Learning on the Job

Jacob Adenbaum[1] Fil Babalievsky[2] William Jungerman[3]

[1]University of Edinburgh
[2]Census Bureau
[3]UNC Chapel Hill

November 2024

Disclaimer: Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau.

Introduction

Question: What are the determinants of on-the-job learning?

- First-order to study sorting, monopsony, and human capital accumulation
- Several potential sources:
  - Intrinsic own learning ability
  - Firm learning environment
  - Composition of coworkers
- Challenges:
  1. Human capital is not observable $\rightarrow$ need a model
  2. Any model with all these features has historically been intractable

What we do

- **Theory:** Extend Postel–Vinay and Robin (2002) to accommodate
  1. Arbitrarily large multi-worker firms
  2. Rich two sided heterogeneity in firm and worker productivities and learning characteristics
  3. Complementarities in production and learning across workers
- **Computation:** Overcome curse of dimensionality by
  - Approximating key model objects with neural networks
  - Exploiting recent advances in deep learning
- **Measurement:** Calibrate to French matched employer-employee admin data (DADS)
  - Observe coworker composition for near-universe of French workers/firms
  - Detailed wage and hours data; granular occupation codes

What we find

- **Learning:** Learning from more skilled coworkers is dominant source of learning on the job
  1. Accounts for more than 50% of the variance in human capital growth rates
  2. Remainder split between learning ability (1/3) and firm effects (2/3)
  3. Switching off learning from coworkers decreases human capital and wages 25%
- Two key **sorting motives:**
  1. Production complementarities (worker/firm **and** worker/coworkers) induce positive assortative matching
  2. Learning complementarities (worker/coworkers) induce negative assortative matching

$\rightarrow$ production motive dominant for low human capital workers

$\rightarrow$ training motive dominates production gains at high human capital levels

Related Literature

- **Peer Effects in Labor Markets:** Jarosch, Oberfield, and Rossi-Hansberg (2021), Freund (2024), Herkenhoff, Lise, Menzio, and Phillips (2024), Ma, Nakab, and Vidart (2024)

  Contribution:

  1. Whole distribution of coworkers matters for learning and wages
  2. Much richer patterns of sorting and selection

- **Machine Learning in Economics:**
  - Methods Papers: Maliar, Maliar, and Winant (2021), Kahou, Fernandez-Villaverde, Perla, and Sood (2022), Azinovic, Gaegauf, and Scheidegger (2022), Duarte, Duarte, and Silva (2023)
  - Applications: Duarte (2022), Jungerman (2023)

  Contribution: *heterogeneously sized* state spaces

Model

Environment

Time is continuous (omit time subscripts) , populated by a continuum of workers and firms:

Workers

- Indexed by $i \in [0, N_w]$
- Linear preferences, discount rate ρ
- Heterogeneous in
  1. General human capital h_i
  2. Fixed learning ability a_i
- Workers "retire" at rate δ_r, replaced with draws from G_w
- New workers start unmatched

Firms

- Indexed by $k \in [0, 1]$
- Linear preferences, discount rate ρ
- Heterogeneous in
  1. Fixed productivity z_k
  2. Fixed learning environment q_k
- Firms die at rate δ_f, replaced with draws from G_f
- New firms start unmatched
- Firms consist of n_k matched workers

The Firm State

- Firm state consists of (z_k, q_k) and the **set** of all the states of its workers:
  - Let $\mathcal{W}_k$ be the set of all workers matched to a firm k
  - Define the state of each worker as $\mathbf{x}_i := (h_i, a_i, w_i)$
  - The firm's workforce is $X_k := \{\mathbf{x}_i \mid i \in \mathcal{W}_k\}$
  - We define the firm state $S_k := (z_k, q_k, X_k)$
- **Helpful notation:**
  - Adding a worker to the firm: $S_k \oplus \mathbf{x}_i := (z_k, q_k, X_k \cup \mathbf{x}_i)$
  - Removing a worker from the firm: $S_k \ominus \mathbf{x}_i := (z_k, q_k, X_k \backslash \mathbf{x}_i)$

Technology

- Augment Postel–Vinay and Robin (2002) to add complementarities in two ways:
  1. **Production:** Flow output $F(S_k)$ depends on all worker/firm states Functional Form
  2. **Learning:** All workers at the firm update human capital at rate γ^E, according to a human capital production function $H(S_k)$ Functional Form

→ **Implication:** values are *not separable* across matches:

- Workers must pay attention to the **entire distribution** of coworkers
- Firms care about the effect worker i has on coworker j's human capital growth in the future
- Wage setting must take account of this

Meetings and Matches

- Workers and firms match in a **frictional** labor market

  Firms can match with many workers, but a worker can match with only one firm at a time

- **Technology:** each worker generates meetings at rate ψ^N if unmatched or ψ^E if matched
  - Meetings are allocated uniformly to workers, proportional to match generation
  - Meetings are allocated to firms proportional to firm size $n_k + 1$

    $\rightarrow$ for Gibrat's law, otherwise large firms could not grow as fast (in proportional terms) as small firms
  - Analogous to balanced matching (Burdett and Vishwanath 1988)
- Firms and workers may agree on a wage w_i and form a match
- Matches can be terminated unilaterally, but only at **stochastic intervals:**
  1. Renegotiation shocks which occur at a rate λ

     $\rightarrow$ avoids multilateral negotiations, but means some matches can persist with negative surplus
  2. When the worker meets another firm (at a rate ψ^E)
- Matches can also exogenously separate at rate δ_m

Bargaining

- **Standard** assumptions following Postel–Vinay and Robin (2002):

(A1) Wages conditional on worker states (h_i, a_i) and incumbent firm states if poaching

(A2) Firms make counter-offers when rival firm attempts to hire one of their workers

(A3) Wages are take-it-or-leave-it offers

$\rightarrow$ ensure the familiar sequential auctions bargaining solution, with bilaterally efficient matches

- **Additional** assumptions:

(A4) Wage contracts only renegotiated by mutual consent, **at stochastic intervals**
$\rightarrow$ avoids firm simultaneously negotiating with multiple workers

(A5) When hiring and firing, firms maximize the joint value of the *full* coalition
$\rightarrow$ abstracts away from incentive compatibility problems between firm and workers and aligns their incentives (similar to Herkenhoff, Lise, Menzio, and Phillips 2024)

(A6) When either worker or firm can credibly threaten to end the match, the wage adjusts to the closest boundary of the bargaining set
$\rightarrow$ minimizes variance of wages and necessary when something could happen between renegotiation events that pushes the worker outside the bounds (Hall 2005 and Thomas and Worrall 1988)

Nonemployed Value

- Unmatched workers receive flow benefits proportional to b times their human capital level
- Take it or leave it offers mean worker values are unchanged when accepting a job out of nonemployment
- Let $U(h_i)$ denote the value of nonemployment

$$U(h_i) = \frac{bh_i}{(\rho + \delta_r)}$$

Note this is independent of learning ability a_i

Separation Policies

- Let $V(S_k)$ denote the present value of a firm and all its matched workers
  - Linear utility and counteroffers $\Rightarrow$ wages are **purely allocative**
- Define the **surplus** of the match between worker $\mathbf{x}_i$ and firm S_k to be

$$\Delta(S_k, \mathbf{x}_i) := V(S_k) - V(S_k \ominus \mathbf{x}_i) - U(\mathbf{x}_i)$$

- There are three ways a match can terminate:
  1. Renegotiation shock, if $\Delta(S_k, \mathbf{x}_i) < 0$
  2. Worker is poached
     - $\rightarrow$ Change in poaching firm's value is $\mathcal{B}$ and depends on incumbent surplus and poacher surplus
     - $\rightarrow$ We characterize this in a proposition Proposition
  3. Exogenous match break shock δ_m

Joint Value

$$
\begin{aligned}
\rho V(S_k) = \; & \underbrace{F(S_k)}_{\text{Flow output}} - \underbrace{\delta_f \left(V(S_k) - \sum_{i \in \mathcal{W}_k} U(\mathbf{x}_i) \right)}_{\text{Firm Death}} + \underbrace{\gamma^E \left[V(H(S_k)) - V(S_k) \right]}_{\text{Learning}} \\
& + (n_k + 1)\omega \left[\underbrace{s^N \int \max\left\{ \Delta(S_k \oplus \mathbf{x}_j, \mathbf{x}_j), 0 \right\} \mathrm{d}\chi^N(\mathbf{x}_j)}_{\text{Meet Unmatched}} + \underbrace{s^E \int \mathcal{B}(S_k, S_{p(j)}, \mathbf{x}_j) \mathrm{d}\chi^E(\mathbf{x}_j, S_{p(j)})}_{\text{Meet Matched}} \right] \\
& + \underbrace{\sum_{i \in \mathcal{W}_k} (\delta_r + \delta_m) \left[V(S_k \ominus \mathbf{x}_i) - V(S_k) \right] + \delta_m U(\mathbf{x}_i)}_{\text{Match Breaks and Retirement}} + \underbrace{\sum_{i \in \mathcal{W}_k} (\lambda + \psi^E) \max\left\{ -\Delta(S_k, \mathbf{x}_i), 0 \right\}}_{\text{Quit Opportunities and Poaching}}
\end{aligned}
$$

where:

- s^N and s^E are the shares of matches generated by employed and nonemployed workers
- ω is the (equilibrium) rate at which each firm employee generates matches for the firm
- χ^E and χ^N are the ergodic distributions for employed and nonemployed workers

Distribution Definitions Quits and Poaching

Worker Value

- Define the worker value $W_i(S_k)$ as NPV of wages of a worker i at firm k
- Value function is very messy to define but follows a similar structure HJB
- Accounts for same events, except:
  - The effect of contacts with poaching firms does not drop out
  - Handle wage negotiations when worker i receives a renegotiation shock, or meets a new firm
    Renegotiation Poaching
- As in Lise and Robin (2017), W is **not needed** to characterize ergodic distribution χ
  All of the real allocations fully characterized by V and χ
- Wages are purely allocative: only need $W_i(S_k)$ to back out the wages implied by the model

Equilibrium

A **stationary** equilibrium is:

1. a set of value functions $\{V, U\}$
2. distributions $\{\chi, \chi^N\}$, and
3. a firm match rate ω

such that

1. the values solve the HJB equations conditional on the distributions
2. the distributions are stationary and consistent with the decisions implied by the values, and
3. the market for matches clears:

$$\underbrace{\omega \int (1 + n(S_k))\mathrm{d}\chi(S_k)}_{\text{meetings received by firms}} = \underbrace{N_w \left[e\psi^E + (1-e)\psi^N\right]}_{\text{meetings generated by workers}}$$

Note: these distributions imply the shares of matches generated: $s^N = \frac{(1-e)\psi^N}{e\psi^E+(1-e)\psi^N}$ and $s^E = \frac{e\psi^E}{e\psi^E+(1-e)\psi^N}$

Distribution Definitions

Computation

Computational Algorithm

Since wages are **allocative**, we can proceed in two steps:

1. Solve for V and χ jointly:
   - Iterate training (updating) V and simulating to approximate χ until jointly converged
   - **Key observation:** We don't need wages at all for this step
   - **Challenge:** *very* high-dimensional *heterogeneously-sized* state space
     Number of states of a firm with n workers is proportional to n
2. Solving for W:
   - **Key observation:** HJB for W is more complicated than V, but we *already have* χ
   - After solving for W can back out wages along simulation path

Neural Networks are Function Approximators

- **Challenges:** Curse of dimensionality and heterogeneously sized state spaces
- **Solution:** approximate V and W with neural networks
- Neural Networks are highly parameterized function approximators with three key features:
  1. Universal approximation theorem (Hornik, Stinchcombe, and White 1989)
  2. Number of parameters required grows *linearly* with the number of states (exponentially for polynomials)
  3. Differentiable and easy to “train”

  Definition | Example | Training | Properties
- Highly effective at solving high dimensional dynamic programs (Maliar, Maliar, and Winant 2021, Azinovic, Gaegauf, and Scheidegger 2022)
- With appropriate architectures, can handle **set valued states** Permutation Invariance

Defining the Loss Function

- Assume a NN approximation parameterized by θ_V
- Need to define a loss function to "train" the neural network to minimize

$$\mathcal{L}_V(\theta_V) := \int R_V(S_k; \theta_V)^2 \mathrm{d}\Omega(S_k)$$

  - $R_V(S_k; \theta_V)$ is the residual of the joint value HJB evaluated at S_k
  - Ω is a distribution over states (in principle, any measure would do)
  - In practice, we want one that prioritizes accuracy in the states we care about

    A natural choice is χ, but want good approximation on states off equilibrium $\rightarrow$ synthetic distribution that augments χ with all states reachable within a single event from χ

- We train θ_V by **stochastic gradient descent** on batches sampled from Ω
  - Works well with Monte Carlo approximations of integrals in HJB

    We find accurate enough with 50-100 draws for each integral

  - Solves HJBs to reasonable degree of accuracy (L^2 errors $< 10^{-5}$) in 25 minutes on a GPU

    Can achieve higher accuracy with more computation time V Convergence W Convergence χ Convergence

Measurement

Data

- French matched employer-employee **administrative data**
- Constructed using mandatory form all businesses must submit every year (DADS)
- Two main datasets:
  1. Short panel: near-universe of workers, but overlapping structure (IDs reshuffled)
     - observe full universe of workers and coworkers
     - use this for descriptive evidence and main estimation targets
  2. Long panel: full employment history of people born in October
     - use this for flow rates and measuring nonemployment
- Key variables: wages, hours, establishment, occupation, demographics
- What we don't have: worker education

Defining a team

- Key decision: how do we define a team?
  - Too narrow $\rightarrow$ omitting relevant coworkers
  - Too broad $\rightarrow$ include coworkers you never interact with
- **Our approach:** teams are set of coworkers at the establishment within same 1-digit occupation
  - Want to be conservative in not excluding relevant interactions
  - Ex: 2-digit occupation would separately categorize "Lawyers" from "Legal Professionals"
  - Ex: 4-digit occupation would separately categorize "Medical Residents" from "Hospital Doctors without an Independent Practice"

Occ Codes Flow Rates

Parameterization ◂ Back to Model

- **Production:** Output produced according to a CES:

$$F(\underbrace{z_k, q_k, X_k}_{S_k}) := z_k \left(\sum_{i \in \mathcal{W}_k} h_i^{\eta} \right)^{\frac{1}{\eta}} \tag{1}$$

where

  - η controls the elasticity of substitution between workers
  - Can accommodate both supermodular and submodular production functions

- **Learning:** Extend Jarosch, Oberfield, and Rossi-Hansberg (2021):

$$\log\left(\frac{h_i'}{h_i}\right) = a_i + q_k + \underbrace{\frac{\theta^+}{n_k - 1} \sum_{j \in \mathcal{W}_i^+} \log\left(\frac{h_j}{h_i}\right)}_{\text{Effect of More Skilled Workers}} + \underbrace{\frac{\theta^-}{n_k - 1} \sum_{j \in \mathcal{W}_i^-} \log\left(\frac{h_j}{h_i}\right)}_{\text{Effect of Less Skilled Workers}} \tag{2}$$

where $\mathcal{W}_i^+ := \{j \in W_{k(i)} | h_j > h_i\}$ and $\mathcal{W}_i^- := \{j \in W_{k(i)} | h_j < h_i\}$

- **Initial Distributions:** jointly log normal Details

Calibration Strategy

- **External:**
  - Retirement rate, discounting set exogenously
  - Learning and renegotiation shocks set for expected waiting time of 1 year
  - Normalize non-separable means to zero

  Externally Set

- **Internally** calibrate remaining parameters to match:
  1. Variances, covariances of wage growth to match initial distributions
  2. Labor market flows to match arrival rates of shocks
  3. Within/between firm variance decomposition to match η

     See Herkenhoff, Lise, Menzio, and Phillips (2024)
  4. Auxiliary regression to target learning function parameters Auxiliary Regression

  We compute all growth rates and regressions in terms of percentile ranks, rather than actual wages to help with scale invariance across parameterizations

Calibration Results (Still rough and in progress!)

	Description	Value	Target	Data	Model
Short panel					
N_w	Workers per firm	5.371	Average employer size (unweighted)	4.660	5.190
b	Nonemployment flow value	0.141	p50 - p25 Wages	3.090	6.780
η	Production elasticity	0.939	Between-firm wage variance share (rank)	0.843	0.463
μ_q	Average Learning Environment	-0.016	Mean wage rank change	1.819	3.737
σ_z	Firm productivity variance	0.342	Correlation firm size vs. wage rank	0.038	0.164
σ_q	Firm learning environment variance	0.013	Variance of firm mean wage rank change	54.773	70.098
			Variance of $\alpha_{k(i)}$ in Equation 6	34.409	67.962
σ_{zq}	Firm learning-productivity covariance	0.013	Firm mean wage level-growth covariance	0.131	0.091
σ_h	Initial worker human capital variance	0.157	p75 - p50 Wages	6.165	3.262
σ_a	Worker learning ability variance	0.008	Wage rank change variance	73.354	174.555
			Variance of $\epsilon_{i,t}$ in Equation 6	7.286	9.428
σ_{ha}	Worker learning-initial productivity covariance	6.505e-04	Worker wage level-growth covariance	0.109	-0.005
θ^+	Learning from higher-ability coworkers	0.165	$\tilde{\theta}_1^+$ in Equation 6	0.340	0.384
			$\tilde{\theta}_2^+$ in Equation 6	0.001	-0.003
θ^-	Learning from lower-ability coworkers	0.034	$\tilde{\theta}_1^-$ in Equation 6	0.003	-0.002
			$\tilde{\theta}_2^-$ in Equation 6	0.000	-0.002
δ_f	Employer death rate	0.001	P50 employer size	1	5
ψ^E	Employed Contact Rate	1.048	P90 employer size	8	7
Long panel					
δ_m	Match break rate	0.127	EN rate	0.149	0.070
ψ^N	Nonemployed Contact Rate	0.314	NE Rate	0.311	0.353
			Inferred employment rate	0.733	0.823

Note: This table reports the internally-calibrated parameters and compares the relevant model-generated empirical targets with those in the data. Unconditional moments are computed before the sample is restricted to stayers.

Results

Drivers of Sorting

Sorting patterns depend on **production** and **learning** complementarities:

1. Complementarities in production b/w worker and firm productivities (h, z)

   $\rightarrow$ motive for **positive assortative matching**

2. Complementarities in production between workers *within* a firm

   - $\eta = 0.939 < 1$ so production function is supermodular

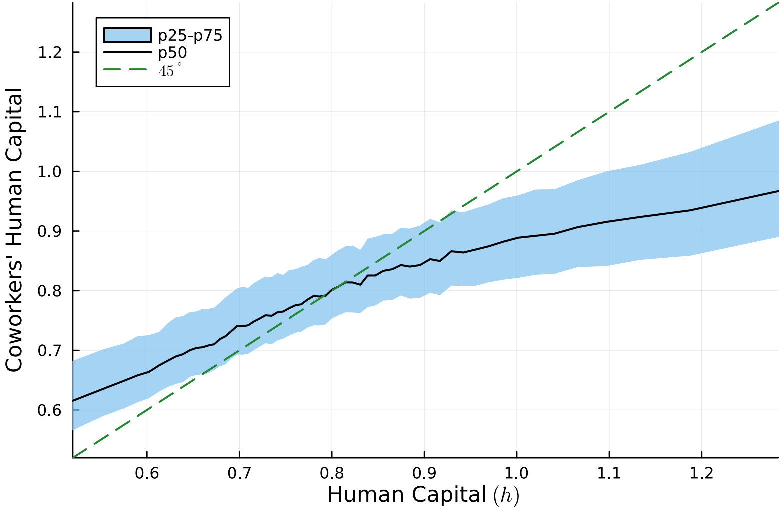
   $\rightarrow$ another motive for **positive assortative matching**

3. Complementarities in learning between workers

   - A worker *training* their coworkers is more valuable when gap to coworkers is larger

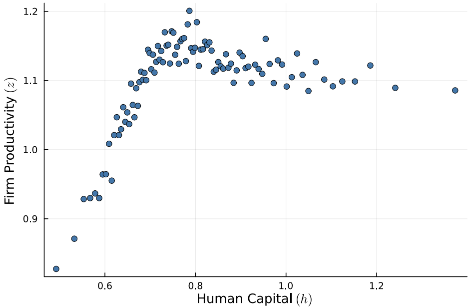
   $\rightarrow$ motive for **negative assortative matching**

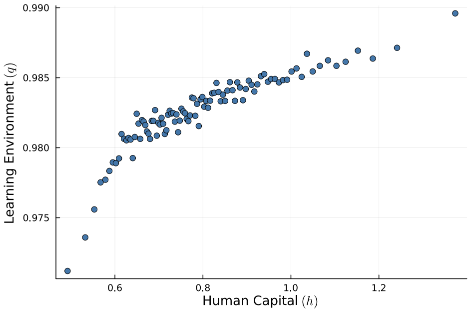
Sorting along coworkers: low-skill learn, high-skill teach



Distance to More and Less Skilled Coworkers

Sorting of Human Capital with Firm Characteristics

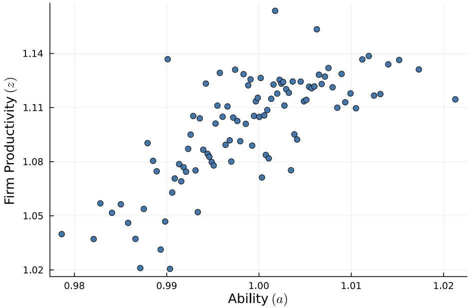


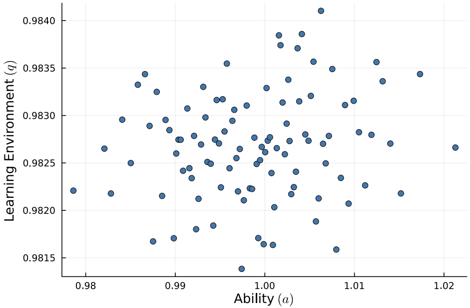


1. Sorting with firm productivity z mirrors coworker composition:
   - For low h, production complementarities induce **positive** assortative matching with z
   - For high h, incentive to train lower h coworkers outweighs the relative losses in production
     - $\rightarrow$ training motive dominates and we see **negative** assortative matching with z
2. Sorting with firm learning environment q is **positive**

Joint Distributions

Sorting of Learning Ability with Firm Characteristics

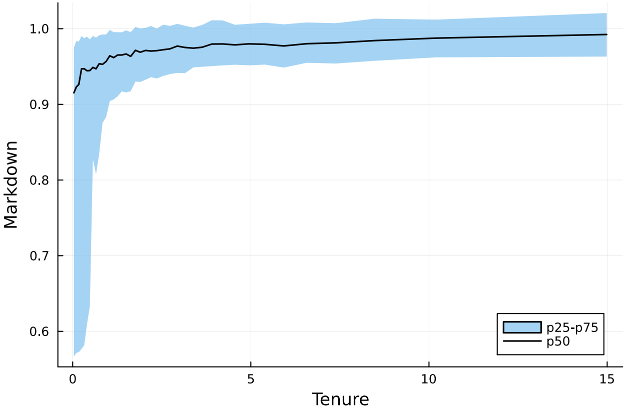




1. Sorting with firm productivity z is positive
2. No clear relationship with firm learning environment q

Joint Distributions

Workers get a higher share of MPL as their tenure increases



This is inherited from Postel–Vinay and Robin (2002)

Markdown Definition

→ as workers get wage offers, their share of the surplus goes up

Statistical Decomposition of Learning

Use structural model to decompose variance of human capital growth:

$$\mathrm{Var}\left(\log\left(\frac{h_i'}{h_i}\right)\right) = \underbrace{\mathrm{Var}(a_i)}_{\text{Learning Ability}} + \underbrace{\mathrm{Var}(q_k)}_{\text{Learning Environment}} + \underbrace{\left(\frac{\theta^+}{n_k - 1}\right)^2 \mathrm{Var}\left(\sum_{j\in\mathcal{W}_{i,k}^+} \log\left(\frac{h_j}{h_i}\right)\right)}_{\text{More Skilled Coworkers}}$$
$$+ \underbrace{\left(\frac{\theta^-}{n_k - 1}\right)^2 \mathrm{Var}\left(\sum_{j\in\mathcal{W}_{i,k}^-} \log\left(\frac{h_j}{h_i}\right)\right)}_{\text{Less Skilled Coworkers}} + \text{Covariance Terms} \tag{3}$$

Statistical Decomposition of Learning

	$\log a_i$	$\log q_i$	$\frac{\theta^+}{n_k-1}\sum_{j\in\mathcal{W}_i^+}\log\left(\frac{h_j}{h_i}\right)$	$\frac{\theta^-}{n_k-1}\sum_{j\in\mathcal{W}_i^-}\log\left(\frac{h_j}{h_i}\right)$
$\log a_i$	0.156	0.002	-0.112	-0.029
$\log q_i$		0.363	-0.011	0.002
$\frac{\theta^+}{n_k-1}\sum_{j\in\mathcal{W}_i^+}\log\left(\frac{h_j}{h_i}\right)$			0.525	0.072
$\frac{\theta^-}{n_k-1}\sum_{j\in\mathcal{W}_i^-}\log\left(\frac{h_j}{h_i}\right)$				0.033

- Most variation in human capital growth is learning from more skilled coworkers (52.5%)
- Learning ability (15.6%) and learning environment (36.3%) are also important
- Negative sorting between a and learning potential from more skilled coworkers (-11.2%)

Structural Decomposition of Learning

- Key parameters driving on-the-job learning are:
  - σ_a: std of worker learning ability
  - σ_q: std of firm learning environment
  - (θ^-, θ^+): learning function parameters
- To quantify the relative importance of each, we turn them off one at a time (and together)
- Resolve the model, and compute statistics about the distributions of h and w
- Normalize baseline to 1, so interpretable as percent deviation

Structural Decomposition of Learning: Individual Effects

		Mean h	Var h	Mean w	Var w
Individual	σ_q	1.050	0.646	1.093	1.615
	σ_a	1.009	0.700	0.956	1.112
	(θ^-, θ^+)	0.686	5.582	0.722	0.698

1. Shutting off learning leads to big decrease in mean h (31.4%) and mean w (27.8%)
   no complementarities in learning removes negative sorting of high h $\rightarrow$ smaller effect on w than on h
2. Mean w **decreases** without learning ability (9.3%), but **increases** without learning environment (4.4%)
   q is an additional dimension of heterogeneity that firms can exploit in setting wages $\rightarrow$ firms with higher q can pay lower w

Structural Decomposition of Learning: Cumulative Effects

		Mean h	Var h	Mean w	Var w
Cumulative	σ_a, σ_q	1.004	0.518	0.924	0.900
	$\sigma_a, \sigma_q, (\theta^-, \theta^+)$	0.858	2.564	0.861	0.904

1. Shutting off a and q *jointly* $\rightarrow$ modest 0.4% increase in mean h, but a larger 7.6% decrease in mean w

   This is because the learning ability channel dominates the learning environment channel

2. Shutting off all channels results in both lower h and w

   This is because the learning function is the dominant source of wage growth

Conclusion

- Developed novel model of large multi-worker firms, accommodating rich heterogeneity in firm and worker characteristics
- Introduced complementarities in production and learning across workers in the firm
- Show how to solve such a model using recent advances in deep learning
- Calibrated model to French administrative data
- In preliminary calibration, the bulk of the variation in human capital and wages across workers is driven by learning from more skilled coworkers

Thank you!

Back Matter

Proposition 1 (Separations)

When a worker j at firm p receives a poaching event with firm $k \neq p$, the increment to the joint value is $\max\{-\Delta(S_p, \mathbf{x}_j), 0\}$. *The change in the poaching firm's value net of their payment to the worker is*

$$\mathcal{B}(S_k, S_p, \mathbf{x}_j) = \max\left\{\Delta(S_k \oplus \mathbf{x}_j, \mathbf{x}_j) - \max\left\{\Delta(S_p, \mathbf{x}_j), 0\right\}, 0\right\}$$

Intuition:

- In standard case, where the surplus is positive at both firms, poacher k:
  - gets surplus $\Delta(S_k \oplus \mathbf{x}_j, \mathbf{x}_j)$ from hiring worker j
  - pays worker j the surplus $\Delta(S_p, \mathbf{x}_j)$ they would have gotten at firm p
- The max operators account for the fact that sometimes the surpluses are negative:
  - outside max operator checks if poaching is efficient
  - inside max operator checks if incumbent match should terminate

Back

Distribution Definitions

1. $\chi(S_k)$ is the distribution of firms across states
2. $\chi^N(\mathbf{x}_i)$ is the distribution of non-employed workers
3. $\chi^E(\mathbf{x}_j, S_{p(j)})$ is the distribution of workers across firms
4. $\Pi(S_p)$ is the size weighted distribution of firm states

$\chi^E(S_k, \mathbf{x}_i)$ is embedded within the distribution over firm states χ, since the worker states are included within the firm states

Back to Joint Value Back to Equilibrium

Quits and Poaching

1. When a **renegotiation** shock hits, either:
   - The match isn't terminated and any changes to w_i don't change V since it is a linear transfer between the firm and the worker
   - The surplus is negative and the worker quits to nonemployment
     - → The match gets refunded the surplus $-\Delta(S_k, \mathbf{x}_i)$
2. When a **poaching** event occurs, either:
   - Stay at incumbent firm and any change to w_i does not change V
   - Move to poaching firm
     - New firm pays worker their marginal product at old firm
     - Old firm loses that marginal product
     - → Cancels out and change to V is 0

Back

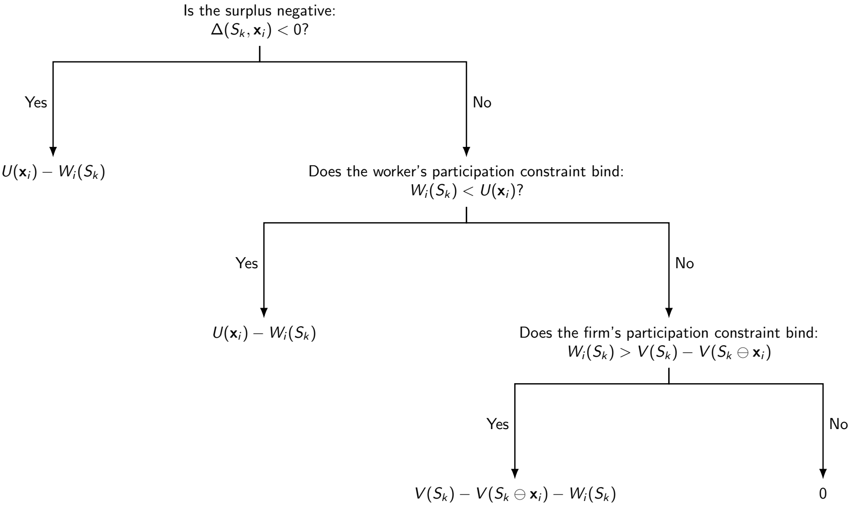
Worker Value

Back

$$
\begin{aligned}
\rho W_i(S_k) = w_i &+ \underbrace{\gamma^E\Big(W_i(H(S_k)) - W_i(S_k)\Big)}_{\text{Learning}} + \underbrace{\delta_f\Big(U(\mathbf{x}_i) - W_i(S_k)\Big)}_{\text{Firm Death}} + \underbrace{\sum_{j\neq i\in\mathcal{W}_k}(\delta_r+\delta_m)\Big(W_i(S_k\ominus\mathbf{x}_j) - W_i(S_k)\Big)}_{\text{Coworker Match Breaks and Retirement}} \\
&+ \underbrace{(n_k+1)\omega s^E\int\Big(\mathbb{1}\left\{\mathcal{B}(S_k, S_{p(j)}, \mathbf{x}_j) > 0\right\}\Big)\Big(W_i(S_k\oplus\mathbf{x}_j) - W_i(S_k)\Big)\mathrm{d}\chi^E(\mathbf{x}_j, S_{p(j)})}_{\text{Potential new co-worker from employment}} \\
&+ \underbrace{(n_k+1)\omega s^N\int\Big(\mathbb{1}\left\{\Delta(S_k\oplus\mathbf{x}_j, \mathbf{x}_j) > 0\right\}\Big)\Big(W_i(S_k\oplus\mathbf{x}_j) - W_i(S_k)\Big)\mathrm{d}\chi^N(\mathbf{x}_j)}_{\text{Potential new co-worker from non-employment}} \\
&+ \underbrace{\lambda\sum_{j\neq i\in\mathcal{W}_k}\Big(\mathbb{1}\left\{\Delta(S_k, \mathbf{x}_j) < 0\right\}\Big)\Big(W_i(S_k\ominus\mathbf{x}_j) - W_i(S_k)\Big)}_{\text{Coworker Quit Opportunities}} \\
&+ \underbrace{\psi^E\int\sum_{j\neq i\in\mathcal{W}_k}\mathbb{1}\left\{\mathcal{B}(S_p, S_k, \mathbf{x}_j) > 0\right\}\Big(W_i(S\ominus\mathbf{x}_j) - W_i(S_k)\Big)\mathrm{d}\Pi(S_p)}_{\text{Coworker Poacher Meetings}} \\
&+ \underbrace{\delta_m\Big(U(\mathbf{x}_i) - W_i(S_k)\Big) - \delta_r W_i(S_k)}_{\text{Own Match Breaks and Retirement}} + \underbrace{\lambda Q_i(S_k)}_{\text{Own Renegotiation Shocks}} + \underbrace{\psi^E\int P_i(S_k, S_p)\mathrm{d}\Pi(S_p)}_{\text{Own Poacher Meetings}}
\end{aligned}
$$

Renegotiation Logic

Back



Poaching Logic

Back

Is the surplus negative: $\Delta(S_k, \mathbf{x}_i) < 0$?

- Yes: $U(\mathbf{x}_i) - W_i(S_k)$
- No: Will the worker leave: $\Delta(S_k, \mathbf{x}_i) < \Delta(S_p \oplus \mathbf{x}_i, \mathbf{x}_i)$?
  - Yes: $\Delta(S_k, \mathbf{x}_i) + U(\mathbf{x}_i) - W_i(S_k)$
  - No: Does the participation constraint hold: $W_i(S_k) > \Delta(S_k, \mathbf{x}_i) + U(\mathbf{x}_i)$?
    - Yes: $\Delta(S_k, \mathbf{x}_i) + U(\mathbf{x}_i) - W_i(S_k)$
    - No: Is the poaching wage offer competitive: $W_i(S_k) < \max\{\Delta(S_p \oplus \mathbf{x}_i, \mathbf{x}_i), 0\} + U(\mathbf{x}_i)$?
      - Yes: $\max\{\Delta(S_p \oplus \mathbf{x}_i, \mathbf{x}_i), 0\} + U(\mathbf{x}_i) - W_i(S_k)$
      - No: 0

Poaching Value Change

We define the cases:

	Condition	Description
C_1	$\Delta(S_k, \mathbf{x}_i) < 0$	Surplus is negative
C_2	$\Delta(S_k, \mathbf{x}_i) < \Delta(S_p \oplus \mathbf{x}_i, \mathbf{x}_i)$	Worker leaves for p
C_3	$W_i(S_k) > \Delta(S_k, \mathbf{x}_i) + U(\mathbf{x}_i)$	Firm participation constraint
C_4	$W_i(S_k) < U(\mathbf{x}_i)$	Worker participation constraint
C_5	$W_i(S_k) < \max\{\Delta(S_p \oplus \mathbf{x}_i, \mathbf{x}_i), 0\} + U(\mathbf{x}_i)$	Poacher offer is competitive

Proposition 2 (Poaching)

When a worker i at firm k receives a poaching event from firm p, Then the change in the worker i's value upon receiving a poaching offer from p is given by:

$$P_i(S_k, S_p) = \begin{cases} U(\mathbf{x}_i) - W_i(S_k) & \text{if } C_1, \\ \Delta(S_k, \mathbf{x}_i) + U(\mathbf{x}_i) - W_i(S_k) & \text{if } \neg C_1 \text{ and } C_2, \\ \Delta(S_k, \mathbf{x}_i) + U(\mathbf{x}_i) - W_i(S_k) & \text{if } \neg C_1, \neg C_2, \text{ and } C_3, \\ \max\{\Delta(S_p \oplus \mathbf{x}_i, \mathbf{x}_i), 0\} + U(\mathbf{x}_i) - W_i(S_k) & \text{if } \neg C_1, \neg C_2, C_4, \text{ and } C_5, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Back

Neural Networks: Definition

- A neural network is a nonlinear function $f : \mathbb{R}^m \to \mathbb{R}^n$ that consists of interconnected nodes, or *neurons*, organized into *layers* (input, hidden, outer).
- Simplest version has no hidden layers: each output $k \in \{1, 2, \ldots, n\}$ is

$$y_k(x, w) = \sum_{i=1}^{m} w_{i,k}^0 x_i$$

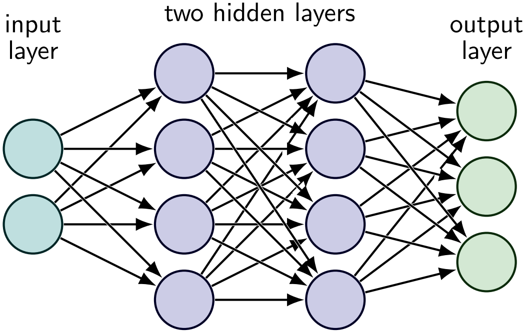
- Add a (hidden) layer with $p \in \mathbb{N}$ nodes and activation function h:

$$y_k(x, w) = \sum_{j=1}^{p} w_{j,k}^1 h\left(\sum_{i=1}^{m} w_{i,j}^0 x_i\right)$$

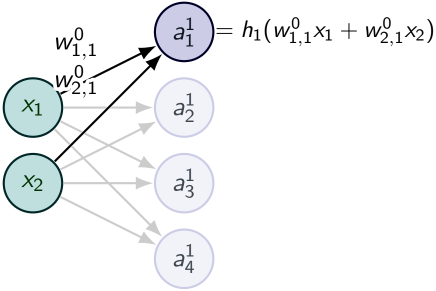
- Can add as many layers (depth) and nodes (width) as we want
- Choice of activation functions is crucial and can be used to enforce constraints

Back NN:example NN:training NN:properties

Neural Networks: Example



(a) Network diagram



(b) Neuron

$$y_k(x, w) = \sum_{j_2=1}^{4} w^2_{j_2,k} \overbrace{h_2\left(\sum_{j_1=1}^{4} w^1_{j_1,j_2} \underbrace{h_1\left(\sum_{i=1}^{2} w^0_{i,j_1} x_i\right)}_{a^1_{j_1}}\right)}^{a^2_{j_2}}, \quad k = 1, 2, 3$$

Back NN:definition NN:training NN:properties

Neural Networks: Training

- Neural network weights are updated by minimizing a *loss* function

$$w^* = \arg\min_w \mathcal{L}(x; w)$$

- A commonly-used loss function is the mean squared error (MSE)

$$\mathcal{L}^{MSE}(x; w) = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

- In practice, the weights are updated using *gradient descent*,

$$w_{new} = w + \eta \frac{\partial \mathcal{L}(x; w)}{\partial w}$$

- $\eta \in \mathbb{R}_+$ is the *learning rate*: not too small (flat spots), not too big (overshoot w^*)

Back NN:definition NN:example NN:properties

Neural Networks: Properties

1. Universal approximation theorem (Hornik, Stinchcombe, and White 1989)
2. Can represent highly complex functions: kinks and ridges, binding constraints, non-differentiabilities, discontinuities, and discrete choices
3. Bypass curse of dimensionality: number of weights to estimate scales linearly with dimension of input
   - 0 hidden layers: $m \times n$
   - 1 hidden layer: $m \times p + p \times n$
   - 2 hidden layers: $m \times p_1 + p_1 \times p_2 + p_2 \times n$

   Series (e.g. Chebyshev or Hermite) scale exponentially
4. Training is fast and easy due to recent advances in computing
5. Deep reinforcement learning: solve dynamic programs without direct optimization

Back NN:definition NN:example NN:training

Permutation Invariance

Proposition 3 (Kahou, Fernandez-Villaverde, Perla, and Sood 2022)

Let $f : \mathbb{R}^{N+1} \to \mathbb{R}$ be a continuous, permutation invariant function under S_N, i.e, for all $(x, X) \in \mathbb{R}^{N+1}$ and all $\pi \in S_N$:

$$f(x, \pi X) = f(x, X)$$

Then there exist $L \leq N$ and continuous functions $\rho : \mathbb{R}^{L+1} \to \mathbb{R}$ and $\phi : \mathbb{R} \to \mathbb{R}^L$ such that

$$f(x, X) = \rho\left(x, \frac{1}{N}\sum_{i=1}^{N} \phi(X_i)\right) \tag{5}$$

where X_i is the ith element of X.

Key Intuition: Permutation invariant functions can be represented as an average of a set of "moments" generated by an inner neural network ϕ

- Similar in spirit to Krusell and Smith (1998)
- Moment selection is automatic, and we have stronger theoretical guarantees Back

Occupation Codes in France

1			**Farmers**
2			**Craftsmen, Tradespeople, and Business Owners**
3			**Executives and High-Level Professionals**
	31		*Independent Professionals*
		311c	Dentists
		311d	Psychologists and Therapists
		311e	Veterinarians
		3121	Lawyers
	34		*Professors, Scientific Professionals*
		342b	Research Professors
		344a	Hospital Doctors Without an Independent Practice
		344c	Residents in Medicine, Dentistry and Pharmacy
		344d	Salaried Pharmacists
	37		*Corporate Administrative and Commercial Managers*
		372e	Legal Professionals
		375a	Advertising Executives
4			**Intermediate Professions**
5			**Clerical Workers**
6			**Manual Laborers**
9			**Non-Coded**

Back

Self-flow Rates

Table: Self-Flow Rates

	Rate (%)
OCC1	89.92
Firm	83.64
Establishment	79.16
Establishment $\times$ OCC1	74.11

Note: This table reports self-flow rates, the empirical probability that a worker stays at the same group from one year to the next. Calculated in the DADS-Postes from 2014 to 2015.

Back

Initial Distributions

- Workers draw their initial human capital h_i^0 and their permanent learning ability a_i from a joint log normal distribution $G_w(h_i^0, a_i)$:

$$\begin{pmatrix} \log h_i^0 \\ \log a_i \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_h \\ \mu_a \end{pmatrix}, \begin{pmatrix} \sigma_h^2 & \sigma_{ha}^2 \\ \sigma_{ha}^2 & \sigma_a^2 \end{pmatrix}\right]$$

- We also assume a joint log normal process $G_f(z_k, q_k)$:

$$\begin{pmatrix} \log z_k \\ \log q_k \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_z \\ \mu_q \end{pmatrix}, \begin{pmatrix} \sigma_z^2 & \sigma_{zq}^2 \\ \sigma_{zq}^2 & \sigma_q^2 \end{pmatrix}\right]$$

Back

Table: Externally-Calibrated Parameters

	Description	**Value**	**Explanation**
δ_r	Worker retirement rate	0.05	40 year career
δ_h	Human capital depreciation rate	1.0	Match data frequency
λ	Renegotiation shock arrival rate	1.0	Match data frequency
γ^E	Learning event arrival rate	1.0	Match data frequency
ρ	Annual discounting rate	0.05	Standard
μ_h	Mean log initial human capital	0.0	Normalization
μ_z	Mean log firm productivity	0.0	Normalization
μ_a	Mean log worker learning ability	0.0	Normalization

Note: This table reports the externally-calibrated parameters and their source.

Back

Learning Regression

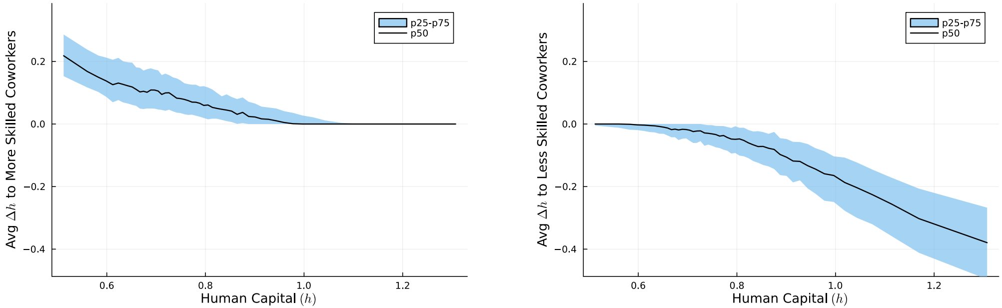
- We cannot directly observe human capital, but we do observe wages
- Run an auxiliary regression in short-panel meant to closely mirror the learning function (replace human capital with percentile ranks of wages)

$$
\begin{aligned}
w_{i,t} - w_{i,t-1} = \alpha_{k(i)} &+ \underbrace{\tilde{\theta}_1^{+} \sum_{j \in \mathbb{W}^{+}_{i,t}} \frac{w_{j,t-1} - w_{i,t-1}}{n_{k(i)} - 1}}_{\text{Higher-Wage Coworkers}} + \underbrace{\tilde{\theta}_1^{-} \sum_{j \in \mathbb{W}^{-}_{i,t}} \frac{w_{j,t-1} - w_{i,t-1}}{n_{k(i)} - 1}}_{\text{Lower-Wage Coworkers}} \\
&+ \underbrace{\tilde{\theta}_2^{+} \sum_{j \in \mathbb{W}^{+}_{i,t}} \frac{(w_{j,t-1} - w_{i,t-1})^2}{n_{k(i)} - 1} + \tilde{\theta}_2^{-} \sum_{j \in \mathbb{W}^{-}_{i,t}} \frac{(w_{j,t-1} - w_{i,t-1})^2}{n_{k(i)} - 1}}_{\text{Nonlinear Effects}} + \epsilon_{i,t}
\end{aligned}
\tag{6}
$$

- Regression coefficient help target model analogues
- Variance of fixed effects targets σ_q
- RMSE targets σ_a

Back

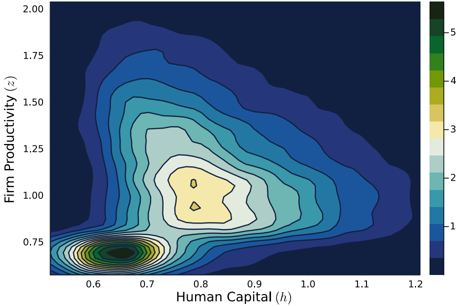
Opportunities for Learning

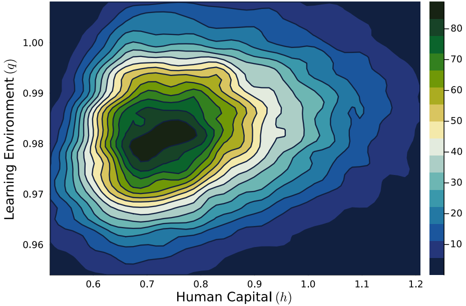


- Low h workers are closer to their coworkers than high h workers
- Few learning opportunities for high h workers as they are much more skilled than their coworkers

Back

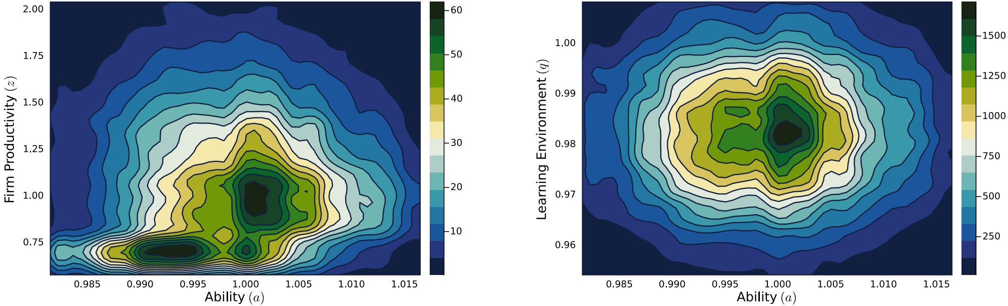
Joint Distributions





◂ Back

Joint Distributions



Back

Markdown Definition

The **dynamic** marginal product of a worker $\mathbf{x}_i$ is the change in the joint value if the worker is removed:

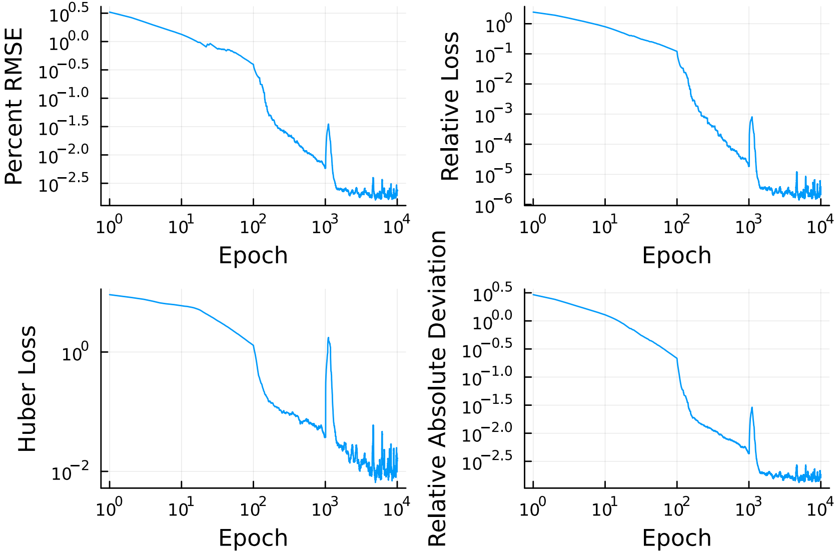
$$J_i(S_k) := V(S_k) - V(S_k \ominus \mathbf{x}_i)$$

The markdown is the ratio of the worker's value to the marginal product:

$$W_i(S_k)/J_i(S_k)$$

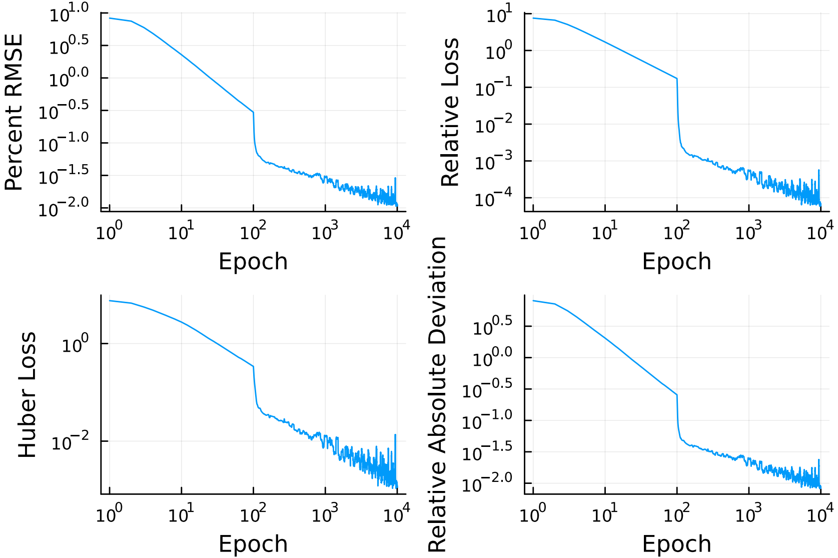
Back

Convergence V



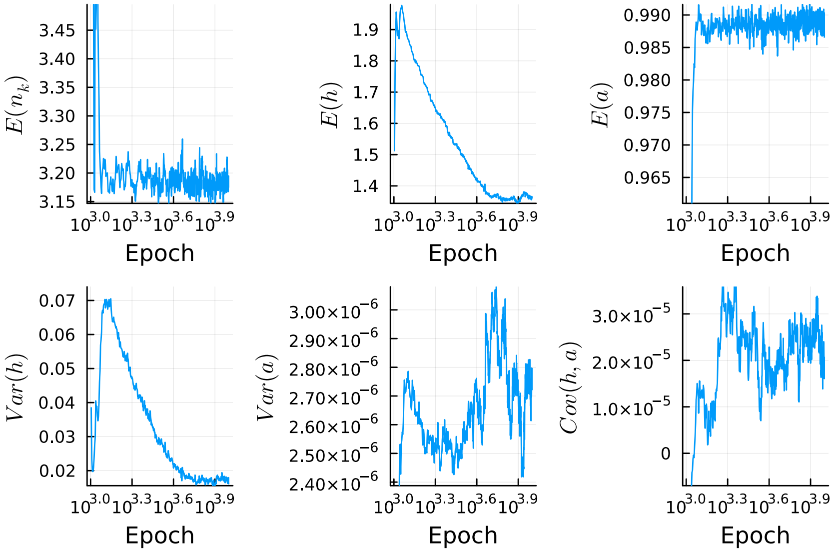
Back

Convergence W



Back

Convergence χ



Back

References

Azinovic, M., L. Gaegauf, and S. Scheidegger (2022): "Deep Equilibrium Nets," *International Economic Review*, 63, 1471–1525.

Burdett, K. and T. Vishwanath (1988): "Balanced matching and labor market equilibrium," *Journal of Political Economy*, 96, 1048–1065.

Duarte, V. (2022): "Simple Allocation Rules and Optimal Portfolio Choice Over the Lifecycle," .

Duarte, V., D. Duarte, and D. H. Silva (2023): "Machine Learning for Continuous-Time Finance," Working Paper.

Freund, L. (2024): "Superstar Teams," Working Paper.

Hall, R. E. (2005): "Employment Fluctuations with Equilibrium Wage Stickiness," *American Economic Review*, 95, 50–65.

Herkenhoff, K., J. Lise, G. Menzio, and G. M. Phillips (2024): "Production and learning in teams," *Econometrica*, 92, 467–504.

References (cont.)

Hornik, K., M. Stinchcombe, and H. White (1989): "Multilayer feedforward networks are universal approximators," *Neural networks*, 2, 359–366.

Jarosch, G., E. Oberfield, and E. Rossi-Hansberg (2021): "Learning From Coworkers," *Econometrica*, 89, 647–676.

Jungerman, W. (2023): "Dynamic Monopsony and Human Capital," Working Paper.

Kahou, M. E., J. Fernandez-Villaverde, J. Perla, and A. Sood (2022): "Exploiting Symmetry in High-Dimensional Dynamic Programming," Working Paper.

Krusell, P. and A. A. Smith, Jr (1998): "Income and wealth heterogeneity in the macroeconomy," *Journal of Political Economy*, 106, 867–896.

Lise, J. and J.-M. Robin (2017): "The Macrodynamics of Sorting between Workers and Firms," *American Economic Review*, 107, 1104–35.

Ma, X., A. Nakab, and D. Vidart (2024): "How do Workers Learn? Theory and Evidence on the Roots of Lifecycle Human Capital Accumulation," Working Paper.

Maliar, L., S. Maliar, and P. Winant (2021): "Deep Learning for Solving Dynamic Economic Models." *Journal of Monetary Economics*, 122, 76–101.

References (cont.)

Postel–Vinay, F. and J. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70, 2295–2350.

Thomas, J. and T. Worrall (1988): "Self-Enforcing Wage Contracts," *The Review of Economic Studies*, 55, 541–554.