Deep Learning for Search And Matching Models

(a.k.a. "DeepSAM")

Jonathan Payne — Princeton

Adam Rebei — Stanford

Yucheng Yang — Zurich and SFI

April 29, 2024

EIEF Seminar

Introduction

- Heterogeneity and aggregate shocks are important in markets with search frictions (e.g. labor and financial markets.)
- Most search and matching (SAM) models with heterogeneous agents study:
  1. Deterministic steady state (e.g. Shimer-Smith '00),
  2. Aggregate fluctuations, but make assumptions to eliminate distribution from state space (e.g. "block recursivity" in Menzio-Shi '11, Lise-Robin '17).
- We present SAM models as high-dim. PDEs with distribution & agg. shocks as states

  ...and develop a new deep learning method, DeepSAM, to solve them globally.
- We use our method to study labor market dynamics across the business cycle. (We also study OTC markets but not for today.)

Deep Learning for Economic Models

- Deep learning has been successful in high dimensional scientific computing problems.

- We can use deep learning to solve high-dimensional value & policy functions in economics:

  1. Use deep neural networks to approximate value function $V : \mathbb{R}^N \to \mathbb{R}$

$$V(\mathbf{x}) \approx \mathcal{L}^P \circ \cdots \circ \mathcal{L}^p \circ \cdots \circ \mathcal{L}^1(\mathbf{x}), \quad \mathbf{x}\text{: high-dim state vector,}$$
$$\mathbf{h}_p = \mathcal{L}^p(\mathbf{h}_{p-1}) = \sigma\big(\mathbf{W}_p \mathbf{h}_{p-1} + \mathbf{b}_p\big), \quad \mathbf{h}_0 = \mathbf{x},$$

σ : element-wise nonlinear fn, e.g. $\mathrm{Tanh}(\cdot)$. Want to solve unknown parameters $\boldsymbol{\Theta} = \{\mathbf{W}_p, \mathbf{b}_p\}_p$.

  2. Cast high-dim function into a loss function, e.g. Bellman equation residual.

  3. Optimize unknown parameters, $\boldsymbol{\Theta}$, to minimize average loss on a "global" state space, using stochastic gradient descent (SGD) method.

- Similar procedure to polynomial "projection", but more efficient in practice.

This Paper

- Apply DeepSAM to canonical models with aggregate shocks:
  1. Shimer-Smith/Mortensen-Pissarides model with two-sided heterogeneity (today's focus).
  2. Shimer-Smith/MP model with two-sided heterogeneity and on-the-job search (at end).
  3. Application to search models in the OTC market.
- High accuracy in "global" state space (including distribution); efficient compute time.
- Economic experiments and findings:
  1. The impact of distribution on aggregate dynamics is large when aggregate shocks affect agents in an uneven way, but is small under symmetric shocks.
  2. Countercyclical sorting over business cycles.
  3. Magnitude of countercyclicality depends on the worker-firm bargaining power.

Literature

- Deep learning in macro; for incomplete market heterogeneous agent models (HAM) (e.g. Han-Yang-E '21 "DeepHAM", Gu-Laurière-Merkel-Payne '23, among many others)
  - This paper: search and matching (SAM) models.

	Distribution	Distribution impact on decisions
HAM	Asset wealth and income	Via aggregate prices
SAM	Type (productivity) of agents in two sides of matching	Via matching probability with other types

- Continuous time formulation of macro models with heterogeneity (e.g. Ahn et al. '18, Schaab '20, Achdou et al. '22, Alvarez et al. '23, Bilal '23.)
  - This paper: global solution with aggregate shocks.

- Search model with business cycle (e.g. Shimer '05, Menzio-Shi '11, Lise-Robin '17.)
  - This paper: keep distribution in the state vector.

Table of Contents

Shimer-Smith/Mortensen-Pissarides with Two-sided Heterogeneity

- Continuous time, infinite horizon environment.
- Workers $x \in [0,1]$ with exog density $g_t^w(x)$; Firms $y \in [0,1]$ with free entry:
  - Unmatched: unemployed workers get benefit b; vacant firms produce nothing.
  - Matched: type x worker and type y firm produce output $z_t f(x,y)$.
  - z_t: follows two-state continuous time Markov Chain (can be generalized).
- Meet randomly at rate $m(\mathcal{U}_t, \mathcal{V}_t)$, $\mathcal{U}_t$ is total unemployment, $\mathcal{V}_t$ is total vacancies.
- Upon meeting, agents choose whether to accept the match:
  - Match surplus $S_t(x,y)$ divided by generalized Nash bargaining: worker get fraction β.
  - Match acceptance decision $\alpha_t(x,y) = \mathbb{1}\{S_t(x,y) > 0\}$. Match dissolve rate $\delta(x,y,z)$.
- Equilibrium object: $g_t(x,y)$ mass function of matches (x,y).

Free entry condition

Recursive Equilibrium Part I: Unemployed Workers & KFE

- Idiosyncratic state $= x$, Aggregate states $= (z, g(x,y))$.

- Hamilton-Jacobi-Bellman equation for an unemployed worker's value $V^u(x,z,g)$:

$$\rho V^u(x,z,g) = b + \frac{m(z,g)}{\mathcal{U}(z,g)} \int \alpha(x,\tilde{y},z,g)(V^e(x,\tilde{y},z,g) - V^u(x,z,g)) \frac{g^v(\tilde{y})}{\mathcal{V}(z,g)} d\tilde{y}$$
$$+ \lambda_{z\tilde{z}}(V^u(x,\tilde{z},g) - V^u(x,z,g)) + D_g V^u(x,z,g) \cdot \mu^g$$

  - $\alpha_t(x,\tilde{y},z,g)$: match acceptance decision.
  - $g_t^u(x)$ and $g_t^v(y)$: mass functions for unemployed workers and vacant firms.
  - $V^e(x,y,z,g)$ is employed worker's value and $D_g V^u(x,z,g)$ is Frechet derivative w.r.t g.

- Kolmogorov forward equation (KFE):

$$\frac{dg_t(x,y)}{dt} := \mu^g(x,y,z,g) = -\delta(x,y,z)g(x,y) + \frac{m(z,g)}{\mathcal{U}(z,g)\mathcal{V}(z,g)} \alpha(x,y,z,g) g^v(y) g^u(x)$$

Recursive Equilibrium Part II: Other Equations

- Hamilton-Jacobi-Bellman equation (HJBE) for employed worker's value $V^e(x, y, z, g)$:

$$\begin{aligned}\rho V^e(x, y, z, g) = w(x, y, z, g) + \delta(x, y, z)\left(V^u(x, z, g) - V^e(x, y, z, g)\right) \\ + \lambda_{z\tilde{z}}(V^e(x, y, \tilde{z}, g) - V^e(x, y, z, g)) + D_g V^e(x, y, z, g) \cdot \mu^g\end{aligned}$$

- HJBE for a vacant firm's value $V^v(y, z, g)$:

$$\begin{aligned}\rho V^v(y, z, g) = \frac{m(z, g)}{\mathcal{V}(z, g)} \int \alpha(\tilde{x}, y, z, g)(V^p(\tilde{x}, y, z, g) - V^v(y, z, g)) \frac{g^u(\tilde{x})}{\mathcal{U}(z, g)} d\tilde{x} \\ + \lambda_{z\tilde{z}}(V^v(x, \tilde{z}, g) - V^v(x, z, g)) + D_g V^v(y, z, g) \cdot \mu^g\end{aligned}$$

- HJBE for a producing firm's value $V^p(x, y, g)$:

$$\begin{aligned}\rho V^p(x, y, z, g) = zf(x, y) - w(x, \tilde{y}, z, g) + \delta(x, y, z)(V^v(y, z, g) - V^p(x, y, z, g)) \\ + \lambda_{z\tilde{z}}(V^p(x, y, \tilde{z}, g) - V^p(x, y, z, g)) + D_g V^p(x, y, z, g) \cdot \mu^g\end{aligned}$$

Recursive Characterization For Equilibrium Surplus

- Surplus from match $S(x,y,z,g) := V^p(x,y,z,g) - V^v(y,z,g) + V^e(x,y) - V^u(x,z,g)$.
- Characterize equilibrium with master equation for surplus: Free entry condition

$$\begin{aligned}
\rho S(x,y,z,g) = {} & zf(x,y) - \delta(x,y,z)S(x,y,z,g) \\
& - (1-\beta)\frac{m(z,g)}{\mathcal{V}(z,g)}\int \alpha(\tilde{x},y,z,g)S(\tilde{x},y,z,g)\frac{g^u(\tilde{x})}{\mathcal{U}(z,g)}d\tilde{x} \\
& - b - \beta\frac{m(z,g)}{\mathcal{U}(z,g)}\int \alpha(x,\tilde{y},z,g)S(x,\tilde{y},z,g)\frac{g^v(\tilde{y})}{\mathcal{V}(z,g)}d\tilde{y} \\
& + \lambda(z)(S(x,y,\tilde{z},g) - S(x,y,z,g)) + D_g S(x,y,z,g)\cdot\mu^g(z,g)
\end{aligned}$$

- Kolmogorov forward equation (KFE):

$$\frac{dg_t(x,y)}{dt} := \mu^g(x,y,z,g) = -\delta(x,y,z)g(x,y) + \frac{m(z,g)}{\mathcal{U}(z,g)\mathcal{V}(z,g)}\alpha(x,y,z,g)g^v(y)g^u(x)$$

- High-dim PDEs with distribution in state: hard to solve with conventional methods.

Comparison to Other Heterogeneous Agent Search Models

- Lise-Robin '17: sets $\beta = 0$ and introduces "free" vacancy creation so that:
  (and Postal-Vinay style Bertrand competition for workers searching on-the-job)

$$\alpha(x, y, z, g) = \alpha(x, y, z), \quad S(x, y, z, g) = S(x, y, z)$$

- Menzio-Shi '11: one-sided heterogeneity, competitive search, and "free" firm entry so:

$$S(x, y, z, g) = S(x, y, z)$$

- We look for a solution for S and α in terms of the distribution g.

Modification 1: Finite Type Approximation

- Approximate $g(x, y)$ on finite types: $x \in \mathcal{X} = \{x_1, \dots, x_{n_x}\}$, $y \in \mathcal{Y} = \{y_1, \dots, y_{n_y}\}$.
- Finite state approximation $\Rightarrow$ analytical (approximate) KFE: $g \approx \{g_{ij}\}_{i \le n_x, j \le n_y}$
- Approximated master equation for surplus:

$$
\begin{aligned}
0 = \mathcal{L}^S S(x, y, z, g) = &-(\rho + \delta) S(x, y, z, g) + z f(x, y) - b \\
&- (1 - \beta) \frac{m(z, g)}{\mathcal{V}(z, g)} \frac{1}{n_x} \sum_{i=1}^{n_x} \alpha(\tilde{x}_i, y, z, g) S(\tilde{x}_i, y, z, g) \frac{g^u(\tilde{x}_i)}{\mathcal{U}(z, g)} \\
&- \beta \frac{m(z, g)}{\mathcal{U}(z, g)} \frac{1}{n_y} \sum_{j=1}^{n_y} \alpha(x, \tilde{y}_j, z, g) S(x, \tilde{y}_j, z, g) \frac{g^v(\tilde{y}_j)}{\mathcal{V}(z, g)} \\
&+ \lambda(z)(S(x, y, \tilde{z}, g) - S(x, y, z, g)) + \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \partial_{g_{ij}} S(x, y, z, \{g_{ij}\}_{i,j}) \mu^g(\tilde{x}_i, \tilde{y}_j, z, g)
\end{aligned}
$$

Modification 2: Approximate Discrete Choice

- In the original model,

$$\alpha(x, y, z, g) = \mathbb{1}\{S(x, y, z, g) > 0\}$$

- Discrete choice $\alpha \Rightarrow$ discontinuity of $S(x, y, z, g)$ at some g.

- To ensure master equation well defined & NN algorithm works, we approximate with

$$\alpha(x, y, z, g) = \frac{1}{1 + e^{-\xi S(x,y,z,g)}}$$

- Interpretation: logit choice model with utility shocks $\sim$ extreme value distribution.
  ($\xi \to \infty \Rightarrow$ discrete choice α.)

DeepSAM algorithm

- Approximate surplus by neural network $S(x, y, z, g) \approx \widehat{S}(x, y, z, g; \boldsymbol{\Theta})$.
- Start with initial parameter guess $\boldsymbol{\Theta}^0$. At iteration n with $\boldsymbol{\Theta}^n$:
  1. Generate K sample points, $Q^n = \left\{\left(x_k, y_k, z_k, \{g_{ij,k}\}_{i \le n_x, j \le n_y}\right)\right\}_{k \le K}$.
  2. Calculate the average mean squared error of surplus master equation on sample points:
  $$L\left(\boldsymbol{\Theta}^n, Q^n\right) := \frac{1}{K} \sum_{k \le K} \left| \mathcal{L}^S \widehat{S}\left(x_k, y_k, z_k, \{g_{ij,k}\}_{i \le n_x, j \le n_y}\right) \right|^2$$
  3. Update NN parameters with stochastic gradient descent (SGD) method:
  $$\boldsymbol{\Theta}^{n+1} = \boldsymbol{\Theta}^n - \zeta^n \nabla_{\boldsymbol{\Theta}} L\left(\boldsymbol{\Theta}^n, Q^n\right)$$
  4. Repeat until $L\left(\boldsymbol{\Theta}^n, Q^n\right) \le \epsilon$ with precision threshold ϵ.
- Once S is solved, we have α and can solve for worker and firm value functions.

The Deep Learning Conundrum

- Algorithm seems straightforward to describe
- ... but hard to implement successfully!
- I will discuss some features that we have found helpful.

Methodology Q & A

- ***Q.*** *How do we choose where to sample?*
  - We start by drawing distributions "between" steady states for different fixed z.
  - Can move to ergodic sampling once error is small.
  - Can increase sampling in regions of the state space where errors are high.
- ***Q.*** *How stable can we make the algorithm?*
  - Most difficult when $\widehat{S}(x, y, z, g; \boldsymbol{\Theta})$ has sharp curvature. We use "homotopy":
    - Step 1: train NN for parameters that give low curvature in $\widehat{S}^1$
    - Step 2: change parameters closer and retrain NN starting from previous $\widehat{S}^2 = \widehat{S}^1$
    - Step 3+: keep changing parameters and retraining until at desired parameters.
- ***Q.*** *How do we calculate derivatives when we evaluate the master equation error?*
  - We take automatic derivatives of neural network $\widehat{S}(x, y, z, g; \boldsymbol{\Theta})$; w.r.t. inputs (x, y, z, g).

Methodology Q & A

- ***Q.*** *What do we mean when we say this is a global solution?*
  - DeepSAM gives a solution across the discretized state space $(x, y, z, \{g_{ij}\}_{i \le n_x, j \le n_y})$,
  - …which is a $3 + n_x \times n_y$ dimensional state space.
  - Not a perturbation in z or g (e.g. Bilal '23).

- ***Q.*** *Could we use an alternative parametric approx. (e.g. Chebyshev polynomials)?*
  - Chebyshev projections require specially chosen grids to avoid oscillation problems
  - Automatic derivatives can easily calculated for neural networks.
  - Effective non-linear optimizers have been developed for neural nets.

- ***Q.*** *What about dimension reduction?*
  - For competitive markets, Krusell-Smith '98 suggest approximating distribution by mean.
  - For random search, not clear what moment enables approximation of:

$$\int \alpha(\tilde{x}, y, z, g) S(\tilde{x}, y, z, g) \frac{g^u(\tilde{x})}{\mathcal{U}(z, g)} d\tilde{x}, \quad \text{and} \quad \int \alpha(x, \tilde{y}, z, g) S(x, \tilde{y}, z, g) \frac{g^v(\tilde{y})}{\mathcal{V}(z, g)} d\tilde{y}$$

Table of Contents

Calibration

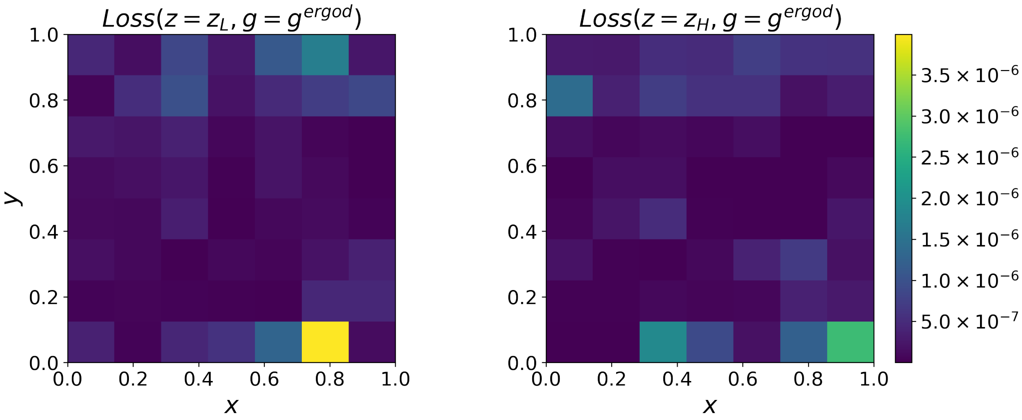
Frequency: annual.

Parameter	Interpretation	Value	Target/Source
ρ	Discount rate	0.05	Kaplan, Moll, Violante '18
δ	Job destruction rate	0.2	BLS job tenure 5 years
ξ	Extreme value distribution for α choice	2.0	
$f(x, y)$	Production function for match (x, y)	$0.6 + 0.4\left(\sqrt{x} + \sqrt{y}\right)^2$	Hagedorn et al '17
β	Surplus division factor	0.72	Shimer '05
$z, \tilde{z}$	TFP shocks	1 ± 0.015	Lise Robin '17
$\lambda_z, \lambda_{\tilde{z}}$	Poisson transition probability	0.08	Shimer '05
$\delta, \tilde{\delta}$	Separation shocks	0.2 ± 0.02	Shimer '05
$\lambda_\delta, \lambda_{\tilde{\delta}}$	Poisson transition probability	0.08	Shimer '05
$m(\mathcal{U}, \mathcal{V})$	Matching function	$\kappa \mathcal{U}^\nu \mathcal{V}^{1-\nu}$	Lise Robin '17
ν	Elasticity parameter for meeting function	0.5	Lise-Robin '17
κ	Scale parameter for meeting function	5.4	Unemployment rate
b	Worker unemployment benefit	0.5	Shimer '05
n_x	Discretization of worker types	7	
n_y	Discretization of firm types	8	

Numerical performance: Accuracy I

Calibration

- Mean squared loss as a function of type in the master equations of S (at ergodic g).



Numerical performance: Accuracy II Calibration

- Compare steady state solution without aggregate shocks to solution using conventional methods.

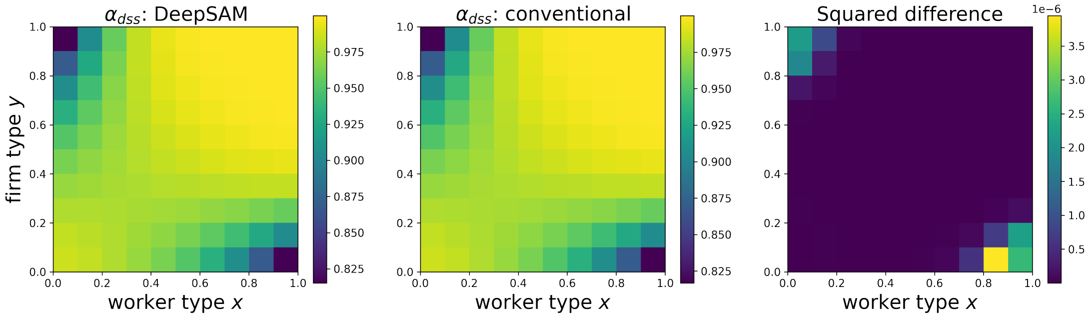


Figure: Comparison with steady-state solution

Comparison for discrete α

Numerical performance: Speed

- Solving the 58-dimensional surplus function takes 57 minutes on an A100 GPU.
- To our knowledge, it's infeasible to use any conventional methods to solve the problem globally with 58 dimensions.

Table of Contents

IRF to negative TFP shock: full vs block recursive model

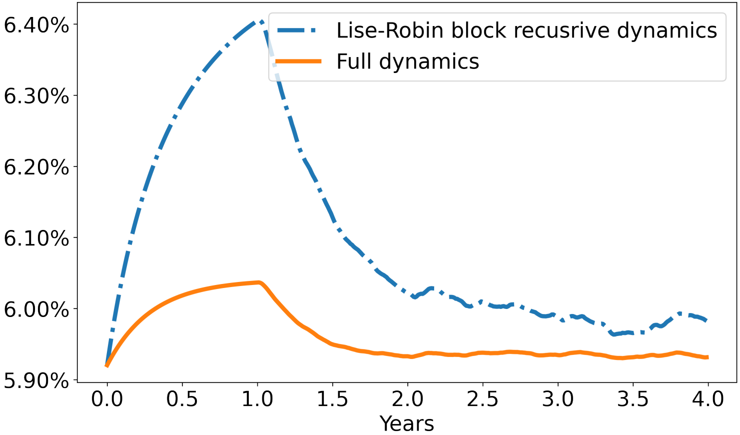


Figure: Our model solved with DeepSAM vs. block-recursive model à la Lise-Robin

Note: we recalibrate the model to match the unemployment rate at steady state when we adopt the Lise-Robin assumption with $\beta = 0$.

Distribution feedback to aggregates: IRF to labor market crisis

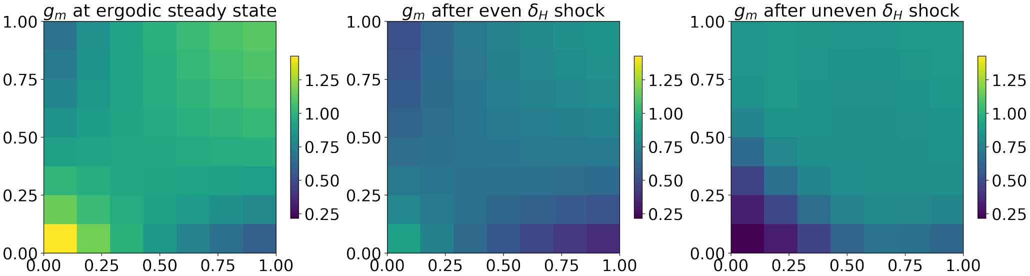
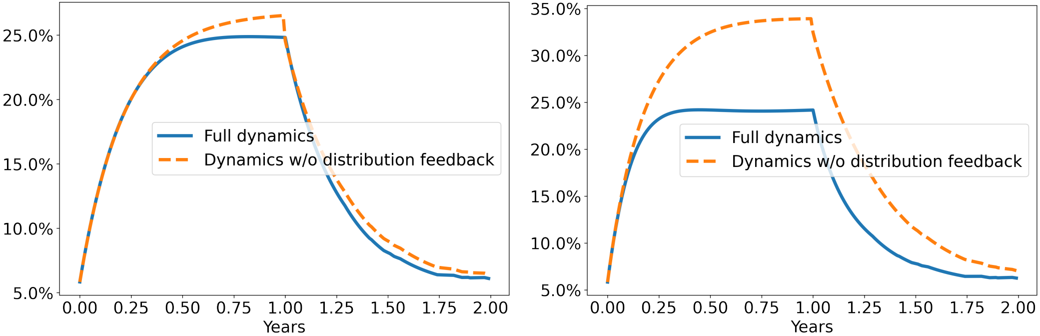


Figure: Ergodic distribution and distribution after the "uneven" and "even" depression

- Depression (25% U_t) due to persistent separation shocks. Uneven shock: more separation for poor worker-poor firm matches.
- **Question**: how recovery dynamics differ under full solution vs under solution with no feedback from distribution g to agents' decision?

Unemployment dynamics after "depression" shocks on g



Left: "even" shock; right: "uneven" shock.

Table of Contents

Environment Features

- Same worker types, firm types, and production function.
- Now all workers search; meeting rate is $m(\mathcal{W}_t, \mathcal{V}_t)$; total search effort is $\mathcal{W}_t := \mathcal{U}_t + \phi\mathcal{E}_t$
- Terms of trade when a vacant $\tilde{y}$-firm meets:
  - Unemployed x-worker: Nash bargaining where workers get surplus fraction β,
  - Worker in (x, y) match: Bertrand competition b/n firms. (Postel-Vinay-Robin '02)

    (If $S_t(x, \tilde{y}) > S_t(x, y)$, then worker moves to firm $\tilde{y}$ and gets $(1-\beta)S_t(x, y)$)
- If worker doesn't move, then continuation value updates to new match surplus. (Contract indexing, as in Lise-Robin.)

Recursive Characterization For Equilibrium Surplus

- Can characterize equilibrium with the master equation for the surplus:

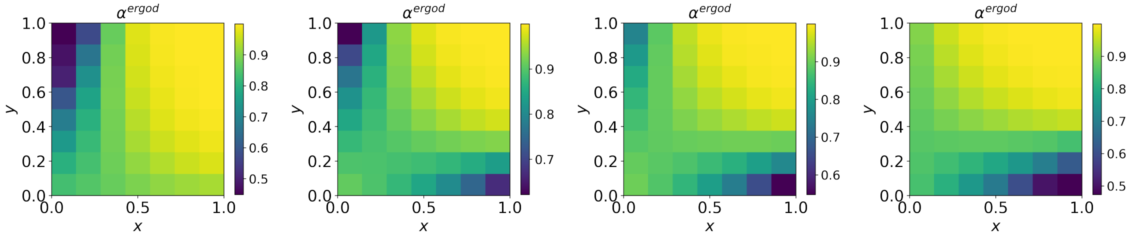
$$
\begin{aligned}
\rho S(x,y,z,g) &= zf(x,y) - \delta S(x,y,z,g) \\
&- (1-\beta)\frac{m(z,g)}{\mathcal{W}(z,g)\mathcal{V}(z,g)}\Bigg[\int \alpha(\tilde{x},y,z,g)S(\tilde{x},y,z,g)g^u(\tilde{x})d\tilde{x} \\
&-\phi\int \alpha^p(\tilde{x},y,\tilde{y},z,g)(S(\tilde{x},y,z,g)-S(\tilde{x},\tilde{y},z,g))g(\tilde{x},\tilde{y})d\tilde{x}d\tilde{y} \\
&+\phi\int \alpha^p(x,\tilde{y},y,z,g)S(x,y,z,g)g^v(\tilde{y})d\tilde{y}\Bigg] \\
&- b - \beta\frac{m(z,g)}{\mathcal{W}(z,g)\mathcal{V}(z,g)}\int \alpha(x,\tilde{y},z,g)S(x,\tilde{y},z,g)g^v(\tilde{y})d\tilde{y} \\
&+ \lambda(z)(S(x,y,\tilde{z},g)-S(x,y,z,g)) + D_gS(x,y,z,g)\cdot\mu^g(z,g)
\end{aligned}
$$

where:

$$\alpha^p(\tilde{x},y,\tilde{y},z,g) := \mathbb{1}\{S(\tilde{x},y,z,g) \geq S_t(\tilde{x},\tilde{y},z,g) \geq 0\}$$

KFE

Worker Bargaining Power Influences Assortative Matching

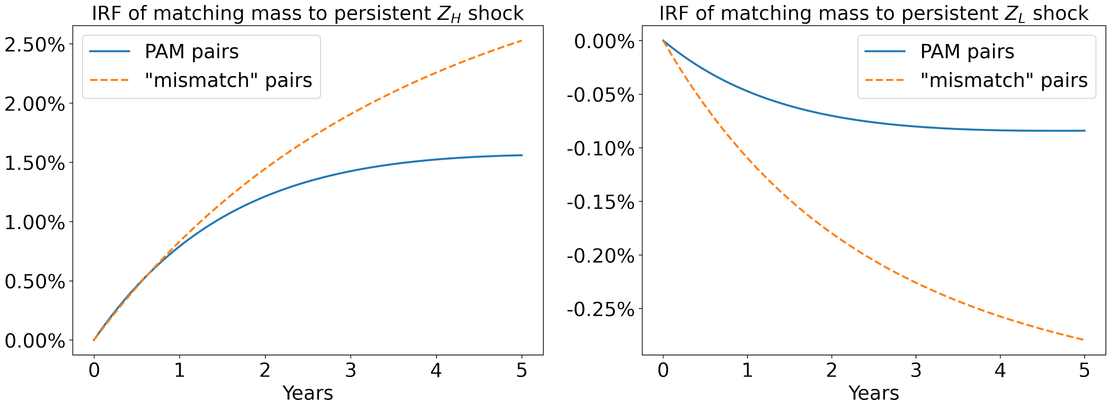


Sorting at the ergodic distribution for different worker bargaining power β. Left to right $\beta = 0$ (Lise-Robin '17), $0.5, 0.72$ (benchmark), 1.

Additional parameter calibration: $\phi = 0.2$.

Sorting Over Business Cycles

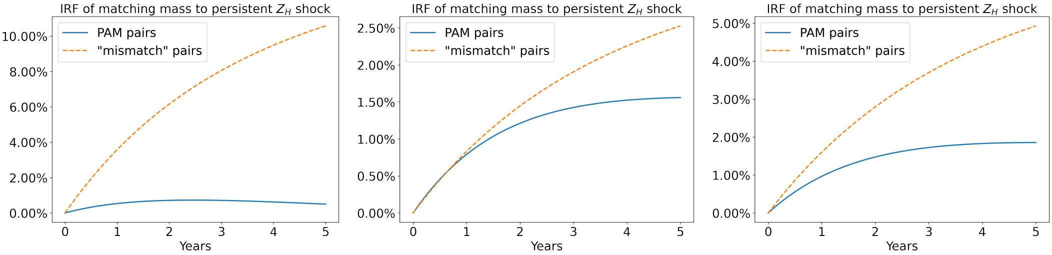
- Study how “mismatch” changes over the business cycle.



“PAM” pairs: pairs where x & y are close. “Mismatch”: pairs where x & y are not close.

Sorting Over Business Cycles

- Countercyclicality of sorting depends on bargaining power.



Left to right $\beta = 0$ (Lise-Robin '17), 0.72 (benchmark), 1.

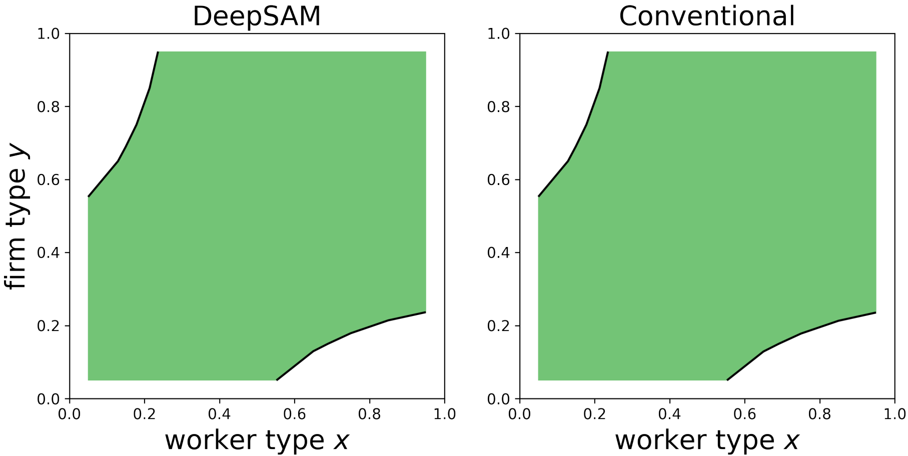
Table of Contents

Conclusion and Future Work

- We develop a global solution method, DeepSAM, to search and matching models with heterogeneity and aggregate shocks.
- We apply DeepSAM to canonical labor search models, and find interaction between heterogeneity and aggregate shocks that we cannot study before.
- A powerful new tool to be combined with rich data of heterogeneous workers and firms over business cycles!
- More applications:
  - Search in OTC market.
  - Spatial and network models with aggregate uncertainty.
  - ...

Thank You!

DeepSAM vs Conventional method at DSS: discrete case



back

Free Entry Condition

- Instead of assuming exogenous g_f, we assume free entry with entry cost c:

$$c = \mathbb{E}[V_t^v] = \int V^v(\tilde{y}, z, g) d\tilde{y}. \tag{1}$$

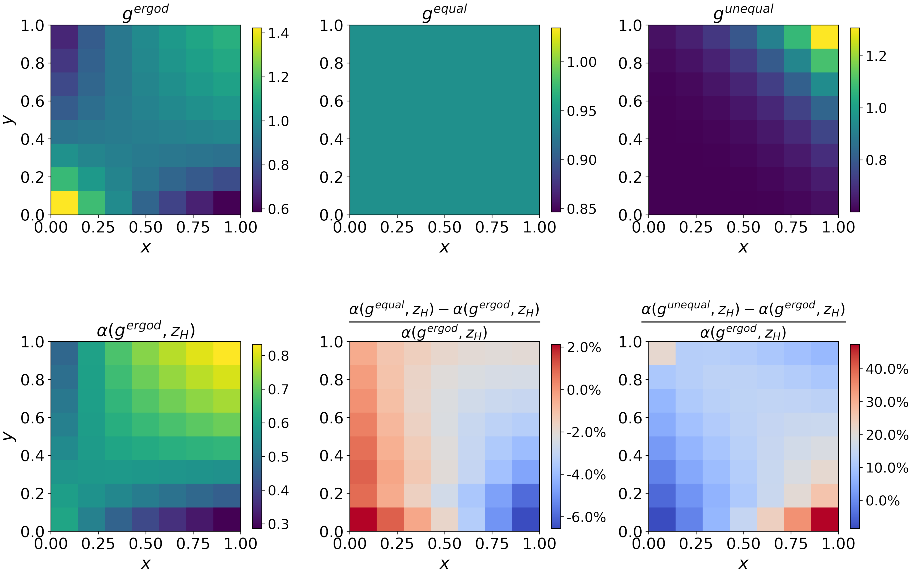
- As the matching function is homothetic $\frac{m(z_t, g_t)}{\mathcal{V}_t} = \widehat{m}\left(\frac{\mathcal{V}_t}{\mathcal{U}_t}\right)$, combining free entry condition with HJB equation for V^v gives:

$$\widehat{m}\left(\frac{\mathcal{V}_t}{\mathcal{U}_t}\right) = \frac{\rho c}{\int\int \alpha(\tilde{x}, \tilde{y}) \frac{g_t^u(\tilde{x})}{\mathcal{U}_t}(1-\beta) S_t(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y}} \Rightarrow \quad \mathcal{V}_t = \mathcal{U}_t \widehat{m}^{-1}(\cdots) \tag{2}$$

where $g_t^u = g_t^w - \int g_t^m(x, y) dy$ and so the RHS can be computed from g_t^m and S_t.
- $g_t^f = \mathcal{V}_t + \mathcal{P}_t$, where $\mathcal{V}_t$ and $\mathcal{P}_t$ can be expressed in terms of g and S.
- With free entry condition, the master equation expression for surplus takes the same form as before but with different expressions of $g^f(y)$.

back

Variation in α as the Distribution Varies



On-the-job-search: KFE

- The KFE becomes:

$$
\begin{aligned}
dg_t^m(x,y) = & -\delta g_t^m(x,y)dt \\
& -\phi\frac{m(\mathcal{W}_t,\mathcal{V}_t)}{\mathcal{W}_t\mathcal{V}_t} g_t^m(x,y)\int \alpha_t^p(x,y,\tilde{y})g_t^v(\tilde{y})d\tilde{y}dt \\
& +\frac{m(\mathcal{W}_t,\mathcal{V}_t)}{\mathcal{W}_t\mathcal{V}_t}\alpha_t(x,y)g_t^u(x)g_t^v(y)dt \\
& +\phi\frac{m(\mathcal{W}_t,\mathcal{V}_t)}{\mathcal{W}_t\mathcal{V}_t}\int \alpha_t^p(\tilde{x},\tilde{y},y)g_t^v(y)\frac{g_t^m(\tilde{x},\tilde{y})}{\mathcal{E}_t}d\tilde{x}d\tilde{y}dt
\end{aligned}
$$

back