey elasticities line up with experiments (Mas & Pallais 2017).
- Not a new idea — it was cool in the 1950s too (Reynolds 1951; Myers & Shultz 1951). Sociologists never stopped.
- **Caveat:** survey design is an art. A discrete-choice experiment beats “how much would you pay to move from A to B?” (see Mas 2025 handbook).

Our approach: just ask 10,000 workers — through the IAB

Suppose you planned to move to a new company in the next {one/three/six} months. Would you consider applying to any of these? Please select all that apply.

- ☐ Company 1
- ☐ Company 2
- ☐ Company 3
- ☐ Company 4
- ☐ Company 5
- ☐ Company 6
- ☐ Company 7
- ☐ I would not apply to any of these

What do you think your gross annual pay would be if you worked at these companies in a position similar to your current one?

Company 2: [Fill in gross pay]

Company 4: [Fill in gross pay]

Company 7: [Fill in gross pay]

Suppose you can remain at your current company or switch to any of the companies listed below and immediately receive the raise specified.

Please rank the following job offers from 1 to 4 where 1 is the offer you are most likely to take and 4 is the offer you are least likely to take.

Company 2 with a X% raise _____

Company 4 with a Y% raise _____

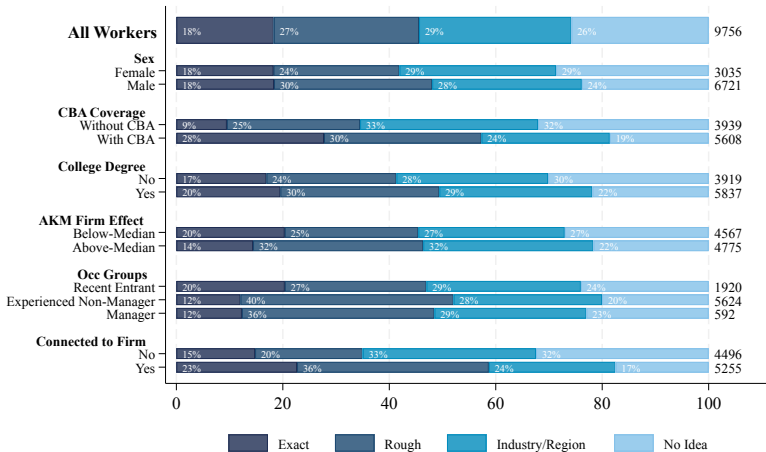
Company 7 with a Z% raise _____

Remain at current firm at current pay

Joint work with **Ingrid Haegele & Jörg Heining (IAB)**.

Many knew the pay *before* they ever applied

At the time that you applied, did you know what salary you would earn?



Workers perceive a firm component to pay

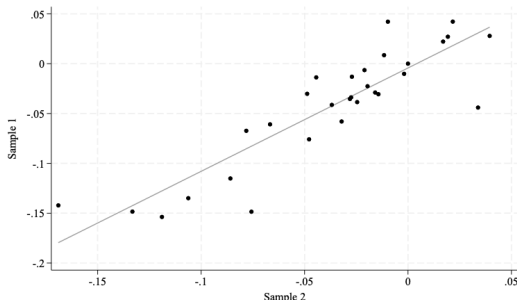
What would your salary be at these companies, in a position like your current one?

Decompose *beliefs* about pay at named firms, within worker across firms:

$$\log \tilde{w}_{ij} = \alpha_i + \psi_j + \epsilon_{ij}$$

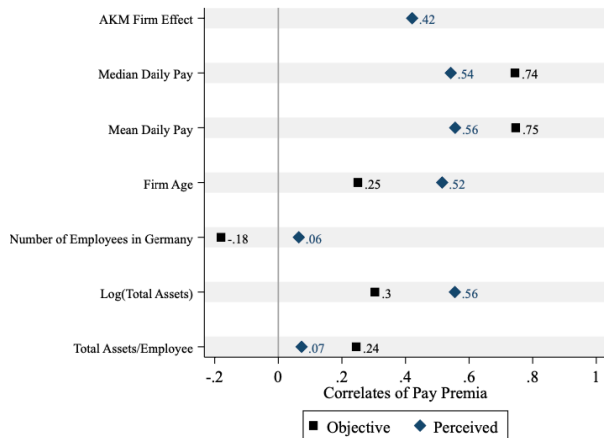
Split-sample ψ_j

Split-Sample Estimates



- ψ_j : perceived *firm* premium (vs. a normalized base firm).
- Estimate ψ_j on two random halves of workers — they line up.
- **The firm component in beliefs is real, not noise.**

People know some things about wages



- Workers' *perceived* pay premia at other firms line up with AKM firm effects (corr. ≈ 0.4).
- Beliefs also track objective fundamentals: pay levels, firm age, assets.

Hard to square with random-search models that assume workers are in the dark about who pays well.

How many workers direct their search? A two-type mixture

Each worker is one of two types — consideration flat in pay (*undirected*) or tilted toward higher-paying firms (*directed*):

$$\Pr(\text{consider}_{ij}) = \Lambda(a + X'_{ij}\delta + \pi_i b \psi_j), \quad \pi_i \in \{0, 1\}$$

- Regressor is the firm's **administrative** pay premium ψ_j — what workers *know*, not what they say.
- Finite-mixture (spike-and-slab) logit, EM on 9,756 workers.

$\hat{\pi} \approx 0.38$ (SE 0.02) direct their search. [$\pi = 0$: pure random; $\pi = 1$: pure directed]

$\eta_j^{\text{Consideration}}$ rises with $\hat{\pi}$ — a meaningful share of firm-level power runs through **directed** search.

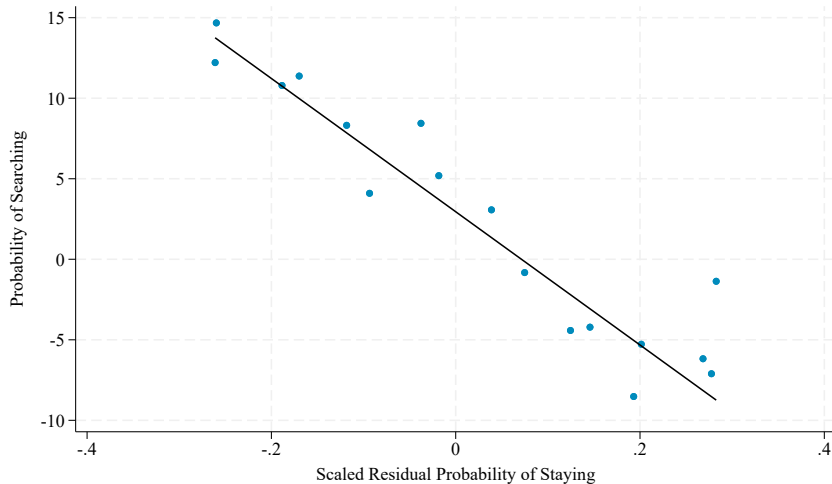
What counts as “the market” affects the measured elasticity

Mis-measuring workers' consideration sets makes workers look inelastic:

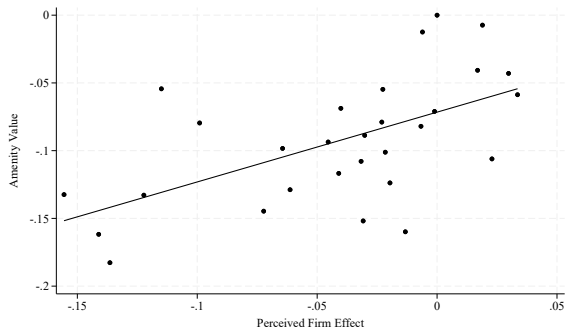
	Move to a Researcher- Provided Firm (1)	Move to a Worker-Provided Firm		
		Baseline (2)	Same Commute (3)	Same Growth (4)
Outside Pay	0.391*** (0.133)	1.566*** (0.236)	1.288*** (0.364)	0.945** (0.368)
Constant	0.321*** (0.018)	0.359*** (0.033)	0.425*** (0.042)	0.408*** (0.043)
Implied Elasticity	1.219** (0.477)	4.364*** (1.037)	3.031*** (1.139)	2.320** (1.130)
Observations	21872	12217	6264	6272

... attachment leads to lack of search

Workers who don't want to leave don't search...



Amenities do not “undo” wage variation



- From ranked choices we recover a firm-specific *amenity* value a_j alongside the pay premium ψ_j .
- a_j is **positively** correlated with perceived pay premia — high-pay firms tend to be high-amenity.
- No compensating-differential trade-off: some firms are just *better to be at*.

Could better information unlock mobility? Mostly not

Calibrate the model to the data — compressed beliefs ($\hat{\mu} \approx 0.3$), estimated switching costs, and $\hat{\pi} \approx 0.38$ directed searchers — then run counterfactuals.

Mobility gain, low-wage workers:

- Full pay **transparency**: +24.7 **pp**.
- Eliminate firm **attachment**: +34.5 **pp**.
- Even handing every worker the median firm's pay lifts mobility just +2.7 **pp** vs. +28.5 **pp** from removing attachment ($\sim 10\times$).

Preferences, not information frictions, are the binding constraint on mobility.

This isn't to say that providing information is a “bad” idea, it's just not sufficient!

Monergony

Are we even on the labor supply curve?

Estimating η assumes the firm sits *on* its labor supply curve — that employment responds to the wage by moving *along* supply.

- But if the wage also buys **effort**, the firm may be *off* it:
 - Screen & ration: willing workers turned away — effective, not raw, supply.
 - Demand-constrained: labor supply *slack*, jobs rationed — **involuntary unemployment**, wage pinned by a Solow condition, not supply.
- Then headcount no longer traces labor supply, and estimated η isn't the supply elasticity.

⇒ efficiency wages meet monopsony: “monergony.”

Efficiency wages meet monopsony: “monergony”

In reality, the wage may also affect **productivity**, not just who shows up.

Labor discipline: effort is binary, costs c ; shirkers are caught & fired with prob. P .

Worker i exerts effort iff

$$w - c \geq (1 - P)w + P b_i \quad \Longleftrightarrow \quad w \geq \phi + b_i, \quad \phi \equiv \frac{c}{P} \text{ (the no-shirking premium).}$$

- Unlike textbook efficiency wages, *one* wage cannot satisfy all workers' IC constraints.
- Output depends on **effective labor** $e(w)L \Rightarrow$ a new margin in wage-setting.
- Effort considerations *generally push wages above* the pure monopsony level.

Monerergy depends on screening and on which constraint binds

Firm solves $\max_{w,L} G(e(w)L) - wL$ s.t. $L \leq F(w)$, with

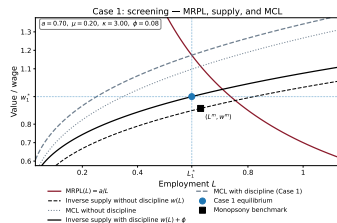
$$e(w) = \begin{cases} \frac{F(w - \phi)}{F(w)} & \text{no screening on } b_i, \\ 1 & \text{screening feasible.} \end{cases}$$

Three cases:

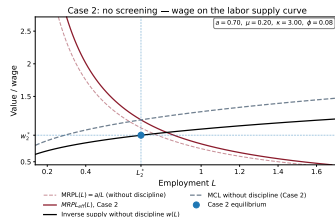
1. Full screening — labor supply binds; hire only non-shirkers.
2. No screening, supply-binding — some hires shirk.
3. No screening, demand-binding — a Solow / efficiency-wage region.

The three cases, in one picture

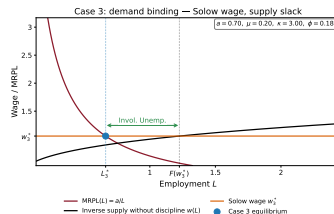
1. Full screening



2. Supply binds, no screening



3. Solow region



Wages *above* pure monopsony;
markdown shrinks. Efficient
rationing.

Markdown on *effort-adjusted*
MRPL; effort can dampen or
amplify it. Equilibrium shirking.

No markdown — but **involuntary**
unemployment; more
competition $\Rightarrow$ higher wages.

Why monopsony matters

- **Involuntary unemployment can coexist with monopsony power** — firms screen on outside options.
- Headcount need not identify η : the relevant margin is *effective* labor.
 - Reinforces focusing on clean behavioral margins: separations, quits, applications.
- Helps square low η with the **labor share**:
 - Markdown rule $\times L/Y$: $\eta = 3$ and output-elasticity 0.63–0.81 $\Rightarrow$ labor share 0.47–0.6 — too low.
 - $\Rightarrow$ markdowns alone can't be the whole story.

Conclusions and Future Directions

Bottom line

- η is a clean summary of the heterogeneity in workers' outside options — and of a firm's **latent capacity** to lower wages.
- Expect *diverse* estimates: across individuals, firms, markets, horizons. Median ≈ 2 .
- Firm wage-setting is richer than the Lerner rule: amenities, recruitment, effort, fairness, mis-optimization, bargaining.
- Monopsony is now big enough to interact with most of labor economics.

Open questions

Measuring power

- More direct estimates of η — replicate gig designs; pair surveys with admin data.
- Trace the *whole* supply curve, away from the current wage; relax constant- η .
- Are we measuring labor supply, or *effective* labor? (monergony, partial screening)

Conduct & incidence

- Are monopsony profits just quasi-rents (recruiting, training, entry)?
- Protocols & collusion: how common are bargaining, discrimination, no-poach?
- Evolution of η and conduct over time — largely unmapped.

Policy & the future

- Which tools actually move η — non-competes, no-poach, mobility limits, transparency?
- AI agents that apply to and target jobs at scale could reshape search and recruiting costs — likely new mechanisms of power, not a new balance.

Thank you!

Sydnee Caldwell | scaldwell@berkeley.edu