Deep Learning for Search and Matching Models*

Jonathan Payne† Adam Rebei‡ Yucheng Yang§

February 1, 2025

First draft: February 15, 2024

Abstract

We develop a new method to globally solve and estimate search and matching models with aggregate shocks and heterogeneous agents. We characterize general equilibrium as a high-dimensional partial differential equation with the distribution as a state variable. We then use deep learning to solve the model and estimate economic parameters using the simulated method of moments. This allows us to study a wide class of search markets where the distribution affects agent decisions and compute variables (e.g. wages and prices) that were previously unattainable. In applications to labor search models, we show that distribution feedback plays an important role in amplification and that positive assortative matching weakens in prolonged expansions, disproportionately benefiting low-wage workers.

Keywords: Search and Matching, Distribution Feedback, Two-sided Heterogeneity, Business Cycles, Sorting, Over-the-Counter Financial Markets, Deep learning.

*We thank Fernando Alvarez, Isaac Baley, Jaroslav Borovička, Katarína Borovičková, Niklas Engbom, Jesús Fernández-Villaverde, Lukas Freund, Pablo Guerron, Lars Hansen, Ji Huang, Zhen Huo, Greg Kaplan, Dirk Krueger, Felix Kübler, Ricardo Lagos, Jeremy Lise, Lukas Mann, Guido Menzio, Giuseppe Moscarini, Andi Mueller, Galo Nuño, Jesse Perla, Xincheng Qiu, Jean-Marc Robin, Tom Sargent, Edouard Schaal, Venky Venkateswaran, Gianluca Violante, David Wiczer, and seminar participants at AMLEDS Webinar, ASU, Atlanta Fed, Chicago, CKGSB, EIEF, Minnesota, Norges Bank, NYU, PKU, PSE, Rice, Tsinghua, Yale, Zurich; ASSA, CEF, DSE, ESIF, OzMac, Sargent Alumni Reading Group, SED, Swiss Winter Macro Finance Workshop, Zurich Workshop on the Frontier of Quantitative Macro for helpful comments and suggestions. We thank Nikolai Kurlovich and Runhuan Wang for outstanding research assistance. We thank the Julis-Rabinowitz Center for Public Policy and Finance and the Swiss National Science Foundation for generous financial support.

†Princeton University, Department of Economics. Email: jepayne@princeton.edu

‡Stanford University, Institute for Computational & Mathematical Engineering. Email: rebei@stanford.edu

§University of Zurich, Department of Finance and SFI. Email: yucheng.yang@uzh.ch

1 Introduction

Heterogeneity in markets with search frictions is important for explaining economic phenomena such as wage dispersion, sorting, job ladders, and asset liquidity (e.g. Burdett and Mortensen (1998); Shimer and Smith (2000); Postel-Vinay and Robin (2002); Duffie, Gârleanu, and Pedersen (2005)). These markets feature volatile distribution dynamics over business cycles, during major crises, and in response to policy reforms (e.g. Moscarini and Postel-Vinay (2023); Chen, Cui, He, and Milbradt (2017)). However, modeling equilibrium in these markets has proven technically challenging because optimizing agents need to understand and forecast time-varying distributions. To ensure tractability, many researchers have simplified the problem by limiting how distributions impact agent decisions (e.g. the "block-recursive" equilibria in Lagos and Rocheteau (2009), Menzio and Shi (2011), Lise and Robin (2017)[1]). Developments in the deep learning literature have opened up the possibility of relaxing these restrictions. In this paper, we present a general formulation of heterogeneous agent search and matching models with aggregate shocks as a collection of high dimensional partial differential equations (PDEs) with the agent distribution explicitly as a state variable. We then develop the first deep learning method for solving these models and estimating economic parameters using the simulated method of moments. We use our method to study three foundational models in labor and financial search that have never been solved in their general form. This allows us to understand the interactions between aggregate shocks and heterogeneity during crises and business cycles.

We focus on models with the following features. The economy is populated by heterogeneous agents (e.g. workers or investors) and heterogeneous institutions (e.g. firms or financial intermediaries) that can be matched or unmatched. Matches generate utility that depends upon the idiosyncratic agent and institution types and an exogenous aggregate variable that follows a continuous time Markov chain. Unmatched agents and institutions engage in random search to meet each other. Upon meeting, they choose whether to accept the match and then bargain over the division of the matching surplus. We show that the equilibrium for this economy can be characterized recursively with a state space consisting of the exogenous aggregate variable and the endogenous distribution of matches across types in the economy. The

[1]Following Menzio and Shi (2011) and Lise and Robin (2017), we call an equilibrium "block-recursive" if the agents' value and policy functions are independent of the endogenous distribution of agents across their idiosyncratic states.

match distribution impacts agent decisions because the opportunity cost of accepting a match depends upon which other agents are looking for matches and which institutions are available to match. The equilibrium is then characterized by a "master" PDE for the surplus function, which includes high dimensional derivatives capturing the impact of distributional changes on surplus.

In Section 2, we propose a deep learning method for solving this class of PDEs and estimating economic parameters using the simulated method of moments, which we refer to as DeepSAM. To solve the model, we approximate the surplus function with a neural network (NN) and then train the neural network parameters to minimize the average loss in the PDE on a collection of sample points. We estimate the model using the simulated method of moments by building on the state-of-the-art practices in the deep learning literature. Instead of solving the model repeatedly across different economic parameter values, we introduce these parameters as pseudo-state variables and solve the resulting "extended master" equations using deep learning. This provides an explicit solution to the problem over a large range of structural parameter values. We then use this solution to build a surrogate model mapping structural parameters to simulated moments that we utilize for estimation.

To our knowledge, we are the first to apply deep learning to study search and matching models, which requires resolving a number of technical challenges. First, unlike some Walrasian market models (e.g. Krusell and Smith (1998)), the distribution does not enter the agent optimization problem only through aggregate prices. Instead, there is complicated feedback between the shape of the distribution and agent decisions because agents need to forecast who they will meet. Second, the free entry condition in search models typically requires solving a non-linear fixed point problem at each training step. Third, the algorithm must be stabilized across unusually shaped surplus functions and the economic parameter space. We resolve these problems by developing novel sampling techniques that help train the model in the economically meaningful state space, deriving an explicit representation of the firm distribution, and gradually introducing curvature into the solution algorithm.

We deploy our method in three search models that represent "canonical" environments in labor economics and finance. In each case, we relax assumptions previously imposed for tractability and study the general environment.

In Section 3, we solve a labor market search model with two-sided heterogeneity that has been extended to include aggregate crisis shocks. This model can be thought

of as either the Shimer and Smith (2000) model with two-sided heterogeneity and aggregate shocks, or as the Mortensen and Pissarides (1994) model with worker and firm heterogeneity. Even though this environment is a key building block in the literature, our method is the first to be able to solve the model without simplifying assumptions. We verify the accuracy of our solution in several ways. In our full model with aggregate shocks, we show that the average numerical error in the differential equation is in the order of 10^{-7}. In the simplified version of our model without aggregate shocks that can be solved with traditional methods, we show the average squared difference between our solution and the fixed point solution is in the order of magnitude of 10^{-6}.

We use our solution to study the role of distributional feedback during the COVID-19 pandemic. We calibrate the model to include a crisis state that generates the heterogeneous employment decline across worker skill groups and firm industries occurred during COVID, as estimated by Cajner, Crane, Decker, Grigsby, Hamins-Puertolas, Hurst, Kurz, and Yildirmaz (2020). We then compute the impulse response following a COVID shock and decompose the time path by comparing the results to the "restricted" dynamics when agents make matching decisions under the "myopic" belief that they are always at the ergodic employment distribution. We find that in the full model unemployment falls approximately 30% faster during the recovery than in the restricted model. This is because, in the full model, firms understand that COVID disproportionately increases unemployment among low-skilled workers, which leads them to forecast a higher opportunity cost of waiting for a high-skilled worker and so offer jobs to a wider range of workers. We also consider a counterfactual crisis where unemployment increases symmetrically across all agent types and show that for this case the restricted dynamics closely approximate the full dynamics. This illustrates that solving the model globally across the distribution is particularly important for understanding large, asymmetric shocks.

In Section 4, we study a labor market with on-the-job search and endogenous separation. Our environment builds on the recent literature (e.g. Lise and Robin (2017) and subsequent papers) but relaxes the assumptions commonly imposed to achieve block recursivity. In particular, we allow workers to possess positive bargaining power during both initial and on-the-job search meetings. We use our deep learning method to estimate the model to match empirical moments of the US labor market. The entire solution and estimation process takes approximately 5 hours, where the model

is solved over the economic parameter space and simulated across 10,000 parameter combinations to build the surrogate model deployed for the simulated method of moments. Our estimated model finds support for the Okun (1973) hypothesis that low-wage workers benefit disproportionately from longer expansions. This occurs because there is counter-cyclical sorting in our model that is more pronounced as expansions and recessions are prolonged. During an expansion (recession) high-type workers become increasingly scarce (plentiful) so high-type firms become less (more) picky in their job offers. Consequently, longer expansions lead to greater weakening of positive assortative matching, which benefits low-skilled workers. This offers an explanation for Okun's hypothesis through the feedback between distribution changes and acceptance decisions, a connection that is shut down in other models.

Our method allows us to study business cycle dynamics that have previously been too difficult to examine. One important example is the dynamics of the wage distribution. As pointed out by Lentz, Lise, and Robin (2017), even though the surplus function is block recursive in the commonly used Lise and Robin (2017) framework, the division of the surplus, i.e. the wage, is not. This means that wages necessarily depend upon the distribution and so cannot be computed with traditional methods. We use DeepSAM to overcome these challenges and solve for the high dimensional wage function. Our solution finds that wages of low-skilled workers are more procyclical over the business cycle. We also use our method to revisit the implications of the assumptions required to generate block-recursivity, e.g. that unemployed workers have no bargaining power when meeting with firms. By solving the model over a wide range of worker bargaining power values, we show that unemployment and vacancies are more responsive to business cycle shocks when worker bargaining power is small, indicating that the block recursive assumptions are quantitatively important.

In Section 5, we use DeepSAM to solve an over-the-counter (OTC) bond market model with heterogeneous investors, different bond maturities, and aggregate default risk. This can be thought of as an extension to Duffie, Gârleanu, and Pedersen (2005) and Weill (2008) that expands investor and asset heterogeneity and allows for aggregate risk. From a technical point of view, relative to the labor models in the earlier sections, this model introduces type switching and asset trade. We use our model to study how liquidity and institutional frictions impact bond prices at different maturities. We show that a financial crisis shock that increases the liquidity needs of hedge funds and increases default risk has more impact on long-maturity bonds. This offers

a search-theoretic rationale for the volatility of the term structure.

Literature Review. Over the past three decades, there have been major advances in solving search and matching models with heterogeneity and aggregate risk. One advance is the Bertrand competition model of wage setting introduced in Postel-Vinay and Robin (2002) and deployed in many papers (e.g. Cahuc, Postel-Vinay, and Robin (2006); Lise and Robin (2017)). Another is the directed search block recursive structure introduced by Moen (1997); Menzio and Shi (2010, 2011). Both these approaches impose contracting and entry assumptions that ensure agent decisions are independent of the distribution of matches. Our paper relaxes these constraints and solves a general class of models where the distribution may impact agents' decisions.

Our labor market models (Sections 3 and 4) are connected to recent papers studying business cycle dynamics in heterogeneous agent labor search models (e.g. Schaal (2017); Mueller (2017); Krusell, Mukoyama, Rogerson, and Şahin (2017); Moscarini and Postel-Vinay (2018); Engbom (2021); Fukui (2020); Baley, Figueiredo, and Ulbricht (2022); Alves (2022); Qiu (2023); Moscarini and Postel-Vinay (2023); Birinci, Karahan, Mercan, and See (2024); Gregory, Menzio, and Wiczer (2024)). Our DeepSAM approach offers a way to expand the range of models used in this literature by enriching agent heterogeneity, relaxing block recursivity, departing from perfect foresight, studying non-linear crisis dynamics, and potentially other extensions. We also show how to deploy deep learning to estimate these models at scale.

We are part of a growing literature using deep learning methods to solve economic models and overcome the limitations of traditional solution techniques. These papers have focused on solving heterogeneous agent macroeconomic models with incomplete but Walrasian markets (e.g. Azinovic, Gaegauf, and Scheidegger (2022), Maliar, Maliar, and Winant (2021), Han, Yang, and E (2021), Kahou, Fernández-Villaverde, Perla, and Sood (2021), Fernández-Villaverde, Hurtado, and Nuno (2023), Gopalakrishna (2021), Sauzet (2021), Huang (2022), Gu, Lauriere, Merkel, and Payne (2023), Azinovic and Žemlička (2023), Duarte, Duarte, and Silva (2024), Huang (2024), among others, see the recent review by Fernández-Villaverde, Nuno, and Perla (2024)). Our contribution is to show how to undertake deep learning to solve search and matching models, which are workhorse models for a large literature in macroeconomics, monetary economics, and finance. What makes these models difficult to solve compared to many competitive incomplete market models (e.g. Krusell and Smith

(1998)) is that the shape of the distribution matters directly for the equilibrium. This is because, as summarized in Table 1, the distribution impacts agents' decisions via the matching probability with other types rather than through aggregate prices. This imposes greater challenges on how we develop our numerical and sampling schemes to get an accurate solution.

	Distribution	How distribution affect agents' decisions
HAM	Asset wealth and income	Via aggregate prices
SAM	Type (productivity) of agents in two sides of matching	Via matching process with other types

Table 1: How distribution matters in heterogeneous agent models (HAM) vs search and matching (SAM) models.

Our estimation approach builds on the idea of introducing structural parameters as pseudo-state variables, first proposed by Norets (2012) and recently extended by Chen, Didisheim, and Scheidegger (2023); Kase, Melosi, and Rottner (2024); Friedl, Kübler, Scheidegger, and Usui (2023); Duarte and Fonseca (2024). Our PDE formulation is related to the master equations in Bilal (2023); Alvarez, Lippi, and Souganidis (2023). Finally, our training approach draws on the "Physics-informed neural networks" (PINN) literature (e.g. Raissi, Perdikaris, and Karniadakis (2019)) in computational science.

The paper is structured as follows. Section 2 describes our DeepSAM methodology for solving and estimating a general class of search and matching models. Section 3 applies DeepSAM to solve a canonical labor market search model with two-sided heterogeneity and aggregate shocks, and studies the impact of the COVID-19 shock. Section 4 applies DeepSAM to estimate a model with on-the-job search and endogenous separation. Section 5 applies DeepSAM to an over-the-counter bond market with heterogeneous investors, bond maturities, and aggregate shocks. Section 6 concludes.

2 Methodology

In this section, we outline a general environment that nests models across the search and matching literature. We introduce our deep learning method to solve the model. Finally, we outline how to use neural networks to undertake efficient estimation to match simulated moments (a deep learning based "simulated method of moments").

2.1 Environment

Setting: The economy is in continuous time with an infinite horizon. It is populated by a continuum of infinitely lived, heterogeneous agents. A unit measure of agents are workers indexed by type $x \in \mathcal{X}$ and a unit measure are firms indexed by type $y \in \mathcal{Y}$. We use workers and firms for notational convenience but we could equivalently label the different agents as investors and financial dealers, or other market participants. Workers are either employed (e) in a match or unmatched (u). Firms are either producing (p) in a match or vacant (v). The distribution of matches between workers and firms is endogenous and determined by agent decisions. All agents have a discount rate ρ. The exogenous aggregate state of the economy is indexed by $z_t \in \mathcal{Z}$, which follows a continuous time Markov chain with transition matrix Σ.

Match utility: If a worker is unmatched (unemployed), they get flow utility b. Workers match with firms but not with each other. If a worker of type x is matched with a firm of type y, then they generate transferable utility $F(x, y, z)$,[2] where F is increasing in each variable and twice differentiable with uniformly bounded first partial derivatives on $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Matches are destroyed at the exogenous rate $\delta(x, y, z)$ that potentially depends upon the idiosyncratic type and the aggregate state.

Distributions: Let $g_t^w(x)$ denote the population measure function of workers and let $g_t^f(y)$ denote the population measure function of firms. For expositional simplicity, here we focus on the case with an exogenous and time-invariant g^w and an entry condition that determines g_t^f.[3] Formally, following Hagedorn et al. (2017), new firms can enter the economy with a draw of y from the uniform distribution $U(0, 1)$. They pay a flow cost of c per period while operating in the economy. Let $g_t(x, y)$ denote the measure function of matched workers and firms, $g_t^e(x)$ denote the measure function of matched workers, $g_t^u(x)$ denote the measure function of unmatched workers, $g_t^p(y)$ denote the measure function of producing firms, and $g_t^v(y)$ denote the measure function of vacant firms. The relationships between the functions are given in Table 2 below. With some abuse of terminology, we follow the literature and refer to g,

[2] Our method can also handle non-transferable utility. Code to solve Smith (2006) with aggregate shocks is available upon request.

[3] Our method can also handle endogenous or time varying g_t^w. For example, in Section 5, we solve an over-the-counter model with investor-type switching.

g^e, g^u, g^p, g^v, g^w, and g^f as the "distributions" in the economy even though they are technically neither densities nor cumulative distribution functions. We define the aggregate worker employment by $\mathcal{E}_t := \int g_t^e(x)dx$, aggregate worker unemployment by $\mathcal{U}_t := \int g_t^u(x)dx$, aggregate producing firms by $\mathcal{P}_t := \int g_t^p(y)dy$, and aggregate vacant firms by $\mathcal{V}_t := \int g_t^v(y)dy$. Later we will show we can calculate all distributions from g_t and z and so (z_t, g_t) is a sufficient aggregate state space for the economy.

Description	Measure Function	Conditional Density
Matches	$g_t(x,y)$	
Employed workers	$g_t^e(x) = \int g_t(x,y)dy$	$g_t^e(x)/\mathcal{E}_t$
Unemployed workers	$g_t^u(x) = g_t^w(x) - g_t^e(x)$	$g_t^u(x)/\mathcal{U}_t$
Producing firms	$g_t^p(y) = \int g_t(x,y)dx$	$g_t^p(y)/\mathcal{P}_t$
Vacant firms	$g_t^v(y) = g_t^f(y) - g_t^p(y)$	$g_t^v(y)/\mathcal{V}_t$

Table 2: Summary of distributions

Search and Matching Technology: Only and all unmatched workers engage in random search. We generalize to include "on-the-job" search in Section 4. A function $m : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+, (\mathcal{U}_t, \mathcal{V}_t) \mapsto m(\mathcal{U}_t, \mathcal{V}_t)$ takes the current level of unemployment and vacancies and generates meetings. The rate at which a worker meets a potential firm is $\mathcal{M}_t^u := m(\mathcal{U}_t, \mathcal{V}_t)/\mathcal{U}_t$, while the rate at which a vacant firm meets a potential hire is $\mathcal{M}_t^v := m(\mathcal{U}_t, \mathcal{V}_t)/\mathcal{V}_t$. The rate at which a worker meets any firm $y \in Y \subset \mathcal{Y}$ equals $\mathcal{M}_t^u(\int_Y (g_t^v(y)/\mathcal{V}_t)dy)$, where $g_t^v(y)/\mathcal{V}_t$ is the firm density conditional on being vacant. The rate at which a firm meets any worker $x \in X \subset \mathcal{X}$ equals $\mathcal{M}_t^v(\int_X (g_t^u(x)/\mathcal{U}_t)dx$, where $g_t^u(y)/\mathcal{U}_t$ is the worker density conditional on being unemployed.

Surplus division: We impose that, upon matching, agents negotiate according to a generalized Nash Bargaining protocol so that workers get a fraction β of surplus and firms get the remaining fraction $1 - \beta$. The contract is implemented by providing a flow "wage" $w(x, y, z, g)$ to the worker and flow "profit" $f(x, y, z) - w(x, y, z, g)$ to the firm.

2.2 Recursive Characterization of Equilibrium

We now define and characterize a recursive equilibrium. The aggregate states are (z, g), where z is the aggregate productivity and g is the measure function over

matches. We guess (and later verify) that the law of motion for g takes the form[4]:

$$dg_t(x,y) = \mu^g(x,y,z,g)dt.$$

We denote agents' belief about the evolution of the match measure g by $\breve{\mu}^g(x,y,z,g)$, which equals to $\mu^g(x,y,z,g)$ in equilibrium.

2.2.1 Surplus Division

Let $V^u(x,z,g)$ denote the value of an unemployed worker of type x, $V^e(x,y,z,g)$ denote the value of a worker of type x employed at a firm of type y, $V^v(y,z,g)$ denote the value of a vacancy for firm y, and $V^p(x,y,z,g)$ denote the value of a producing firm y employing a worker of type x. The surplus of a match is defined as:

$$S(x,y,z,g) := V^p(x,y,z,g) - V^v(y,z,g) + V^e(x,y,z,g) - V^u(x,z,g)$$

The Nash Bargaining protocol implies that the division of surplus is given by:

$$\begin{aligned} \beta S(x,y,z,g) &= V^e(x,y,z,g) - V^u(x,z,g) \\ (1-\beta) S(x,y,z,g) &= V^p(x,y,z,g) - V^v(y,z,g) \end{aligned} \tag{2.1}$$

which implicitly determine the wage $w(x,y,z,g)$. Optimizing workers and firms accept the match when the surplus is positive so α is given by:

$$\alpha(x,y,z,g) := \begin{cases} 1, & \text{if } S(x,y,z,g) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.2}$$

Following the literature, we assume that contract terms are indexed to the contracts of new hires so that $V^e(x,y,z,g) - V^u(x,z,g)$ is the same for all workers with a particular (x,y) and $V^p(x,y,z,g) - V^v(y,z,g)$ is the same for all firms with a particular (x,y).

[4]The evolution of g does not have a Poisson shock from the continuous time Markov chain process for the exogenous aggregate state z because the change in z does not directly break up matches. Instead, it changes agents decisions and so the rate of the change in g. For further discussion, see Gu et al. (2023).

2.2.2 Agent Hamilton-Jacobi-Bellman (HJB) Equations

Given beliefs and optimal decisions, the value functions V^u, V^e, V^v, and V^p satisfy the Hamilton Jacobi Bellman (HJB) Equations:

$$\begin{aligned}
\rho V^u(x,z,g) &= b + \mathcal{M}^u \int \alpha(x,\tilde{y},z,g)(V^e(x,\tilde{y},z,g) - V^u(x,z,g))\frac{g^v(\tilde{y})}{\mathcal{V}}d\tilde{y} \\
&\quad + \sum_{\check{z}\neq z} \lambda(z,\check{z})(V^u(x,\check{z},g) - V^u(x,z,g)) + \langle D_g V^u, \breve{\mu}^g \rangle \qquad (2.3) \\
\rho V^e(x,y,z,g) &= w(x,y,z,g) - \delta(x,y,z)(V^u(x,z,g) - V^e(x,y,z,g)) \\
&\quad + \sum_{\check{z}\neq z} \lambda(z,\check{z})(V^e(x,y,\check{z},g) - V^e(x,y,z,g)) + \langle D_g V^e, \breve{\mu}^g \rangle \qquad (2.4) \\
\rho V^v(y,z,g) &= -c + \mathcal{M}^v \int \alpha(\tilde{x},y,z,g)(V^p(\tilde{x},y,z,g) - V^v(y,z,g))\frac{g^u(\tilde{x})}{\mathcal{U}}d\tilde{x} \\
&\quad + \sum_{\check{z}\neq z} \lambda(z,\check{z})(V^v(y,\check{z},g) - V^v(y,z,g)) + \langle D_g V^v, \breve{\mu}^g \rangle \qquad (2.5) \\
\rho V^p(x,y,z,g) &= F(x,y,z) - w(x,y,z,g) - \delta(V^v(y,z,g) - V^p(x,y,z,g)) \\
&\quad + \sum_{\check{z}\neq z} \lambda(z,\check{z})(V^p(x,y,\check{z},g) - V^p(x,y,z,g)) + \langle D_g V^p, \breve{\mu}^g \rangle \qquad (2.6)
\end{aligned}$$

where for $j \in \{u,e,v,p\}$, $D_g V^j$ is the Frechet derivative of V^j with respect to the measure function g,[5] $\langle f(y), h(y) \rangle = \int f(y)h(y)dy$ is the inner product, and α is an indicator function for the optimal match acceptance decision.

The HJB equation for unemployed agents can be interpreted in the following way. The left-hand side of (2.3) is the flow value of being unemployed. On the right-hand-side, the first term is the flow utility benefit, the second term is the meeting rate multiplied by the expected gain in a meeting, the third term is the value function shift when the exogenous aggregate state changes and the final term governs how the value function is impacted by distribution changes. The other HJB equations have a similar interpretation.

[5]The mean field game literature has studied the mathematical difficulties involved in defining a Frechet derivative for HJB equations with an infinite dimensional state (Cardaliaguet et al. (2019)). It has not been established whether these characterizations are appropriate for economic models. However, our numerical examples will always work with a finite type space where the derivatives are clearly defined.

2.2.3 Match Distribution Evolution

Given the agent matching decisions, the measure function of matches evolves according to:

$$dg_t(x,y) = -\delta(x,y,z)g_t(x,y)dt + \mathcal{M}_t^u g_t^u(x)\alpha(x,y,z,g)\frac{g_t^v(y)}{\mathcal{V}_t}dt$$

where the first term is the outflow from the breakup of matches and the second term is the inflow from the creation of new matches. Given g, and the firm measure function g^f, we can recover the other measures from Table 2. So, the Kolmogorov Forward Equation (KFE) for the match measure function can be expressed as:

$$\begin{aligned} dg_t(x,y) = &-\delta(x,y,z)g_t(x,y)dt + \frac{m(\mathcal{U}_t,\mathcal{V}_t)}{\mathcal{U}_t\mathcal{V}_t}\alpha_t(x,y)\left(g^w(x) - \int g_t(x,y)dy\right) \\ &\times \left(g^f(y) - \int g_t(x,y)dx\right)dt = \mu^g(x,y,z,g)dt. \end{aligned} \tag{2.7}$$

2.2.4 Free Entry and the Firm Distribution

Like in Hagedorn et al. (2017), the firm measure function g_t^f is determined by the "free-entry" condition[6]:

$$0 = \mathbb{E}[V_t^v] = \int V^v(\tilde{y},z,g)d\tilde{y}. \tag{2.8}$$

Combining the free-entry condition with equations (2.1) and (2.5) gives:

$$\frac{m(\mathcal{U}_t,\mathcal{V}_t)}{\mathcal{V}_t} = \frac{c}{\int\int \alpha(\tilde{x},\tilde{y},z_t,g_t)\frac{g_t^u(\tilde{x})}{\mathcal{U}_t}(1-\beta)S(\tilde{x},\tilde{y},z_t,g_t)d\tilde{x}d\tilde{y}}$$

where $g_t^u = g_t^w - \int g_t(x,y)dy$ and so the RHS can be computed from g_t and S_t. Conceptually, the matching rate depends upon the average surplus because new firms enter the model until it is no longer profitable to do so. If the matching function is homothetic in $\mathcal{U}_t$ and $\mathcal{V}_t$ (as is common in the literature), then we have $\frac{m(z_t,g_t)}{\mathcal{V}_t} = \widehat{m}\left(\frac{\mathcal{V}_t}{\mathcal{U}_t}\right)$ and so $\mathcal{V}_t$ can be solved for explicitly. Otherwise, we can deploy a non-linear fixed point solver for $\mathcal{V}_t$ in the loop of our algorithm. Since firm y draws are uniformly distributed, g_t^f is then given by $g_t^f = \mathcal{V}_t + \mathcal{P}_t$.

[6] As is common in the labor search literature, this makes g_t^f a jump variable. Our approach could be extended to consider alternative entry arrangements for which g_t^f is a state variable.

2.2.5 Equilibrium and Master Equation

Definition 1. A **(recursive) equilibrium** is a collection of functions $\{V^u, V^e, V^v, V^p, w, \alpha, g^f\}$ of the state variables (z, g) such that: (i) given beliefs about the evolution of g_t, $(V^u, V^e, V^v, V^p, \alpha)$ solve the HJB equations (2.3)-(2.6), (ii) the division of surplus satisfies (2.1), (iii) g^f satisfies the free entry condition (2.8), and (iv) agent beliefs about the evolution of g_t are consistent in the sense that $\breve{\mu}^g = \mu^g$, where μ^g is given by equation (2.7).

After combining the HJB equations and imposing belief consistency, the equilibrium can be characterized by the "master equation" for the surplus:

$$\begin{aligned} 0 = \mathcal{L}^S S =: & -\rho S(x,y,z,g) + F(x,y,z) + c - \delta(x,y,z)S(x,y,z,g) \\ & - (1-\beta)\frac{m(z,g)}{\mathcal{V}(z,g)} \int \alpha(\tilde{x},y,z,g) S(\tilde{x},y,z,g) \frac{g^u(\tilde{x})}{\mathcal{U}(z,g)} d\tilde{x} \\ & - b - \beta \frac{m(z,g)}{\mathcal{U}(z,g)} \int \alpha(x,\tilde{y},z,g) S(x,\tilde{y},z,g) \frac{g^v(\tilde{y})}{\mathcal{V}(z,g)} d\tilde{y} \\ & + \sum_{\check{z} \neq z} \lambda(z)(S(x,y,\check{z},g) - S(x,y,z,g)) + \langle D_g S(x,y,z,g), \mu^g(x,y,z,g) \rangle \end{aligned} \tag{2.9}$$

where μ^g is given by (2.7), $(\mathcal{U}, \mathcal{V}, g^u, g^v)$ can be calculated by Table 2, g^f is pinned down by (2.8), and α is given by equation (2.2). Once we obtain $S(x,y,z,g)$ and $\alpha(x,y,z,g)$ by solving Equation (2.9), we can obtain $\{V^u, V^e, V^v, V^p, w\}$ by solving equations (2.3), (2.4), (2.5), (2.6) together with the surplus division equation (2.1).

2.2.6 Relation to Environments in Other Papers

Block recursivity: We can compare our set-up to well-known papers in the search literature with block recursive equilibria. Lise and Robin (2017) sets $\beta = 0$, introduces a vacancy creation condition at each y, and assumes that all unmatched vacancies will be destroyed each period. They show this implies that α and S do not depend upon g[7], i.e. $\alpha(x,y,z,g) = \alpha(x,y,z)$ and $S(x,y,z,g) = S(x,y,z)$, which means their model is block-recursive in total surplus and acceptance decisions. In Appendix D.2, we derive these results using our continuous time environment and notation. Menzio and Shi (2011) introduces competitive search and free entry on one side of the market, which also implies that the value functions and agent decisions do not depend upon

[7]They also introduce on-the-job search, which we compare to in Section 4.

g. Our method allows us to relax the assumptions that lead to block recursivity and study a wide class of models in which α and S can explicitly depend upon g.

Dimension reduction: For models with incomplete but Walrasian markets, Krusell and Smith (1998) suggests replacing the distribution in the state space by its mean (and potentially other low dimensional moments). This may be an effective approach for some Walrasian market models because the distribution impacts agents' decisions by changing aggregate prices, which may primarily depend upon the low dimensional moments of the distribution. By contrast, in a search model, the distribution impacts agents' decisions by changing the probability distribution over which agents they meet. This ultimately enters the master equation on lines 3 and 4 of equation (2.9). There are no obvious low-dimensional moments of the distribution that are sufficient for evaluating these terms because we need to integrate across the surplus function, weighted by the acceptance decision and the density of searching agents.

2.2.7 Model Extensions

In Section 4, we extend the model to incorporate endogenous separation and on-the-job search with positive worker bargaining power. In Section 5, we extend the model to incorporate type switching and asset trade in an over-the-counter market. Our solution approach offers a foundation that can be used to study other important models in the search and matching literature such as within-firm heterogeneity for large firms, multi-dimensional sorting, non-transferable utility, and consumption-saving decisions.

2.3 Approximation with Finite Types

Our goal is to solve the master equation (2.9) numerically to obtain $S(x, y, z, g)$ and $\alpha(x, y, z, g)$. We can then solve for the value and wage functions using Equations (2.1) to (2.6). Solving Equation (2.9) is difficult because the state space contains an infinite dimensional variable, g, and so the master equation contains Frechet derivatives with respect to g. To make progress on this problem, we approximate g on a discretized type space so that equation (2.9) becomes a high, but finite-dimensional partial differential equation that can be solved using deep learning.[8]

[8]There are alternative discretization approaches that could also be deployed (e.g. a projection onto a finite set of basis functions). We choose to work with a finite type space because it generates

Discrete type space and KFE: To represent g, we restrict the possible types to finite collections: $x \in \mathcal{X} = \{x_1, \ldots, x_{n_x}\}$ and $y \in \mathcal{Y} = \{y_1, \ldots, y_{n_y}\}$. With some abuse of notation, we let $\underline{\boldsymbol{g}}_t$ denote the vector of measures of matched agents at the points $(\mathcal{X}, \mathcal{Y})$, where $g_{t,ij} = g_t(x_i, y_j)$ denotes g evaluated at type (x_i, y_j). We also let $g^w_{t,i} := g^w_t(x_i)$ and $g^f_{t,j} := g^f_t(y_j)$. The discretized aggregate state variables are: $\{z, \underline{\boldsymbol{g}}_t\}$, the aggregate productivity, and the vector of measures of agent matches. The Riemann approximation to the KFE on the discretized type space is given by:

$$\begin{aligned}\frac{dg_{t,ij}}{dt} = \mu^g(x_i, y_j, z_t, \underline{\boldsymbol{g}}_t) = -\delta(x_i, y_j, z) g_{t,ij} + \frac{m(z_t, \underline{\boldsymbol{g}}_t)}{\mathcal{U}(z_t, \underline{\boldsymbol{g}}_t)\mathcal{V}(z_t, \underline{\boldsymbol{g}}_t)} \alpha(x_i, y_j, z_t, \underline{\boldsymbol{g}}_t) \\ \times \left(g^w_{t,i} - \frac{1}{n_y} \sum_{k=1}^{n_y} g_{t,ik} \right) \left(g^f_{t,j} - \frac{1}{n_x} \sum_{l=1}^{n_x} g_{t,lj} \right), \qquad \forall i \leq n_x, j \leq n_y\end{aligned} \tag{2.10}$$

Master equation: The discretized Master equation for the Surplus is given by:

$$\begin{aligned}0 = \mathcal{L}^S S = &-(\rho + \delta(x,y,z)) S(x,y,z,\underline{\boldsymbol{g}}) + F(x,y,z) + c - b \\ &- (1-\beta) \frac{m(z,\underline{\boldsymbol{g}})}{\mathcal{U}(z,\underline{\boldsymbol{g}})\mathcal{V}(z,\underline{\boldsymbol{g}})} \frac{1}{n_x} \sum_{k=1}^{n_x} \alpha(x_k, y, z, \underline{\boldsymbol{g}}) S(x_k, y, z, \underline{\boldsymbol{g}}) g^u_k \\ &- \beta \frac{m(z,\underline{\boldsymbol{g}})}{\mathcal{U}(z,\underline{\boldsymbol{g}})\mathcal{V}(z,\underline{\boldsymbol{g}})} \frac{1}{n_y} \sum_{l=1}^{n_y} \alpha(x, y_l, z, \underline{\boldsymbol{g}}) S(x, y_l, z, \underline{\boldsymbol{g}}) g^v_l \\ &+ \sum_{k=1}^{n_x} \sum_{l=1}^{n_y} \partial_{g_{kl}} S(x_k, y_l, z, \underline{\boldsymbol{g}}) \mu^g(x_k, y_l, z, \underline{\boldsymbol{g}}) \\ &+ \sum_{\check{z} \neq z} \lambda(z) (S(x, y, \check{z}, \underline{\boldsymbol{g}}) - S(x, y, z, \underline{\boldsymbol{g}})),\end{aligned} \tag{2.11}$$

where μ^g is given by the discretized KFE (2.10), $(\mathcal{U}, \mathcal{V}, g^u, g^v)$ are calculated from Table 2 (after appropriate discretization), g^f comes from $g^f_t = \mathcal{V}_t + \mathcal{P}_t$, and we approximate $\alpha(x, y, z, \underline{\boldsymbol{g}})$ by $\alpha(x, y, z, \underline{\boldsymbol{g}}) = \left(1 + e^{-\xi S(x,y,z,\underline{\boldsymbol{g}})}\right)^{-1}$ to ensure the differentiability of the value function when there is a finite number of types.[9] The differential equations for the value functions, (2.3) to (2.6), have analogous discretized forms.

a very tractable approximation (2.10) to the KF equation.

[9]The "softened" α function can be interpreted as a logit choice model where utility shocks come from an extreme value distribution with parameter ξ.

2.4 The DeepSAM Method

In this section, we present the DeepSAM method for using deep learning to solve and estimate the general class of search and matching models outlined in Subsections 2.1 to 2.3. We first present, for given structural parameters, how to solve the high dimensional master equation (2.11). We then present how to estimate the model by introducing structural parameters as pseudo-state variables in the neural network and using a surrogate model to find the parameters that match simulated moments to the empirical data.

2.4.1 Solution Algorithm for Given Structural Parameters

Let $\boldsymbol{\omega} = (x, y, z, \underline{\boldsymbol{g}}) \in \Omega^{\omega}$ denote the state space. We approximate the surplus function, S, by a neural network, $\hat{S}: \ \Omega^{\omega} \times \Omega^{\Theta} \to \mathbb{R}, (\boldsymbol{\omega}, \boldsymbol{\Theta}) \mapsto \hat{S}(\boldsymbol{\omega}; \boldsymbol{\Theta})$, of the form:

$$
\begin{aligned}
\boldsymbol{h}^{(1)} &= \phi^{(1)}(W^{(1)}\boldsymbol{\omega} + \boldsymbol{b}^{(1)}) && \dots \text{Hidden layer 1} \\
\boldsymbol{h}^{(2)} &= \phi^{(2)}(W^{(2)}\boldsymbol{h}^{(1)} + \boldsymbol{b}^{(2)}) && \dots \text{Hidden layer 2} \\
&\vdots \\
\hat{S} &= \phi^{(H)}(W^{(H)}\boldsymbol{h}^{(H-1)} + \boldsymbol{b}^{(H)}) && \dots \text{Layer H, Surplus}
\end{aligned}
$$

where, using the terminology of the deep learning literature, H is the number of layers, the length of vector $\boldsymbol{h}^{(i)}$ is the number of neurons in layer i, $\phi^{(i)}$ is the activation function for layer i, and the collection $\boldsymbol{\Theta} = (W^1, \dots, W^{(H)}, b^{(1)}, \dots, b^{(H)}) \in \Omega^{\Theta}$ are the parameters for the neural network.

Our goal is to find the parameters of the neural network to approximately solve the master differential equation (2.11) globally across the state space.[10] Our approach is summarized in Algorithm 1. Essentially, we use stochastic gradient descent algorithms or their variants to train the neural network to minimize the average loss in the master equation on a collection of sample points. As with other neural network approaches, there are many implementation details involved with these generic steps. We discuss two: sampling and stability.

Procedure to construct the sample points Q^n: It is trivial to sample across different

[10] We follow the computational economics literature and refer to our approach as a "global solution", in contrast to "local solution" approaches which typically perturb around a steady state.

Algorithm 1: Generic Solution Algorithm

1. Approximate the surplus function by the neural network (NN) $S(x,y,z,g) \approx \widehat{S}(x,y,z,g;\boldsymbol{\Theta})$ with parameters $\boldsymbol{\Theta}$.
2. Start with initial NN parameter guess $\boldsymbol{\Theta}^0$.
3. At iteration n with NN parameter vector $\boldsymbol{\Theta}^n$:
   (a) Generate K sample points, $Q^n = \left\{\left(x_k, y_k, z_k, \{g_{ij,k}\}_{i\leq n_x, j\leq n_y}\right)\right\}_{k\leq K}$.
   (b) Calculate the average mean squared error of the surplus master equation (2.11) on the sample points:
   $$L\left(\boldsymbol{\Theta}^n, Q^n\right) := \frac{1}{K}\sum_{k\leq K}\left|\mathcal{L}^S\widehat{S}\left(x_k, y_k, z_k, \{g_{ij,k}\}_{i\leq n_x, j\leq n_y}\right)\right|^2$$
   (c) Update NN parameters using stochastic gradient descent (SGD) or a variant. For example by $\boldsymbol{\Theta}^{n+1} = \boldsymbol{\Theta}^n - \zeta^n \nabla_{\boldsymbol{\Theta}} L\left(\boldsymbol{\Theta}^n, Q^n\right)$, for $\zeta^n > 0$
   (d) Repeat until $L\left(\boldsymbol{\Theta}^n, Q^n\right) \leq \epsilon$ with precision threshold ϵ.
4. Once S has converged, we can train neural network approximations to the value and wage functions ($\widehat{V}^u(x,z,g;\boldsymbol{\Theta^u})$, $\widehat{V}^e(x,y,z,g;\boldsymbol{\Theta^e})$, $\widehat{V}^v(y,z,g;\boldsymbol{\Theta^v})$, $\widehat{V}^p(x,y,z,g;\boldsymbol{\Theta^p})$, $\widehat{w}(x,y,z,g;\boldsymbol{\Theta^w})$) with NN parameters $(\boldsymbol{\Theta^u}, \boldsymbol{\Theta^e}, \boldsymbol{\Theta^v}, \boldsymbol{\Theta^p}, \boldsymbol{\Theta^w})$ that solve the discretized versions of equations (2.1) to (2.6).

values of x, y, z because they are low dimensional. We can draw values of (x, y, z) using a Latin hypercube sampler or a multinomial sampler on discretized grids. Sampling across the discretized distribution $(g_{i,j})_{i\leq n_x, j\leq n_y}$ is more complicated because it is high dimensional. We first solve the model at the steady state for the different fixed values of z. We then draw distributions $((g_{ij,k})_{i\leq n_x, j\leq n_y})_{k\leq K}$ that are perturbed random combinations of the steady state distributions for the different z. This approach samples values of g that are likely to be similar to the ergodic distribution and so increases the solution accuracy in regions of the state space that are economically interesting. Once the error is small, we can use our current solution guess to simulate the economy and then use the simulated distribution path as the sample points.

Algorithm stability: It is difficult to stabilize the algorithm when $\widehat{S}(x,y,z,g;\boldsymbol{\Theta})$ has sharp curvature (e.g., at regions where agents are marginal about whether to accept

a match or not). When these problems appear, we use the following "homotopy"-style procedure. We start by solving the problem for economic parameters that give low curvature in $\widehat{S}$. We then change parameters to move closer to the desired values and retrain the neural network starting from the previous solution. We keep changing parameters and retraining until we have reached a solution at the desired parameters.

2.4.2 Solution and Estimation

A key goal of quantitative macroeconomics is to estimate structural model parameters so that the model generates output that is consistent with observed data moments. Traditionally, this requires solving the model repeatedly for different parameter values, which can be computationally intensive. Deep learning based solution methods offer an advantage for model estimation (e.g., Chen et al. (2023); Kase et al. (2024)). By introducing economic parameters as pseudo-state variables in the neural network, we can efficiently solve models simultaneously across both the state space and the economic parameter space. We can then estimate the model parameters using the method of simulated moments.

In this section, we outline our algorithm for solving and estimating the model. Let $\mathbf{\Psi} \in \Omega^{\Psi}$ denote the structural parameters that will be estimated (or calibrated internally). Let $\tilde{\varphi} = (\tilde{\varphi}_1, \ldots, \tilde{\varphi}_N)$ denote the $N \times 1$ data moments that we want to match. Let $\varphi(\mathbf{\Psi}) = (\varphi_1(\mathbf{\Psi}), \ldots, \varphi_N(\mathbf{\Psi}))$ denote the corresponding model moments for given structural parameters $\mathbf{\Psi}$, that will be generated by simulating the model. Our goal is to estimate the model by choosing parameters $\mathbf{\Psi}$ to solve:

$$\hat{\mathbf{\Psi}} = \arg\min_{\mathbf{\Psi}} \sum_{i=1}^{N} \gamma_i \left(\frac{\tilde{\varphi}_i - \varphi_i(\mathbf{\Psi})}{\tilde{\varphi}_i} \right)^2, \tag{2.12}$$

where $\{\gamma_i\}_{i \leq N}$ are fixed weights.

The main computational challenge of estimation is to solve $\varphi_i(\mathbf{\Psi})$ for many different values of the structural parameters. We overcome this challenge by introducing $\mathbf{\Psi}$ as a pseudo-state vector for the surplus function, and solving the resulting extended

master equation, which in its discretized form is given by:

$$
\begin{aligned}
0 = \mathcal{L}^S S = &-(\rho + \delta(x,y,z))S(x,y,z,\underline{\boldsymbol{g}},\boldsymbol{\Psi}) + F(x_i,y_j,z) + c - b \\
&- (1-\beta)\frac{m(z,\underline{\boldsymbol{g}})}{\mathcal{U}(z,\underline{\boldsymbol{g}})\mathcal{V}(z,\underline{\boldsymbol{g}})}\frac{1}{n_x}\sum_{k=1}^{n_x}\alpha(x_k,y,z,\underline{\boldsymbol{g}},\boldsymbol{\Psi})S(x_k,y,z,\underline{\boldsymbol{g}},\boldsymbol{\Psi})g_i^u \\
&- \beta\frac{m(z,\underline{\boldsymbol{g}})}{\mathcal{U}(z,\underline{\boldsymbol{g}})\mathcal{V}(z,\underline{\boldsymbol{g}})}\frac{1}{n_y}\sum_{l=1}^{n_y}\alpha(x,y_l,z,\underline{\boldsymbol{g}},\boldsymbol{\Psi})S(x,y_l,z,\underline{\boldsymbol{g}},\boldsymbol{\Psi})g_j^v \\
&+ \sum_{k=1}^{n_x}\sum_{l=1}^{n_y}\partial_{g_{kl}}S(x_k,y_l,z,\underline{\boldsymbol{g}},\boldsymbol{\Psi})\mu^g(x_k,y_l,z,\underline{\boldsymbol{g}},\boldsymbol{\Psi}) \\
&+ \sum_{\check{z}\neq z}\lambda(z)(S(x,y,\check{z},\underline{\boldsymbol{g}},\boldsymbol{\Psi}) - S(x,y,z,\underline{\boldsymbol{g}},\boldsymbol{\Psi})),
\end{aligned}
$$

We approximate the extended surplus function S by a neural network $\hat{S}: \Omega^{\omega}\times\Omega^{\Psi}\times\Omega^{\Theta}\to\mathbb{R}, (\boldsymbol{\omega},\boldsymbol{\Psi},\boldsymbol{\Theta})\mapsto\hat{S}(\boldsymbol{\omega},\boldsymbol{\Psi};\boldsymbol{\Theta})$ and solve it in a similar way as for equation (2.11). The main additional complexity is that the structural parameter vector $\boldsymbol{\Psi}$ is now also part of the state space. This means that we need to sample from the parameter space and expand the range of sampled distributions to incorporate the variation in the parameter space. For the former, we draw economic parameters from a Latin hypercube. For the latter, we first solve the model at the steady state for different fixed values of z and economic parameters. Then, as before, we draw distributions that are perturbed random combinations of these steady state distributions.

We then use the solution to the extended master equation to build a surrogate model that maps economic parameters to simulated moments: $\Phi: \Omega^{\Psi}\times\boldsymbol{\Theta}^{\Phi}\to\mathbb{R}^N, \Psi\mapsto\Phi(\Psi;\boldsymbol{\Theta}^{\Phi})$. We do this by simulating the model under many different structural parameter vectors $\{\boldsymbol{\Psi}_l\}_l$ (using parallelized processing) and computing the resulting model moments $\{\varphi(\boldsymbol{\Psi}_l) = (\varphi_1(\boldsymbol{\Psi}_l),\ldots,\varphi_N(\boldsymbol{\Psi}_l))\}_l$. We then construct the surrogate model by training an additional neural network to approximate the relationship $\Phi(\Psi)$ using the simulated data. With $\Phi(\boldsymbol{\Psi})$, we can obtain the estimated parameters by solving the optimization problem (2.12). Our approach is summarized in Algorithm 2.

Algorithm 2: Generic Solution and Estimation Algorithm

1. Train the model across the state and structural parameter spaces:
   (a) Approximate the surplus function by a neural network that includes the economic parameters $\mathbf{\Psi}$ as inputs $S(x, y, z, g, \mathbf{\Psi}) \approx \widehat{S}(x, y, z, g, \mathbf{\Psi}; \mathbf{\Theta})$.
   (b) Train the neural network, as in Algorithm 1 but with the expanded sampling scheme described in the text.
2. Compute a surrogate model mapping structural parameters to model moments $\Omega^{\Psi} \times \Omega^{\Phi} : (\Psi, \mathbf{\Theta}^{\Phi}) \mapsto \Phi(\Psi; \mathbf{\Theta}^{\Phi})$ and optimize over Φ to solve the distance minimization problem (2.12).

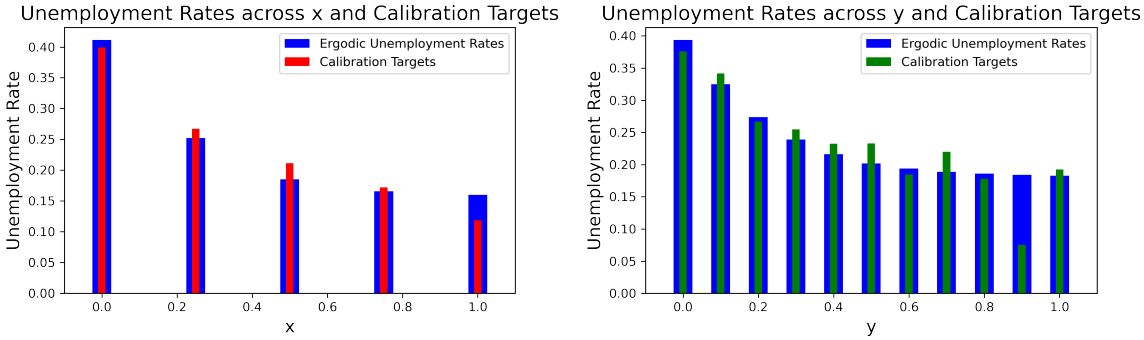
3 Labor Market With Two-Sided Heterogeneity

In this section, we use our method to study a simple search and matching model with two-sided heterogeneity and aggregate shocks. Conceptually, this model extends Shimer and Smith (2000) by incorporating aggregate shocks and Mortensen and Pissarides (1994) by introducing worker and firm heterogeneity. Even though this setup is a widely used building block in the literature, to our knowledge, we are the first to solve this model without simplifying assumptions. The goal of this Section is three-fold: first, we illustrate our DeepSAM solution method; second, we evaluate its numerical performance, highlighting its accuracy and computational efficiency in addressing high-dimensional problems even with highly non-linear crises dynamics; lastly, we use the model to study the impact of the COVID-19 shock and show the importance of distribution feedback in aggregate dynamics. We leave the demonstration of our deep learning estimation approach to the richer setup in Section 4.

3.1 Model Details and Parameters

Our environment is a special case of Section 2.1 with the following features. The exogenous aggregate state z follows a three-state Markov chain, corresponding to expansion (z_H), normal recession (z_L), and major crisis (z_D), the latter of which we calibrate to the COVID-19 crisis but could also be interpreted as a depression shock. The match output $F(z, x, y) = A_z f(x, y)$ and the separation rate $\delta_z(x, y)$ depend upon the state z, where we use the notation that A and δ are indexed by a z subscript.

Economic parameters: The calibrated economic parameters are presented in Table 7 in Appendix A. All parameters are calibrated at the annual frequency. For non-crisis states, where possible, we take standard parameters from the literature. We calibrate the matching efficiency κ to target an ergodic average unemployment rate of 6%. For the crisis state, we calibrate the separation rate $\delta_D(x, y)$ across workers and firms to match the observed peak declines in employment levels during the COVID-19 recession, as calculated in Cajner et al. (2020). Cajner et al. (2020) uses detailed data from a major US payroll company to estimate the employment drop for workers in different skill groups (corresponding to five groups in our model) and firms in different two-digit NAICS industries (corresponding to 11 groups in our model) during the COVID-19 recession. These drops peaked in April 2020, about 0.2 years after the onset of the pandemic in the US. As shown in Figure 1, we calibrate $\delta_D(x, y)$ such that the model's simulated declines in employment across worker and firm groups following a change from the ergodic state to the crisis state z_D for $t = 0.2$ years match the corresponding empirical moments. The detailed values of $\delta_D(x, y)$ are provided in Appendix C.1.



(a) Calibration along worker dimension

(b) Calibration along firm dimension

Figure 1: Calibrating $\delta_D(x, y)$ to match effect of COVID-19 shock on heterogeneous workers and firms in Cajner et al. (2020).

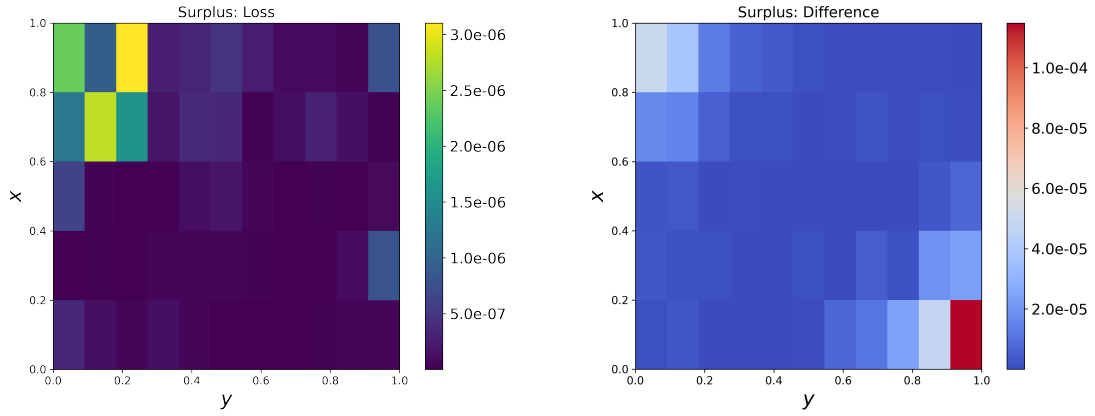
Neural network parameters: We use a fully connected feed-forward network with 4 layers, 50 neurons per layer, and a $tanh(\cdot)$ activation function. We describe additional details of the neural network approximation and sampling in Table 6 in Appendix A.

3.2 Numerical Performance

Table 3 summarizes the numerical performance of the DeepSAM method. The numerical loss and computational time for solving the model are reported in the first column. We achieve small errors, on the order of $\mathcal{O}(10^{-6})$, across our sample. This level of accuracy was reached within 4 hours and 20 minutes using an A100 GPU on Google Colab, a platform accessible to all researchers. Notably, no existing method has been able to solve our 58-dimensional PDE within such a timeframe. We discuss the detailed numerical scheme and the numerical stability of the results in Appendix C.3.[11]

	Model with agg. shock	Model without agg. shock
PDE Training Loss	2×10^{-6}	3.9×10^{-6}
MSE to Existing Solution	No existing solution	5×10^{-6}
Computational Time	4h 20min	57 min

Table 3: Numerical performance of DeepSAM for models with and without aggregate shocks. Computations are performed on the A100 GPU at Google Colab.



(a) Model with aggregate shock: loss across state space

(b) Model without aggregate shock: difference from conventional solution

Figure 2: Numerical accuracy across state space.

We also depict the numerical accuracy visually. Figure 2a shows the mean squared loss in the surplus master equation at a given distribution and exogenous aggregate

[11] The computation time varies with different calibrations. For example, if we set a relatively small $\kappa = 0.4$, it only takes less than 30 minutes to solve the 58-dimensional PDE. That's because the high dimensional function approximated by neural networks is flatter in the curvature, which makes it easier to "learn".

state realization. The loss is in the order of magnitude of 10^{-6} and not highly biased to a particular part of the state space.

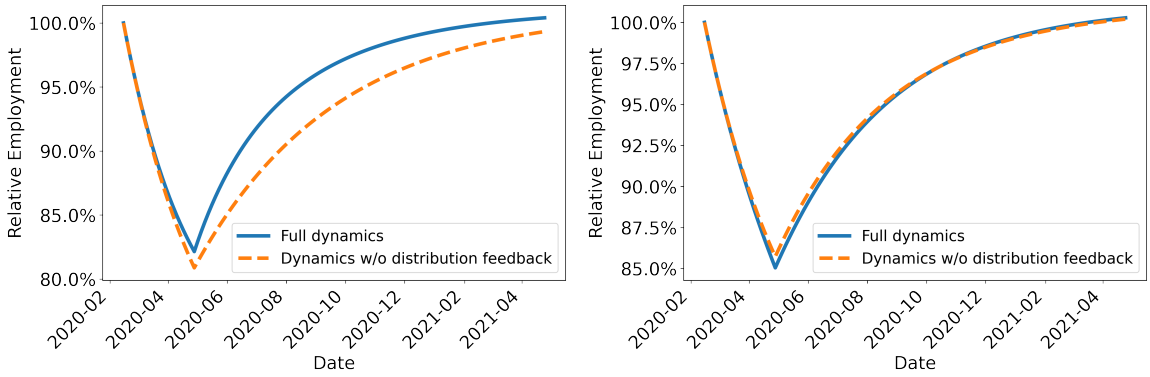
We further validate the DeepSAM method by solving a simpler, non-aggregate shock version of the model for which there are conventional solution techniques that can be used for comparison. We set $(z_t, \delta_t) \equiv (\bar{z}, \bar{\delta})$ and use the fixed point approach in Shimer and Smith (2000) and Hagedorn, Law, and Manovskii (2017) to solve for the deterministic steady state (DSS) solution $S^{\text{DSS}}(x, y)$. We then use DeepSAM to solve the 57-dimensional PDE for $S(x, y, \underline{\boldsymbol{g}})$ and compute the DeepSAM solution at the DSS by: $S^{\text{DSS}}_{\text{DeepSAM}}(x, y) = S(x, y, \underline{\boldsymbol{g}} = \underline{\boldsymbol{g}}^{DSS})$. We define the squared difference of the two methods for each (x, y) pair as $\|S^{\text{DSS}}_{\text{DeepSAM}}(x, y) - S^{\text{DSS}}_{\text{Conventional}}(x, y)\|^2$, where the mean squared difference takes the average of the squared difference across all (x, y) pairs. We report the average numerical performance in the second column of Table 3 and show the mean squared difference is of the order of 10^{-6}. Furthermore, Figure 2b plots the squared difference between the DeepSAM solution and conventional solution at the DSS. The difference is in the order of magnitude of 10^{-5} across the state space. Overall, we interpret our results in Table 3 and Figure 2 as evidence that our solution method has high accuracy.

3.3 Distribution Feedback to Aggregate Dynamics

An advantage of our DeepSAM method is that it can be used to study non-linear distributional dynamics in response to large aggregate shocks. This is because, even with aggregate risk, we can explicitly solve for the agent acceptance policy, α, as a function of the distribution and so can fully capture the feedback between distribution changes and aggregate dynamics. This is particularly valuable when aggregate shocks have asymmetric effects across the population, a phenomenon well-documented in a large empirical literature, such as by Guvenen, Schulhofer-Wohl, Song, and Yogo (2017). In this subsection, we illustrate these points by studying the recent COVID-19 crisis and disentangling the role of distribution feedback.

As discussed above, we model the COVID-19 shock as a change to the disaster state z_D that lasts for $t = 0.2$ years, followed by the recovery phase with stochastic aggregate shocks $z_t \in \{z_H, z_L, z_D\}$. Using our DeepSAM solution, we can simulate the dynamics of the distribution g_t with the Kolmogorov forward equation (2.10) and compute aggregate dynamics for unemployment, employment, average wage, and

other variables. The average path of employment change is plotted as the blue line of Figure 3a. As targeted in the data, we see a sharp drop in employment until April 2020, followed by the recovery phase.



(a) After COVID shock (b) After counterfactual symmetric shock

Figure 3: Relative change of employment compared to the initial level after the COVID-19 shock: full dynamics calculated with the KFE (2.10) vs "restricted" dynamics in which the decision of workers and firms are myopic to change in distribution g_t as in Equation (3.1). The left panel shows the dynamics after the true asymmetric COVID-19 shock, while the right one corresponds to the counterfactual symmetric shock.

To understand the contribution of distribution feedback to aggregate dynamics, we decompose the impulse responses of employment into two channels:

(i) the change of employment when agents' acceptance decision is always evaluated at the long-run ergodic employment distribution, and

(ii) the additional change of employment when the acceptance function reacts to the changing distribution of matches.

We refer to the former as the "restricted" dynamics for the experiment without the additional feedback from the distribution to the acceptance function. Mathematically, under the restricted dynamics, the distribution g_t^R evolves according to:

$$\frac{dg_t^R(x,y)}{dt} = -\delta(x,y,z_t)g_t^R(x,y) + \frac{m_t(z_t,\underline{\boldsymbol{g}}_t^R)}{\mathcal{U}_t(\underline{\boldsymbol{g}}_t^R)\mathcal{V}_t(\underline{\boldsymbol{g}}_t^R)}\alpha(x,y,z_t,\underline{\boldsymbol{g}}^{\text{ergodic}})g_t^{u,R}(x)g_t^{v,R}(y) \quad (3.1)$$

We plot the restricted dynamics with the dashed orange line of Figure 3a. Compared to the full dynamics calculated with the KFE (2.10), the difference in Equation (3.1)

is that agents make decisions assuming the distribution is always at the erdogic state. Thus the gap between the full and restricted dynamics can be interpreted as the contribution of distribution feedback to aggregate impulse responses.

From Figure 3a, we can see that distribution feedback accounts for approximately 30% of the change in employment dynamics during the recovery phase. When agents make decisions based on the true distribution change, the aggregate employment recovers faster than in the economy where agents ignore distribution changes. Under the full dynamics, the economy recovers approximately 2/3 of the employment losses by three months after the end of COVID-19, while under the restricted dynamics it only recovers approximately 1/3 of the losses.

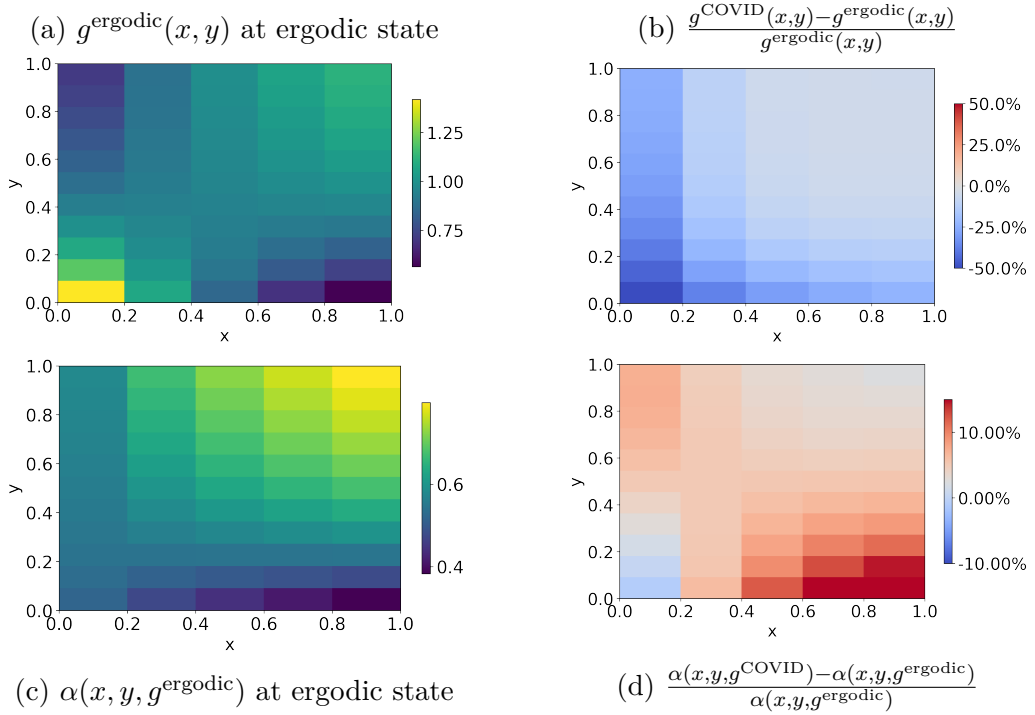


Figure 4: Distribution of matches and associated acceptance decision. Panels (a) and (c): distribution and acceptance at ergodic steady state. Panels (b) and (d): relative difference of distribution and acceptance at the end of the asymmetric COVID-19 shock, compared to the ergodic steady state.

To understand why distribution feedback is significant, in Figure 4 we plot how the agents' decision functions before and after the COVID-19 shock. The left panels of Figure 4 depict the ergodic match distribution and the acceptance functions at the ergodic match distribution. Evidently, the matching function exhibits positive

assortative matching (PAM) in the sense that workers and firms tend to match with counterparts with similar productivity levels. The right panels of Figure 4 depict the change in the distribution and the matching function at the end of the COVID-19 recession. Matches are disproportionately destroyed at the lower ends of both the worker and firm distributions. For high-type workers and firms, this decreases the relative probability of finding a match with a high-type counterpart, namely $\frac{g_v(y)}{\mathcal{V}}$ and $\frac{g_u(x)}{\mathcal{U}}$ decrease for large x and y. As a result, high-type workers and firms become less "picky" and are more willing to accept matches with low types. This is reflected in the increase in α for high x to low y pairs and low x to high y pairs. For low-type workers and firms, the probability of finding a high-type match also decreases but the probability a high-type will accept the match significantly increases. The net result is that the two effects balance out and α for low x to low y matches stays the approximately same. Collectively, these changes show that the COVID-19 shock weakens positive assortative matching and speeds up the recovery.

To further explore the role of asymmetric shocks, we conduct a counterfactual analysis with a disaster shock that has a symmetric impact across workers and firms through the separation rate, namely $\widetilde{\delta}_D(x, y) \equiv \widetilde{\delta}_D$ for all x, y. We calibrate $\widetilde{\delta}_D$ such that the shock generates the same aggregate employment rate as our calibrated $\delta_D(x, y)$ at the peak of the COVID-19 recession. We plot the full and restricted dynamics of the economy after this counterfactual "symmetric" shock in Figure 3b. In this case, the trajectories of the "restricted" dynamics and the full solution are closely aligned, implying that the distribution feedback is small when aggregate shocks affect workers and firms symmetrically.

These findings underscore the significance of distribution feedback on aggregate dynamics in the labor market, particularly when workers and firms have heterogeneous exposure to large aggregate shocks. They also suggest that policy design should account for the asymmetric nature of shocks and the heterogeneous effects of the policies themselves. For example, the design of unemployment insurance should consider feedback from the distribution changes to fully understand its implications and efficacy.

4 On-The-Job Search (OTJS) and Business Cycles

In this section, we study business cycle shocks in a labor market with endogenous job-to-job and job-to-unemployment transitions. Conceptually, the model extends Lise and Robin (2017) by giving workers non-zero bargaining power, allowing for on-the-job bargaining beyond Bertrand competition, and generalizing the free entry condition, all of which are changes that break "block-recursivity". We estimate the model using the deep learning based simulated method of moments outlined in Section 2.4.2. We use the estimated model to study how the business cycle impacts wages and employment for workers and firms with different skill levels, the full analysis of which has previously been unattainable. We find support for the Okun (1973) hypothesis that low-wage workers benefit disproportionately from longer expansions.

4.1 Environment Changes to Section 3.1

Search and matching: All workers now engage in random search. The matching function becomes $m(\mathcal{W}_t, \mathcal{V}_t)$ where $\mathcal{W}_t := \mathcal{U}_t + \phi\mathcal{E}_t$ denotes the total mass of searchers and ϕ is the exogenous relative search intensity of employed workers. The probabilities that an unemployed or an employed worker meets a potential employer are given by $\mathcal{M}_t^u = \frac{m(\mathcal{W}_t,\mathcal{V}_t)}{\mathcal{W}_t}$ and $\mathcal{M}_t^e = \phi\frac{m(\mathcal{W}_t,\mathcal{V}_t)}{\mathcal{W}_t}$, while the probability that a vacant firm meets a potential hire is $\mathcal{M}_t^v = \frac{m(\mathcal{W}_t,\mathcal{V}_t)}{\mathcal{V}_t}$. Conditional on a firm meeting a worker, the probabilities that the worker is unemployed or employed are defined as $\mathcal{C}_u = \frac{\mathcal{U}}{\mathcal{U}+\phi\mathcal{E}}$ and $\mathcal{C}_e = \frac{\phi\mathcal{E}}{\mathcal{U}+\phi\mathcal{E}}$ respectively. So, the probability for a firm to meet an unemployed worker $x \in \mathcal{X}$ is $\mathcal{M}_v\mathcal{C}_u \int_X g^u(x)dx$ and the probability for a firm to meet an employed worker $x \in \mathcal{X}$ is $\mathcal{M}_v\mathcal{C}_u \int_X g^e(x)dx$.

Bargaining between unemployed workers and firms: As before, the surplus of a match between an unemployed worker and a vacant firm is defined as $S_t^u(x,y) := V_t^p(x,y) - V_t^v(y) + V_t^e(x,y) - V_t^u(x)$ and the division of surplus is given by generalized Nash bargaining: $\beta S_t^u(x,y) = V_t^e(x,y) - V_t^u(x)$ and $(1-\beta)S_t^u(x,y) = V_t^p(x,y) - V_t^v(y)$.

Bargaining on the job: If a worker in match (x,y) moves to another firm with productivity $\tilde{y}$, then the worker gets β share of the incremental surplus $S_t(x,\tilde{y}) - S_t(x,y)$, the new firm gets $1-\beta$ share of the incremental surplus, and the incumbent firm keeps their surplus $(1-\beta)S_t(x,y)$. Conceptually, this is a model where the incum-

bent firm cannot be made worse off by the move (e.g. because they have a veto over worker moves), which makes this setup both technically and economically appealing.[12]

Endogenous separation: In addition to allowing workers to search, we also allow matches to break up endogenously. If a worker decides to leave a match, then they return to unemployment at the rate η.

4.2 Equilibrium

Agent HJB equations: As before, the aggregate state variables are (z, g), where z is the aggregate productivity and g is the measure function of matches. Let $\breve{\mu}^g(x, y, z, g)$ denote the agents' belief about the law of motion for the distribution g. Unemployed workers once again have an HJB equation for V^u given by:

$$\begin{aligned}\rho V^u(x,z,g) = b + \mathcal{M}^u \int \alpha(x,\tilde{y},z,g)(V^e(x,\tilde{y},z,g) - V^u(x,z,g))\frac{g^v(\tilde{y})}{\mathcal{V}}d\tilde{y} \\ + \sum_{\check{z}\neq z} \lambda(z,\check{z})(V^u(x,\check{z},g) - V^u(x,z,g)) + \langle D_g V^u, \breve{\mu}^g\rangle\end{aligned}$$

where α is an indicator function for whether workers and firms accept the match, which under generalized Nash Bargaining is again given by (2.2).

Employed workers now choose whether to accept matches from other firms and whether to leave their jobs, which leads to the HJB equation for V^e:

$$\begin{aligned}&\rho V^e(x,y,z,g) = (\delta(x,y,z) + \eta\alpha^b(x,y,z,g))(V^u(x,z,g) - V^e(x,y,z,g)) \\ &+ w(x,y,z,g) + \mathcal{M}^e \int \alpha^e(x,y,\tilde{y},z,g)\beta(S(x,\tilde{y},z,g) - S(x,y,z,g))\frac{g^v(\tilde{y})}{\mathcal{V}}d\tilde{y} \\ &+ \sum_{\check{z}\neq z} \lambda(z,\check{z})(V^e(x,\check{z},g) - V^e(x,z,g)) + \langle D_g V^e, \breve{\mu}^g\rangle\end{aligned}$$

where α^b and α^e are indicator functions for whether a worker or firm chooses to break

[12] Our assumption that workers get a share β of incremental surplus from on-the-job transitions is different to Lise and Robin (2017), which imposes that workers get (i) no surplus when moving from unemployment to employment and (ii) the outcome of Bertrand competition between firms when moving from one job to another. Technically, our setup allows us to relax the $\beta = 0$ restriction in Lise and Robin (2017) while still allowing the problem to be characterized recursively without needing to add match history as a state variable. We also believe our setup makes economic sense because it ensures that the worker has similar bargaining power in all their matches.

up the match and whether workers and firms both accept an on-the-job match. Under generalized Nash Bargaining, these satisfy:

$$\alpha^b(x,\tilde{y}) := \begin{cases} 1, & \text{if } S(x,\tilde{y},z,g) < 0 \\ 0, & \text{otherwise} \end{cases}$$
$$\alpha^e(x,y,\tilde{y},z,g) := \begin{cases} 1, & \text{if } S(x,\tilde{y},z,g) \geq S(x,y,z,g) \text{ and } S(x,\tilde{y},z,g) \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Vacant firms can now hire both unemployed workers and employed workers, which leads to the HJB equation for V^v:

$$\begin{aligned} \rho V^v(y,z,g) = &-c + \mathcal{M}^v \mathcal{C}^u \int \alpha(\tilde{x},y,z,g)(V^p(\tilde{x},y,z,g) - V^v(y,z,g)) \frac{g^u(\tilde{x})}{\mathcal{U}} d\tilde{x} \\ &+ \mathcal{M}^v \mathcal{C}^e \int\int \alpha^p(y,\tilde{x},\tilde{y},z,g)(1-\beta)(S(\tilde{x},y,z,g) - S(\tilde{x},\tilde{y},z,g)) \frac{g_m(\tilde{x},\tilde{y})}{\mathcal{E}} d\tilde{x} d\tilde{y} \\ &+ \sum_{\check{z} \neq z} \lambda(z,\check{z})(V^v(y,\check{z},g) - V^v(y,z,g)) + \langle D_g V^v, \breve{\mu}^g \rangle \end{aligned}$$

where α^p is an indicator function for whether a firm and an employed worker accept a match. This is given by the analogous function to α^e:

$$\alpha^p(y,\tilde{x},\tilde{y},z,g) := \begin{cases} 1, & \text{if } S(\tilde{x},y,z,g) \geq S(\tilde{x},\tilde{y},z,g) \text{ and } S(\tilde{x},y,z,g) \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Finally, producing firms can now endogenously break up matches, which leads to the HJB equation for V^p:

$$\begin{aligned} &\rho V^p(x,y,z,g) = (\delta(x,y,z) + \eta \alpha^b(x,y,z,g))(V^v(x,z,g) - V^p(x,y,z,g)) \\ &+ F(x,y,z) - w(x,y,z,g) + \sum_{\check{z} \neq z} \lambda(z,\check{z})(V^p(x,y,\check{z},g) - V^p(x,y,z,g)) + \langle D_g V^p, \breve{\mu}^g \rangle \end{aligned}$$

Equilibrium: In Appendix D.1, we complete the equilibrium characterization by specifying the KFE and the surplus master equation (the OTJS analogs of (2.7) and (2.9)).

4.3 Deep Learning-Based Estimation

We estimate the parameters $\{\beta, \kappa, c, b, \delta\}$ (respectively the worker bargaining power, scale factor in the meeting function, entry cost, unemployment benefit, and exogenous

separation rate) to match the ergodic unemployment rate, vacancy rate, employment-to-employment transition rate, unemployment-to-employment transition rate, and employment-to-unemployment transition rate, denoted by $\mathbb{E}[U]$, $\mathbb{E}[V]$, $\mathbb{E}[EE]$, $\mathbb{E}[UE]$, $\mathbb{E}[EU]$ respectively, in the US data. The other parameters are taken from the literature and outlined in Table 8 in Appendix A. We discretize workers into seven productivity types ($n_x = 7$) and firms into eight productivity types ($n_y = 8$), which is sufficiently fine to match the empirical regularities in the data.

Parameter	Interpretation	Value	Fitted moment	Data	Model
β	Surplus division factor	0.727	$\mathbb{E}[U]$	0.058	0.058
κ	Scale for meeting function	15.88	$\mathbb{E}[V]$	0.037	0.037
c	Entry cost	9.46	$\mathbb{E}[EE]$	0.025	0.026
b	Worker unemployment benefit	0.03	$\mathbb{E}[UE]$	0.468	0.431
δ	Separation rates	0.02	$\mathbb{E}[EU]$	0.025	0.026

Table 4: Estimated Parameters and Targeted Moments.

We undertake the deep learning based estimation outlined in Section 2.4.2. The estimated parameters, the data moments, and the model moments are shown in Table 4. The target moments are taken from Lise and Robin (2017) and the transition flow moments (including expected rate of employment-to-employment, employment-to-unemployment, and unemployment-to-employment transitions) are presented as monthly values to match with the original paper.[13] Evidently, we are able to get a close match to all five moments. To help illustrate this visually, Figure 5 shows the surrogate neural network mapping $\Phi(\Psi)$ from economic parameters to the aggregate moment loss for the dimensions β and κ (holding the other parameters fixed at their optimal values). This illustrates part of the curvature in the surrogate function that allows DeepSAM to find the optimal parameters.

The computational times for each step of the estimation, as well as the associated numerical losses, are shown in Table 5. The entire solution and estimation process takes 5 hours and 5 minutes, where the model is solved over the economic parameter space and simulated across 10,000 parameter combinations to build the surrogate model deployed for the simulated method of moments. For reference, solving the problem for given structural parameter values, which is a 59-dimensional PDE, will take 55 minutes. To our knowledge, it's infeasible to solve and estimate such a high-

[13] Lise and Robin (2017) use the BLS data 1951:I–2012:IV to construct most of the moments, except for EE transition which is constructed with the CPS data 1994:I–2011:III.

dimensional problem within such a time frame using other approaches. Furthermore, the full estimation only takes 5 hours and 5 minutes, making our method practically useful for quantitative analysis.

	Solution Given the Value of Structural Parameters	Solution with Structural Parameters as Pseudo-states	Training Surrogate Model	Simulated Method of Moments	Entire Estimation
MSE Loss	1.97×10^{-6}	4.8×10^{-6}	6.13×10^{-7}	1.24×10^{-4}	-
Time	55min	4h 1min	1h 3min	1.4min	5h 5min

Table 5: Training loss and computational time for solving vs estimating the model. Computations are performed on the A100 GPU at Google Colab.

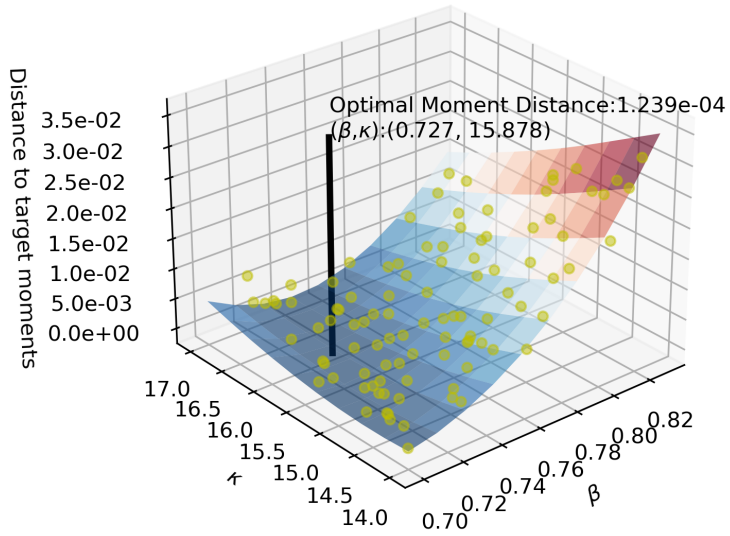


Figure 5: Target moments: $\mathbb{E}[U], \mathbb{E}[V]$. Parameters: matching efficiency κ, worker bargaining power β.

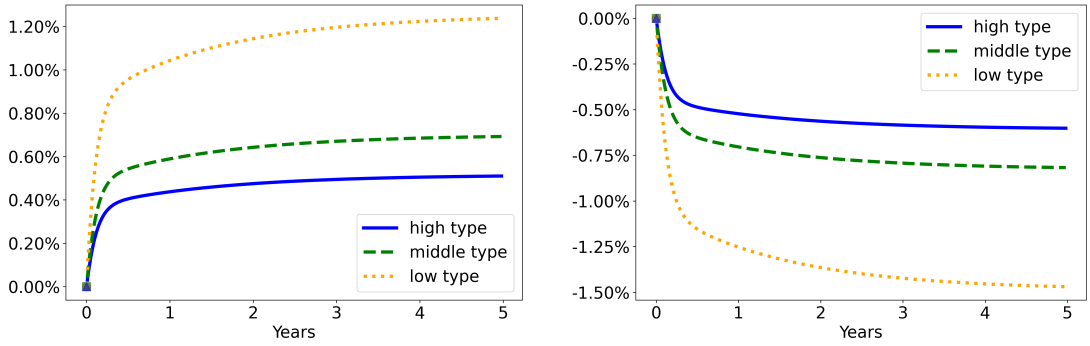
4.4 Dynamics of the Distribution Over Business Cycles

We use our estimated model to study how the business cycle impacts different workers and firms. We start by studying the cyclicality of wages across worker-firm matches (i.e. the variation in the division of surplus). We then study the cyclicality of employment across worker-firm matches (i.e. the variation along the extensive margin).

We find support for the hypothesis proposed by Okun (1973) that longer expansions disproportionally improve labor market outcomes for low-wage workers.

4.4.1 Dynamics of the Wage Distribution

A well-known difficulty in the heterogeneous agent random search literature is that "block-recursive" models cannot solve for wage dynamics. This is because the surplus division does not inherit the block recursive property of total surplus. As Lentz, Lise, and Robin (2017) write: "wages cannot be solved for exactly, indeed one needs to solve for a fixed point in worker values where the distribution of workers across jobs is a state variable." By contrast, our DeepSAM method can solve for wage dynamics because it solves for the surplus function explicitly as a function of the match distribution.



(a) IRF to positive shocks

(b) IRF to negative shocks

Figure 6: Wage change after aggregate shocks. Low, middle, and high types refer to the lowest, middle, and highest types in our $n_x = 7$-type discretization for workers.

Figure 6 shows the wage change across the worker distribution following positive or negative aggregate shocks that last for 5 years. Evidently, the wages of low-type workers are more procyclical.

4.4.2 Dynamics of Employment: A Search-Theoretical Explanation for Okun's Hypothesis

We also use our solution to study employment dynamics across heterogeneous workers and firms. Figure 7 shows the impulse responses for an economy that goes into recession for half a year and then recovers to the high state. The left panel shows that the low and high type workers experience a similar decline in unemployment for

the first quarter of the expansion. However, in subsequent quarters, low-type workers exhibit a more substantial decrease in unemployment. This demonstrates that low-type workers disproportionately benefit from longer expansions, consistent with the observations in Okun (1973).

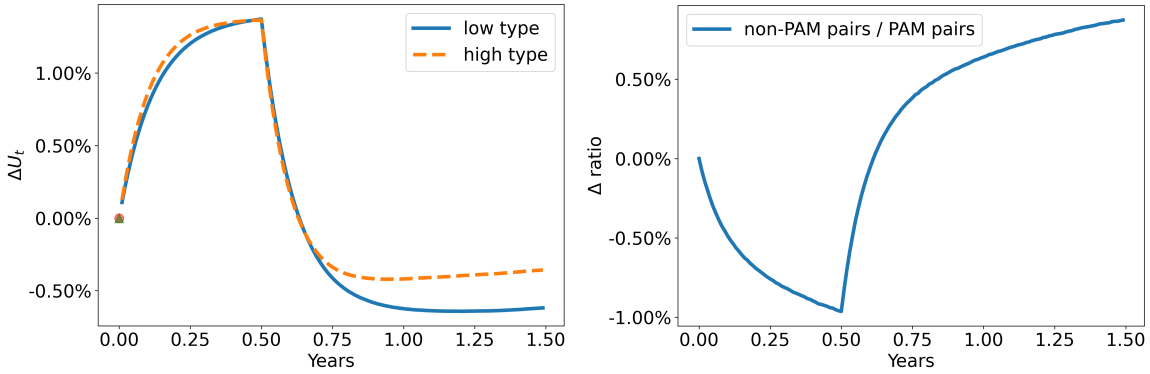


Figure 7: Impulse Responses. Left panel: Change of unemployment U_t for different type of workers. Low and high types refer to the lowest and highest types in our $n_x = 7$-type discretization. Right panel: Change in the ratio of "non-PAM pairs" to "PAM pairs".

In our model, Okun's hypothesis is generated through the impact of distribution dynamics on positive assortative matching (PAM, or positive sorting) during expansions of varying durations. To see this, consider the right panel of Figure 7, which depicts the ratio of "non-PAM pairs" to "PAM pairs". A matched pair (x, y) is called a "PAM pair" if $|x-y| \leq \frac{1}{2}$, and otherwise called a "non-PAM" pair. PAM pairs can be interpreted as worker-firm combinations with similar relative levels of idiosyncratic types. Evidently, the ratio of non-PAM to PAM pairs decreases during recessions and increases during expansions. This indicates that sorting is countercyclical, which aligns with the literature (e.g. Lise and Robin (2017)). More interestingly, assortative matching weakens further as expansions lengthen. This is because high-type workers become increasingly scarce during expansions so high-type firms become less picky. This leads to a progressive weakening of positive assortative matching over time, which disproportionately benefits low-type workers in prolonged expansions and so generates the patterns predicted by Okun. In other words, the distributional changes, which increase the relative employment of high and low type workers, determine the worker types who will benefit most from an additional period of economic expansion. Therefore, our model captures Okun's hypothesis just through the feedback between distribution changes and agent matching.

Due to its policy relevance, Okun's hypothesis has attracted recent attention. Existing literature (e.g. Alves and Violante (2023)) introduces complex model elements such as human capital and skill accumulation to explain Okun's observation. Interestingly, we can generate similar model predictions with a simpler model setup that drives dynamics through the interaction between distribution changes and acceptance decisions. This offers a complementary story that potentially has different policy implications. This is because, in our model, the decrease in low-skilled unemployment during long booms results from the formation of "non-PAM" pairs, which are more likely to dissolve during subsequent recessions (as in the right panel of Figure 7).

4.5 Revisiting "Block Recursivity" with On-the-job Search

With the DeepSAM method, we can solve heterogeneous agent random search models with on-the-job search and endogenous separation without imposing the assumptions needed to get block recursivity. A key assumption in Lise and Robin (2017) for getting block recursivity is setting $\beta = 0$ so unemployed workers get zero surplus when bargaining with a firm (see Appendix D.2). In Figure 8, we study how this restriction impacts impulse response functions. Keeping all other structural parameters the same, we vary β and plot the time paths for changes in the unemployment level (U_t), the vacancy level (V_t), the total quantity of poaching, and the total hires from unemployment in response to a 1.5% negative productivity shock that lasts for one year and then subsequently reverts to the stochastic process for aggregate TFP.[14]

We find that aggregate unemployment, vacancies, poaching, and hiring from unemployment all become more sensitive to aggregate shocks as $\beta \to 0$. Under our preferred estimation of $\beta = 0.727$, a 1.5% productivity shock leads to an increase in the unemployment rate of about 0.9%. By contrast, a block recursive setup predicts a more substantial 2.2% increase in the unemployment rate. This is because firm posting is more elastic when firms get the entire surplus from matches with unemployed workers, as shown in the upper right panel in Figure 8. We interpret these observations as evidence that the assumptions required to establish block recursivity are quantitatively important and place implicit restrictions on model estimation.

[14]The total quantity of *poaching* is defined as $\phi \frac{m(\mathcal{W}_t, \mathcal{V}_t)}{\mathcal{W}_t \mathcal{V}_t} \int \int \int \alpha_t^e(x, \tilde{y}, y) g_t^v(y) g_t(x, \tilde{y}) d\tilde{y} dx dy$. The total quantity of *hires from unemployment* is defined as $\frac{m(\mathcal{W}_t, \mathcal{V}_t)}{\mathcal{W}_t \mathcal{V}_t} \int \int \alpha_t(x, y) g_t^u(x) g_t^v(y) dx dy$.

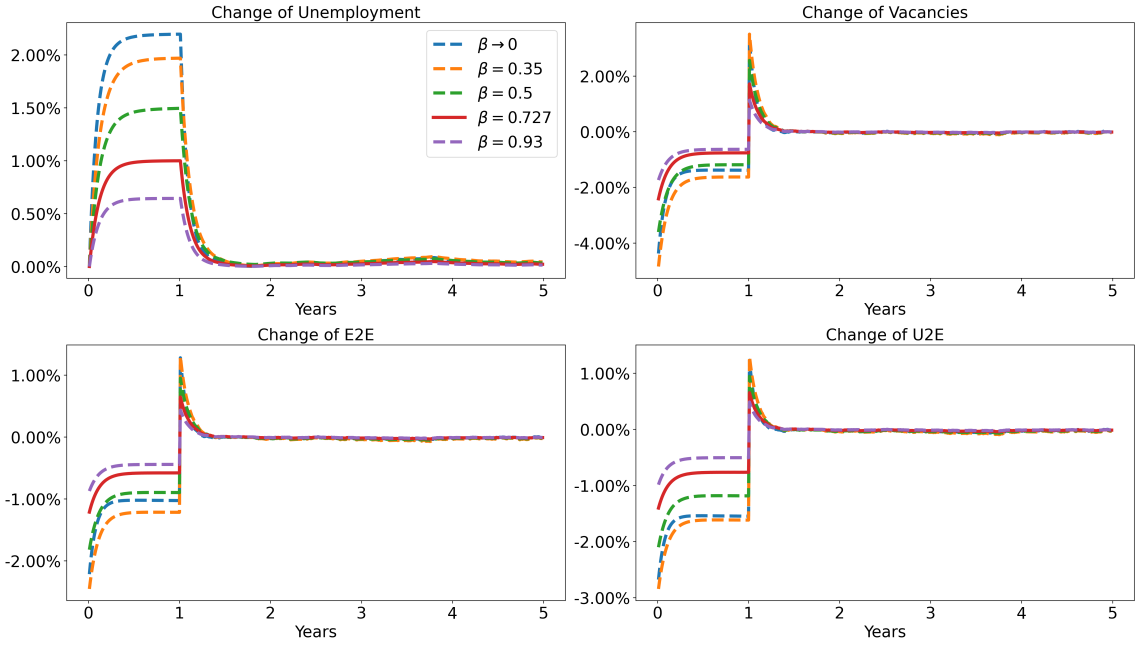


Figure 8: IRF following negative TFP Shock with different β's. After one year, the plots show the average time path across 1000 simulations. E2E and U2E are the employment-to-employment and unemployment-to-employment transition rates.

5 Over-the-Counter Financial Markets

In this final section, we consider financial crises in an over-the-counter (OTC) bond market with search and matching frictions. The model can be thought of as an extension to Duffie, Gârleanu, and Pedersen (2005) and Weill (2008) that expands the investor and asset heterogeneity and incorporates aggregate shocks.[15] We model bond duration explicitly and discuss how investor composition influences the emergent yield curve and bond market responses to financial crises. From a technical point of view, relative to the labor market models in Sections 3 and 4, this section introduces idiosyncratic type switching and asset trade.

5.1 Environment

Setting: Time is continuous with an infinite horizon. The economy has a collection of assets, indexed by $k \in \{1, \ldots, K\}$, which we interpret as bonds. Each asset k

[15] To ensure tractability, Duffie, Gârleanu, and Pedersen (2005) restricts the model to two types of investors and one type of asset. Weill (2008) expands the set of investors and assets but simplifies trading between agents who have different assets and does not include aggregate shocks.

has positive net supply s_k, pays a flow dividend $\delta > 0$ each period and one good at maturity, and matures at the rate $1/\tau_k$ (implying an average maturity of τ_k).

Investors and preferences: The economy is populated by a unit-mass continuum of infinitely-lived and risk-neutral investors indexed by $j \in \{1, \ldots, J\}$. Investors are heterogeneous in their discount rates and asset holding costs, which can be different types of asset market participants such as pension funds, hedge funds, or dealers. Investors of type j discount the future at rate $\rho_j > 0$. An investor gets a marginal utility of 1 from a non-storable numeraire good. In order to make payments, investors are endowed with a technology that instantly produces numeraire goods, at a unit marginal cost. An investor can hold either zero or one share of at most one type of asset. When investor j holds asset k, they get flow utility $\delta - \psi(j, k)$, where $\psi(j, k)$ is a "holding cost" that reflects institutional constraints. Investors are subject to independent idiosyncratic shocks of switching between types that follow a continuous time Markov chain. Let $\lambda_{i,j}$ denote the rate of switching from type i to j and let Λ denote the matrix of switching rates.

Financial crisis risk: The aggregate state in the economy is $z \in \{z_1, \ldots, z_n\}$, which follows a continuous time Markov process where $\zeta_{z,z'}$ denotes the rate at which the process switches from z to z'. We allow the aggregate state to affect agent switching rates and haircuts. Formally, at state z, the switching rate from agent type i to agent type j is given by $\lambda_{ij}(z)$. Changes to $\lambda_{ij}(z)$ impact the fraction of agents with low or high holding costs so, as in Duffie, Gârleanu, and Pedersen (2005), we interpret these shocks as changes to the "liquidity" or "institutional" constraints in the investor population. In addition, at state z, asset k pays a fraction $\phi(k, z)$ of the coupon and the principal. We interpret $1 - \phi(k, z) > 1$ as a "haircut" on the bond.

Primary market: When bonds mature, they are replaced by new bonds in the economy. We assume there is an exogenous primary market that allocates new bonds with maturity τ_k to type j investors who are not holding assets at the rate $\xi_{j,k}$.

Distribution: An investor's idiosyncratic state is made up of her type $j \in \{1, \ldots, J\}$ and her ownership status for each asset $k \in \{1, \ldots, K\}$. Hence the set of investor

idiosyncratic states is:

$$A = \{1n, 2n, \ldots, Jn, 1o1, \ldots, 1oK, 2o1, \ldots 2oK, Jo1, \ldots, JoK\}$$

in which jn denotes an investor with type j who does not hold any asset, and jok denotes an type-j investor holding an asset k. For each $a \in A$, let $g(a)$ denote the fraction of total investors that have state a and $g = (g(a))_{a \in A}$ denote the mass function. Relative to the labor model, the support of the distribution has been expanded to account for type switching.

Meeting and Bargaining: The contact rate between investors with idiosyncratic states a and b is: $\mathcal{M}_{a,b} = \kappa_{a,b} g_a g_b$. When agents with states a and b meet, they engage in generalized Nash bargaining with bargaining power $\beta_{a,b}$ for the agent in state a.

5.2 Equilibrium

The aggregate states are (z, g). We denote the law of motion for the cross-sectional idiosyncratic state distribution, g, by the form $dg_t(a) = \mu^g(a, z, g)dt$ and agents' belief about the law of motion by $\breve{\mu}^g(a, z, g)$.

Trade and surplus division: Upon meeting, agents negotiate asset trades according to a generalized Nash Bargaining protocol. Let $V^n(in, z, g)$ denote the value function for an investor of type i without an asset, and let $V^o(iok, z, g)$ denote the value function for an investor of type i with an asset. If a type-i agent holding asset k meets a type-j agent without an asset, the total surplus from asset trade is:

$$S(iok, jn, z, g) = V^n(in, z, g) - V^o(iok, z, g) + V^o(jok, z, g) - V^n(jn, z, g)$$

The generalized Nash bargaining protocol implies that, if agents trade, the price paid for the asset, $p_k(in, jok, z, g)$, solves:

$$\begin{aligned} \Delta V_{i[ok \to n]} + p_k(in, jok, z, g) &= \beta_{iok,jn} S(iok, jn, z, g) \\ \Delta V_{j[n \to ok]} - p_k(in, jok, z, g) &= (1 - \beta_{iok,jn}) S(iok, jn, z, g) \end{aligned}$$

where $\Delta V_{i[ok \to n]}(z, g) := V^n(in, z, g) - V^o(iok, z, g)$ and $\Delta V_{j[n \to ok]}(z, g) := V^o(jok, z, g) -$

$V^n(jn, z, g)$. This implies that agents choose to trade if the surplus is positive.

Similarly, if a type-i agent holding asset k meets a type-j agent holding asset l, then the total surplus if the investors exchange assets is:

$$S(iok, jol, z, g) = V^o(iol, z, g) - V^o(iok, z, g) + V^o(jok, z, g) - V^o(jol, z, g)$$

Once again, the agents trade if the surplus is positive and the generalized Nash bargaining protocol implies that the net transfer between the agents $\Delta p_{k,l}$ satisfies:

$$\begin{aligned}\Delta V_{i[ok\to ol]} + \Delta p_{k,l}(iol, jok, z, g) &= \beta_{iok,jol} S(iok, jol, z, g)\\ \Delta V_{j[ol\to ok]} - \Delta p_{k,l}(iol, jok, z, g) &= (1 - \beta_{iok,jol}) S(iok, jol, z, g).\end{aligned}$$

Hamilton-Jacobi-Bellman Equations: Given their belief $\tilde{\mu}$, the value function for a non-owner with type i satisfies the following HJB equation:

$$\begin{aligned}\rho_i V^n(in, g, z) = &\sum_a \kappa_{in,a}\alpha(in, a, g, z)\beta_{in,a} S(in, a, z, g) \\ &+ \sum_k \xi_{i,k}(V^o(iok, g, z) - V^n(in, g, z)) + \sum_{j\neq i} \lambda_{i,j}(z)(V^n(jn, g, z) - V^n(in, g, z)) \\ &+ \sum_{z'} \zeta_{z,z'}(V^n(in, g, z') - V^n(in, g, z)) + \sum_{a\in A} \partial_{g_a} V^n(in, g, z)\breve{\mu}^g(a, z)\end{aligned} \tag{5.1}$$

where $\alpha(in, jok, g, z)$ is an indicator function for whether the trade is accepted upon matching, which in equilibrium occurs if the surplus from the trade is positive $S(in, jok, g, z) > 0$. Likewise, the value function for an investor of type i holding asset k, $V^o(iok, g, z)$, is given by the following HJB equation:

$$\begin{aligned}\rho_i V^o(iok, g, z) = &\delta\phi(k, z) - \psi(i, k) + \frac{1}{\tau_k}(V^n(in, g, z) + \phi(k, z) - V^o(iok, g, z)) \\ &+ \sum_a \kappa_{iok,a}\alpha(iok, a, g, z) g_a \beta_{iok,a} S(iok, a, g, z) + \sum_{a\in A} \partial_{g_a} V^o(iok, g, z)\breve{\mu}^g(a, z) \\ &+ \sum_{j\neq i} \lambda_{i,j}(z)(V^o(jok, g, z) - V^o(iok, g, z)) + \sum_{z'} \zeta_{z,z'}(V^o(iok, g, z') - V^o(iok, g, z)).\end{aligned} \tag{5.2}$$

Kolmogorov Forward Equation: The law of motion for the distribution of non-owner

states and owner states are given respectively by:

$$\frac{dg_{in}}{dt} = \mu^g(in, z, g) = \sum_{j\neq i} \lambda_{j,i}(z) g_{jn} + \sum_{j\neq i}\sum_{k} \kappa_{jn,iok} g_{jn} g_{iok} \alpha(jn, iok, g, z) \tag{5.3}$$
$$- \sum_{j\neq i} \lambda_{i,j}(z) g_{in} - \sum_{j\neq i}\sum_{k} \kappa_{in,jok} g_{in} g_{jok} \alpha(in, jok, g, z) + \sum_{k} \frac{1}{\tau_k} g_{iok} - \sum_{k} \xi_{i,k} g_{in}$$

$$\frac{dg_{iok}}{dt} = \mu^g(iok, z, g) = \sum_{j\neq i} \lambda_{j,i}(z) g_{jok} - \sum_{j\neq i} \kappa_{jn,iok} g_{jn} g_{iok} \alpha(jn, iok, g, z)$$
$$+ \sum_{j\neq i} \kappa_{in,jok} g_{in} g_{jok} \alpha(in, jok, g, z) - \sum_{j\neq i}\sum_{l\neq k} \kappa_{iok,jol} g_{iok} g_{jol} \alpha(iok, jol, g, z)$$
$$+ \sum_{j\neq i}\sum_{l\neq k} \kappa_{iol,jok} g_{iol} g_{jok} \alpha(iol, jok, g, z) - \sum_{j\neq i} \lambda_{i,j}(z) g_{iok} - \frac{1}{\tau_k} g_{iok} + \xi_{i,k} g_{in} \tag{5.4}$$

In equilibrium, the flows from assets maturing are equal to the flows from new assets being created so: $\sum_i \xi_{i,k} g_{in} = \frac{1}{\tau_k} \sum_i g_{iok} =: \frac{1}{\tau_k} s_k$.

Master equation: Unlike in Section 2.2.5, we cannot characterize equilibrium in the OTC market using a differential equation for surplus. Instead, we solve the two HJB equations (5.1) and (5.2) combined with belief consistency $\tilde{\mu}^g = \mu^g$ and the KF equations (5.3) and (5.4). To solve the problem with the DeepSAM method, we use neural networks to parameterize $V^n(in, g, z)$ and $V^o(iok, g, z)$, and solve the problem to minimize the weighted loss of equations (5.1) and (5.2) on the sampling data.

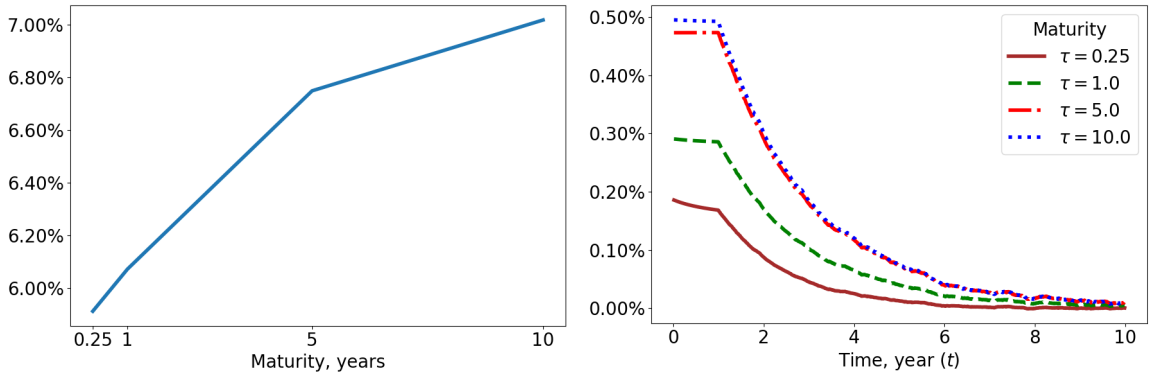
5.3 Endogenous Yield Curve and Financial Crises

We calibrate the model with four types of investors: $\{D, C, U, P\}$, where type D are dealers in the primary bond market, type C are liquidity constrained hedge funds, type U are unconstrained hedge funds, and type P are pension/insurance funds with a long investment horizon. This is reflected in their holding costs. On the asset side, we have four types of bonds with maturities $\tau = 0.25, 1, 5, 10$ years. Unconstrained hedge funds have no holding cost while liquidity-constrained hedge funds have a holding cost of 0.02 across all assets. Pension/insurance funds face holding costs of 0.02 for short maturity bonds ($\tau_1 = 0.25, 1.0$), 0.01 for bonds with $\tau = 5.0$, and no holding cost for long term bonds $\tau = 10.0$. We interpret this as reflecting regulatory constraints or financial frictions that encourage pension/insurance funds to hold long-term bonds.

We consider three aggregate states: good, normal, and bad, where the bad state

is interpreted as a financial crisis. We impose that the dealers and pension/insurance funds have constant types over time. By contrast, hedge funds switch from type U to C (i.e. become liquidity constrained) at the rate 0.3 in the good aggregate state, 0.5 in the normal aggregate state, and 0.7 in the crisis aggregate state. In all aggregate states, they switch from type C to U (i.e. become unconstrained) at the rate 0.1. We calibrate other parameters so that the ergodic yield curve matches the average high-grade corporate yield curve over the past 50 years documented by Payne and Szőke (2024) and the haircut rates in the crisis state to match Chen, Cui, He, and Milbradt (2017). We explain the calibration in more detail in Appendix E.

Figure 9a shows the ergodic mean bond yields as a function of maturity at the ergodic steady state of our economy. Evidently, longer maturity bonds have higher yields indicating an upward sloping yield curve. This shape reflects relative investor willingness to hold short and long maturity bonds in the economy. Hedge funds prefer to hold short-maturity bonds because they are worried that they will end up stuck with long-maturity bonds if they become liquidity-constrained. By contrast, the pension/insurance fund prefers to hold long-maturity bonds. Under our calibration, it is the first effect that dominates and so the yield curve is upward-sloping.

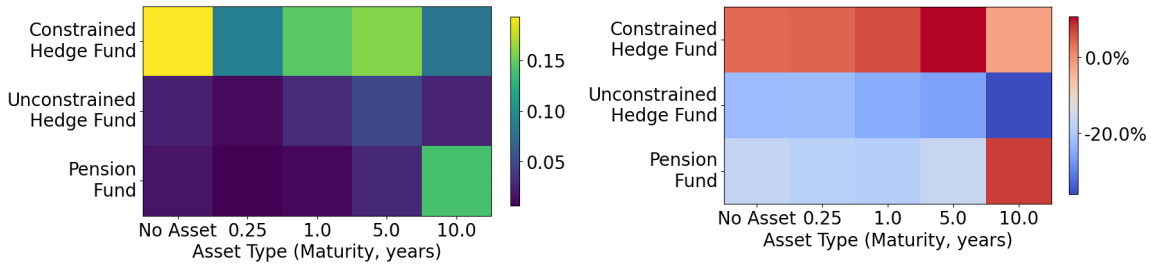


(a) Ergodic Yield Curve

(b) IRF of Yield to Maturity by Bond

Figure 9: Yield curve at ergodic steady state and impulse responses. Plot (b) shows the proportional bond yield change compared to the ergodic yield at each maturity following a one-year recession. To calculate the figures, we simulate 3000 paths and calculate the mean.

We use our model to examine the impact of a financial crisis in our OTC bond market. Specifically, the economy starts at the ergodic mean, then moves to the bad "financial crisis" state z_B for one year, and then follows the stochastic z_t process. Figure 9b shows the impulse responses for bond yields following the shock. For



(a) Ergodic Distribution g^{ergodic} (b) Change after Crisis $\frac{g(z_B)-g^{\text{ergodic}}}{g^{\text{ergodic}}}$

Figure 10: Distribution response to a crisis shock.

the short-maturity bonds, the yields move very little whereas the yields of long-maturity bonds increase significantly. Figure 10b shows the change in the distribution of investors (relative to the ergodic mean in Figure 10a) when the economy stays in the crisis state for a long time. Evidently, the crisis increases the likelihood that a hedge fund becomes constrained and so increases the proportion of liquidity-constrained hedge funds. This heightens hedge fund concern that they will end up stuck with long-maturity bonds while liquidity-constrained. Consequently, the relative demand for long-term bonds falls during the crisis leading to the relatively large yield increase for long-term bonds in Figure 9b.

Although our OTC model is highly stylized, it illustrates DeepSAM's potential for connecting different asset pricing literatures. By incorporating aggregate crisis risk, institutional heterogeneity, and bond maturities into Duffie et al. (2005), our model introduces default risk premia and term premia into the original OTC model with liquidity premia. This connects OTC models to both traditional bond default models and yield curve models with institutional investors (e.g. Vayanos and Vila (2021)) where, similar to our model, it is the variation in institutional willingness to hold the asset that generates the yield curve. We make these connections while maintaining endogenous trading patterns in our OTC market. That is, agent preferences lead to bonds with particular maturities endogenously emerging as being more or less actively traded and having more of less volatile yields. Ultimately, we believe DeepSAM opens up the possibility to study how a wide range of risks are priced in OTC environments.

6 Conclusion

In this paper, we developed a new method for globally solving and estimating search and matching models with aggregate shocks and heterogeneous agents. This allows us to study dynamics in models where agent decisions depend upon the distribution so the model is not "block-recursive". We believe our methodology is a major breakthrough in the literature of search and matching models and will open up many potential applications in the labor, finance, and spatial literature.

References

Alvarez, F., F. Lippi, and P. Souganidis (2023): "Price setting with strategic complementarities as a mean field game," *Econometrica*, 91, 2005–2039.

Alves, F. (2022): "Job ladder and business cycles," Tech. rep., Bank of Canada.

Alves, F. and G. L. Violante (2023): "Some Like It Hot: Monetary Policy Under Okun's Hypothesis," .

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

Azinovic, M. and J. Žemličkа (2023): "Economics-inspired neural networks with stabilizing homotopies," *arXiv preprint arXiv:2303.14802*.

Baley, I., A. Figueiredo, and R. Ulbricht (2022): "Mismatch cycles," *Journal of Political Economy*, 130, 2943–2984.

Bilal, A. (2023): "Solving heterogeneous agent models with the master equation," Tech. rep., National Bureau of Economic Research.

Birinci, S., F. Karahan, Y. Mercan, and K. See (2024): "Labor market shocks and monetary policy," *Working Paper*.

Burdett, K. and D. T. Mortensen (1998): "Wage differentials, employer size, and unemployment," *International Economic Review*, 257–273.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006): "Wage bargaining with on-the-job search: Theory and evidence," *Econometrica*, 74, 323–364.

Cajner, T., L. D. Crane, R. A. Decker, J. Grigsby, A. Hamins-Puertolas, E. Hurst, C. Kurz, and A. Yildirmaz (2020): "The US Labor Market during the Beginning of the Pandemic Recession," *Brookings Papers on Economic Activity*, 3–33.

Cardaliaguet, P., F. Delarue, J.-M. Lasry, and P.-L. Lions (2019): *The master equation and the convergence problem in mean field games*, Princeton University Press.

Chen, H., R. Cui, Z. He, and K. Milbradt (2017): "Quantifying Liquidity and Default Risks of Corporate Bonds over the Business Cycle," *The Review of Financial Studies*.

Chen, H., A. Didisheim, and S. Scheidegger (2023): "Deep surrogates for finance: With an application to option pricing," *Available at SSRN 3782722*.

Duarte, V., D. Duarte, and D. H. Silva (2024): "Machine learning for continuous-time finance," *The Review of Financial Studies*, 37, 3217–3271.

Duarte, V. and J. Fonseca (2024): "Global identification with gradient-based structural estimation," .

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005): "Over-the-counter markets," *Econometrica*, 73, 1815–1847.

Engbom, N. (2021): "Contagious unemployment," Tech. rep., National Bureau of Economic Research.

Fernández-Villaverde, J., S. Hurtado, and G. Nuno (2023): "Financial frictions and the wealth distribution," *Econometrica*, 91, 869–901.

Fernández-Villaverde, J., G. Nuno, and J. Perla (2024): "Taming the Curse of Dimensionality: Quantitative Economics with Deep Learning," Tech. rep.

Friedl, A., F. Kübler, S. Scheidegger, and T. Usui (2023): "Deep uncertainty quantification: with an application to integrated assessment models," .

Fukui, M. (2020): "A theory of wage rigidity and unemployment fluctuations with on-the-job search," .

Gopalakrishna, G. (2021): "Aliens and continuous time economies," Tech. rep.

Gregory, V., G. Menzio, and D. Wiczer (2024): "The alpha beta gamma of the labor market," *Journal of Monetary Economics*, 103695.

Gu, Z., M. Lauriere, S. Merkel, and J. Payne (2023): "Deep Learning Solutions to Master Equations for Continuous Time Heterogeneous Agent Macroeconomic Models," Tech. rep.

Guvenen, F., S. Schulhofer-Wohl, J. Song, and M. Yogo (2017): "Worker betas: Five facts about systematic earnings risk," *American Economic Review*, 107, 398–403.

Hagedorn, M., T. H. Law, and I. Manovskii (2017): "Identifying equilibrium models of labor market sorting," *Econometrica*, 85, 29–65.

Han, J., Y. Yang, and W. E (2021): "DeepHAM: A Global Solution Method for Heterogeneous Agent Models with Aggregate Shocks," .

Huang, J. (2022): "A Probabilistic Solution to High-Dimensional Continuous-Time Macro-Finance Models," *Available at SSRN 4122454*.

——— (2024): "Breaking the Curse of Dimensionality in Heterogeneous-Agent Models: A Deep Learning-Based Probabilistic Approach," *Available at SSRN 4649043*.

Kahou, M. E., J. Fernández-Villaverde, J. Perla, and A. Sood (2021): "Exploiting symmetry in high-dimensional dynamic programming," Tech. rep.

Kase, H., L. Melosi, and M. Rottner (2024): "Estimating Nonlinear Heterogeneous Agent Models with Neural Networks," .

Krusell, P., T. Mukoyama, R. Rogerson, and A. Şahin (2017): "Gross worker flows over the business cycle," *American Economic Review*, 107, 3447–3476.

Krusell, P. and A. A. Smith (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106, 867–896.

Lagos, R. and G. Rocheteau (2009): "Liquidity in asset markets with search frictions," *Econometrica*, 77, 403–426.

Lentz, R., J. Lise, and J.-M. Robin (2017): "The Macro Dynamics of the Wage Distribution," in *2017 Meeting Papers*, vol. 1220.

Lise, J. and J.-M. Robin (2017): "The macrodynamics of sorting between workers and firms," *American Economic Review*, 107, 1104–1135.

Maliar, L., S. Maliar, and P. Winant (2021): "Deep learning for solving dynamic economic models." *Journal of Monetary Economics*, 122, 76–101.

Menzio, G. and S. Shi (2010): "Block recursive equilibria for stochastic models of search on the job," *Journal of Economic Theory*, 1.

——— (2011): "Efficient Search on the Job and the Business Cycle," *Journal of Political Economy*, 119, 468–510.

Moen, E. R. (1997): "Competitive Search Equilibrium," *Journal of Political Economy*, 105, 385–411.

Mortensen, D. T. and C. A. Pissarides (1994): "Job creation and job destruction in the theory of unemployment," *The review of economic studies*, 61, 397–415.

Moscarini, G. and F. Postel-Vinay (2018): "The cyclical job ladder," *Annual Review of Economics*, 10, 165–188.

——— (2023): "The job ladder: Inflation vs. reallocation," Tech. rep., National Bureau of Economic Research.

Mueller, A. I. (2017): "Separations, sorting, and cyclical unemployment," *American Economic Review*, 107, 2081–2107.

Norets, A. (2012): "Estimation of dynamic discrete choice models using artificial neural network approximations," *Econometric Reviews*, 31, 84–106.

Okun, A. M. (1973): "Upward mobility in a high-pressure economy," *Brookings Papers on Economic Activity*, 1973, 207–261.

Payne, J. and B. Szőke (2024): "Convenience Yields and Financial Repression," .

Postel-Vinay, F. and J.-M. Robin (2002): "Equilibrium wage dispersion with worker and employer heterogeneity," *Econometrica*, 70, 2295–2350.

Qiu, X. (2023): "Vacant Jobs," .

Raissi, M., P. Perdikaris, and G. E. Karniadakis (2019): "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, 378, 686–707.

Sauzet, M. (2021): "Projection methods via neural networks for continuous-time models," *Available at SSRN 3981838*.

Schaal, E. (2017): "Uncertainty, Productivity and Unemployment in the Great Recession," *Econometrica*, 85, 250–274.

Shimer, R. (2005): "The cyclical behavior of equilibrium unemployment and vacancies," *American economic review*, 95, 25–49.

Shimer, R. and L. Smith (2000): "Assortative matching and search," *Econometrica*, 68, 343–369.

Smith, L. (2006): "The marriage model with search frictions," *Journal of political Economy*, 114, 1124–1144.

Vayanos, D. and J.-L. Vila (2021): "A preferred-habitat model of the term structure of interest rates," *Econometrica*, 89, 77–112.

Weill, P.-O. (2008): "Liquidity premia in dynamic bargaining markets," *Journal of Economic Theory*, 140, 66–96.

Online Appendix

A Details of the Models

Parameter	Value	Parameter	Value
Number of layers	4	Final learning rate	10^{-5}
Neurons per layer	50	Initial sample size per epoch	256
Activation function	$\tanh(\cdot)$	Final sample size per epoch	512
Initial learning rate	10^{-4}	Convergence threshold	10^{-6}

Table 6: Neural Network Parameters of the Labor Search Model in Section 3

Parameter	Interpretation	Value	Target/Source
ρ	Discount rate	0.05	Interest rate
ξ	Extreme value for α choice	2.0	
$f(x,y)$	Production for match (x,y)	$0.6 + 0.4\left(\sqrt{x} + \sqrt{y}\right)^2$	Hagedorn et al. (2017)
β	Surplus division factor	0.72	Shimer (2005)
$m(\mathcal{U},\mathcal{V})$	Matching function	$\kappa\mathcal{U}^{\nu}\mathcal{V}^{1-\nu}$	Hagedorn et al. (2017)
ν	Elasticity in meeting function	0.5	Hagedorn et al. (2017)
κ	Scale for meeting function	5.4	Unemployment rate
b	Worker unemployment benefit	0.5	Shimer (2005)
c	Entry cost	4.86	Steady state $\mathcal{V}/\mathcal{U} = 1$
Steady State:			
$\bar{z}$	Steady state TFP	1	Shimer (2005)
$\bar{\delta}$	Steady state separation rate	0.2	BLS job tenure 5 years
Exogenous Aggregate Shock Process:			
A_D, A_L, A_H	TFP levels	0.985, 0.985, 1.015	Lise and Robin (2017)
δ_L, δ_H	Separation rates	0.18, 0.22	Shimer (2005)
$\delta_D(x,y)$	TFP and separation at crisis state	See Appendix C.1	Match Cajner et al. (2020)
λ_z	Poisson transition probability	See Appendix C.1	Shimer (2005)
n_x	Discretization of worker types	5	Match Cajner et al. (2020)
n_y	Discretization of firm types	11	Match Cajner et al. (2020)

Table 7: Economic Parameters of the Labor Search Model in Section 3.

Parameter	Interpretation	Value
ρ	Discount rate	0.05
ξ, ξ^e	Extreme value for α and α^e choices	2.5, 0.02
$f(x,y)$	Production for match (x,y)	$0.6 + 0.4\left(\sqrt{x} + \sqrt{y}\right)^2$
$m(\mathcal{U}, \mathcal{V})$	Matching function	$\kappa \mathcal{U}^\nu \mathcal{V}^{1-\nu}$
ν	Elasticity in meeting function	0.5
ϕ	Relative intensity	0.075
A_L, A_H	TFP levels	0.985, 1.015
λ_z	Poisson transition probability	0.08
n_x	Discretization of worker types	7
n_y	Discretization of firm types	8

Table 8: Externally Calibrated Parameters of the OJS Model in Section 4.

B Appendix for the General Methodology

B.1 Master Equation with the Free Entry Condition

As in Section 2.2.4, the free entry condition is

$$0 = \int V_t^v(\tilde{y}) d\tilde{y}$$

Recall from (2.5) the HJB equation for a vacant institution with productivity y is

$$\rho V_t^v(y) = -c + \mathcal{M}_t^v \int \alpha(\tilde{x}, y) \frac{g_t^u(\tilde{x})}{\mathcal{U}_t} (1-\beta) S_t(\tilde{x}, y) d\tilde{x} + \partial_t V_t^v(y)$$

Integrating and combining these equations, we have that:

$$\begin{aligned}
\rho \int V_t^v(\tilde{y}) d\tilde{y} = & -c + \mathcal{M}_t^v \int \int \alpha(\tilde{x}, \tilde{y}) \frac{g_t^u(\tilde{x})}{\mathcal{U}_t} (1-\beta) S_t(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y} + \partial_t \int V_t^v(\tilde{y}) d\tilde{y} \\
\Rightarrow c = & \frac{m(\mathcal{U}_t, \mathcal{V}_t)}{\mathcal{V}_t} \int \int \alpha(\tilde{x}, \tilde{y}) \frac{g_t^u(\tilde{x})}{\mathcal{U}_t} (1-\beta) S_t(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y} \\
\Rightarrow \frac{m(\mathcal{U}_t, \mathcal{V}_t)}{\mathcal{V}_t} = & \frac{c}{\int \int \alpha(\tilde{x}, \tilde{y}) \frac{g_t^u(\tilde{x})}{\mathcal{U}_t} (1-\beta) S_t(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y}}
\end{aligned}$$

Assume $m(\mathcal{U}_t, \mathcal{V}_t)$ is homothetic[16], then $m(\mathcal{U}_t, \mathcal{V}_t)/(\mathcal{V}_t) = \widehat{m}(\mathcal{V}_t/\mathcal{U}_t)$, we have:

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

where $g_t^u = g_t^w - \int g_t(x,y)dy$ and so the RHS can be computed from g_t and S_t. Since firm y draws are uniformly distributed, we have that g_t^f is given by:

$$\begin{aligned} g_t^f &= \mathcal{V}_t + \mathcal{P}_t \\ &= \mathcal{U}_t \widehat{m}^{-1}\left(\frac{c}{\int\int \alpha(\tilde{x},\tilde{y})(g_t^u(\tilde{x})/\mathcal{U}_t)(1-\beta)S_t(\tilde{x},\tilde{y})d\tilde{x}d\tilde{y}}\right) + \int\int g_t(x,y)dydx \end{aligned} \tag{B.2}$$

where $\mathcal{V}_t$ is from (B.1), $\mathcal{P}_t = \int\int g_t(x,y)dydx$, and $\mathcal{U}_t = \int(g_t^w(x) - \int g_t(x,y)dy)$. This means that g_t^f can be computed from g_t and S_t, and

$$\begin{aligned} g_t^v(y) &= g_t^f(y) - g_t^p(y) \\ &= \mathcal{U}_t \widehat{m}^{-1}\left(\frac{c}{\int\int \alpha(\tilde{x},\tilde{y})(g_t^u(\tilde{x})/\mathcal{U}_t)(1-\beta)S_t(\tilde{x},\tilde{y})d\tilde{x}d\tilde{y}}\right) + \int\int g_t(x,y)dydx \\ &\quad - \int g_t(x,y)dx \end{aligned}$$

The master equation for surplus:

$$\begin{aligned} \rho S_t(x,y) &= \rho(V_t^p(x,y) - V_t^v(y) + V_t^e(x,y) - V_t^u(x)) \\ &= f_t(x,y) - w_t(x,y) - \delta(1-\beta)S_t(x,y) + \partial_t V_t^p(x,y) \\ &\quad - \left(\mathcal{M}_t^v \int \alpha(\tilde{x},y)\frac{g_t^u(\tilde{x})}{\mathcal{U}_t}(1-\beta)S_t(\tilde{x},y)d\tilde{x} + \partial_t V_t^v(y)\right) \\ &\quad + w_t(x,y) - \beta\delta S_t(x,y) + \partial_t V_t^e(x,y) \\ &\quad - \left(b + \mathcal{M}_t^u \int \alpha_t(x,\tilde{y})\frac{g_t^v(\tilde{y})}{V_t}\beta S_t(x,\tilde{y})d\tilde{y} + \partial_t V_t^u(x)\right) \\ &= f_t(x,y) - \delta S_t(x,y) - \mathcal{M}_t^v \int \alpha(\tilde{x},y)\frac{g_t^u(\tilde{x})}{\mathcal{U}_t}(1-\beta)S_t(\tilde{x},y)d\tilde{x} \\ &\quad - b - \mathcal{M}_t^u \int \alpha_t(x,\tilde{y})\frac{g_t^v(\tilde{y})}{\mathcal{V}_t}\beta S_t(x,\tilde{y})d\tilde{y} + \partial_t S_t(x,y) \end{aligned}$$

[16]For example, if $m(\mathcal{U},\mathcal{V}) = \kappa\mathcal{U}^\nu\mathcal{V}^{1-\nu}$, then $\mathcal{M}_t^v = m(\mathcal{U},\mathcal{V})/\mathcal{V} = \widehat{m}(\mathcal{V}_t/\mathcal{U}_t) = \kappa(\mathcal{U}/\mathcal{V})^\nu$ and $\mathcal{M}_t^u = m(\mathcal{U},\mathcal{V})/\mathcal{U} = \kappa(\mathcal{V}/\mathcal{U})^{1-\nu}$.

where:

$$
\begin{aligned}
\mathcal{V}_t &= \mathcal{U}_t \widehat{m}^{-1}\left(\frac{c}{\int\int \alpha(\tilde{x},\tilde{y})\frac{g_t^u(\tilde{x})}{\mathcal{U}_t}(1-\beta)S_t(\tilde{x},\tilde{y})d\tilde{x}d\tilde{y}}\right)\\
\mathcal{M}_t^v &= m(\mathcal{U},\mathcal{V})/\mathcal{V} = \widehat{m}(\mathcal{V}_t/\mathcal{U}_t) = \kappa(\mathcal{U}/\mathcal{V})^{\nu},\\
\mathcal{M}_t^u &= m(\mathcal{U},\mathcal{V})/\mathcal{U} = \kappa(\mathcal{V}/\mathcal{U})^{1-\nu}\\
g_t^v(y) &= g_t^f(y) - g_t^p(y) = \mathcal{V}_t + \mathcal{P}_t - g_t^p(y)\\
&= \mathcal{U}_t \widehat{m}^{-1}\left(\frac{c}{\int\int \alpha(\tilde{x},\tilde{y})(g_t^u(\tilde{x})/\mathcal{U}_t)(1-\beta)S_t(\tilde{x},\tilde{y})d\tilde{x}d\tilde{y}}\right) + \int\int g_t(x,y)dydx - \int g_t(x,y)dx
\end{aligned}
$$

The KF equation is in the same form as (2.7) with $g^f(y)$ and $\mathcal{V}$ coming from equation (B.2).

B.2 PDEs for Solving Value and Wage Functions

In this subsection, we present the details of Algorithm 1 to solve for value and wage functions after we obtained the solution of surplus and acceptance functions. Once we have $S(x,y,z,\underline{\boldsymbol{g}})$ and $\alpha(x,y,z,\underline{\boldsymbol{g}})$, we can then solve $V^u(x,z,\underline{\boldsymbol{g}})$ and $V^v(x,z,\underline{\boldsymbol{g}})$ by solving the following equations:

$$
\begin{aligned}
0 = \mathcal{L}^{V^u}V^u = & -\rho V^u(x,z,\underline{\boldsymbol{g}}) - c + b\\
&+ \mathcal{M}^u(z,\underline{\boldsymbol{g}})\frac{1}{n_y}\sum_{j=1}^{n_y}\alpha(x,y_j,z,\underline{\boldsymbol{g}})\beta S(x,y_j,z,\underline{\boldsymbol{g}})\frac{g^v(y_j)}{V(z,\underline{\boldsymbol{g}})}\\
&+ \sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\partial_{g_{ij}}V^u(x,y,z,\underline{\boldsymbol{g}})\mu^g(x_i,y_j,z,\underline{\boldsymbol{g}})\\
&+ \lambda(z)(V^u(x,\tilde{z},\underline{\boldsymbol{g}}) - V^u(x,z,\underline{\boldsymbol{g}})) \qquad \text{(B.3)}\\
0 = \mathcal{L}^{V^v}V^v = & -\rho V^v(y,z,\underline{\boldsymbol{g}})\\
&+ \mathcal{M}^v(z,\underline{\boldsymbol{g}})\frac{1}{n_x}\sum_{i=1}^{n_x}\alpha(x_i,y,z,\underline{\boldsymbol{g}})(1-\beta)S(x_i,y,z,\underline{\boldsymbol{g}})\frac{g^u(x_i)}{U(z,\underline{\boldsymbol{g}})}\\
&+ \sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\partial_{g_{ij}}V^v(x,y,z,\underline{\boldsymbol{g}})\mu^g(x_i,y_j,z,\underline{\boldsymbol{g}})\\
&+ \lambda(z)(V^v(x,\tilde{z},\underline{\boldsymbol{g}}) - V^v(x,z,\underline{\boldsymbol{g}})) \qquad \text{(B.4)}
\end{aligned}
$$

Equations (B.3) (B.4) are also high dimensional PDEs, and we use NN to parameterize the value functions as $\widehat{V}^u(x,z,g;\boldsymbol{\Theta^u})$, $\widehat{V}^v(y,z,g;\boldsymbol{\Theta^v})$ with NN parameters $\boldsymbol{\Theta^u},\boldsymbol{\Theta^v}$, and solve them using deep learning similar to Steps 1-3 in Algorithm 1.

After solving $V^u(x,z,\underline{\boldsymbol{g}})$ and $V^v(x,z,\underline{\boldsymbol{g}})$, we can obtain $V^e(x,y,z,\underline{\boldsymbol{g}})$ and $V^p(x,y,z,\underline{\boldsymbol{g}})$ directly with

$$\begin{aligned}V^e(x,y,z,\underline{\boldsymbol{g}}) &= \beta S(x,y,z,\underline{\boldsymbol{g}}) + V^u(x,z,\underline{\boldsymbol{g}})\\ V^p(x,y,z,\underline{\boldsymbol{g}}) &= (1-\beta) S(x,y,z,\underline{\boldsymbol{g}}) + V^v(y,z,\underline{\boldsymbol{g}})\end{aligned}$$

Finally, we calculate the wage from:

$$\begin{aligned}w(x,y,z,\underline{\boldsymbol{g}}) = \rho(1-\beta)V^u(x,z,\underline{\boldsymbol{g}})&\\ + \beta(zf(x,y) + \sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\partial_{g_{ij}}V^p(x,y,z,\underline{\boldsymbol{g}})\mu^g(x_i,y_j,z,\underline{\boldsymbol{g}}) - \rho V^v(y,z,\underline{\boldsymbol{g}}))&\\ + (\beta-1)\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\partial_{g_{ij}}V^e(x,y,z,\underline{\boldsymbol{g}})\mu^g(x_i,y_j,z,\underline{\boldsymbol{g}})&\end{aligned}$$

C Appendix For the Labor Search Model in Section 3

C.1 Calibration Details

The Poisson transition rate for aggregate shocks across high, low, and disaster states are:

$$\lambda_z = \begin{bmatrix} - & \lambda_{HL} & \lambda_{HD} \\ \lambda_{LH} & - & \lambda_{LD} \\ \lambda_{DH} & \lambda_{DL} & - \end{bmatrix} = \begin{bmatrix} - & 0.4 & 0.001 \\ 0.4 & - & 0.001 \\ 0.0995 & 0.0995 & - \end{bmatrix}$$

The calibrated separation rate across worker and firm types $\delta_D(x,y) =$

5.2834	4.3853	3.7621	3.3420	3.0671	2.8927	2.7848	2.7191	2.6787	2.6534	2.6370
3.3734	2.5752	2.0345	1.6818	1.4614	1.3303	1.2565	1.2169	1.1963	1.1855	1.1794
2.6337	1.9001	1.4115	1.1002	0.9121	0.8058	0.7505	0.7245	0.7136	0.7094	0.7077
2.3878	1.6936	1.2358	0.9478	0.7773	0.6837	0.6374	0.6175	0.6106	0.6087	0.6084
2.3072	1.6352	1.1938	0.9178	0.7555	0.6676	0.6249	0.6072	0.6014	0.6001	0.6000

We calibrate $\delta_D(x, y)$ to match the observed peak declines in employment levels during the COVID-19 recession calculated in Cajner et al. (2020). Using detailed data from a major US payroll company, Cajner et al. (2020, Figure 4) estimates the employment drop of workers in different wage quintiles, which corresponds to five worker groups in our calibration. They also estimate the employment drop for 15 two-digit NAICS industries (Cajner et al., 2020, Table 1). We use the industry-level output and hours worked data from the Bureau of Labor Statistics (BLS)[17] to compute the labor productivity of each industry in 2020. We then consolidate smaller industries to get 11 "composite" sectors, ensuring that after merging, these sectors exhibit similar levels of sectoral output. These sectors are mapped into the 11 groups of firms in our model. We calibrate $\delta_D(x, y)$ such that the model's simulated declines in employment for these heterogeneous groups match these empirical moments after the disaster shock z_D hits the ergodic state of the economy for $t = 0.2$ years.

C.2 Master equation and loss function for the model without aggregate shocks

To verify the accuracy of the DeepSAM method, we apply it to solve a labor search model without aggregate shocks, which can also be solved with a conventional numerical method such as that in Hagedorn, Law, and Manovskii (2017). The master equation and loss function for the Surplus is given by:

$$
\begin{aligned}
0 = \mathcal{L}^S S = & -(\rho + \delta(x, y))S(x, y, \underline{\boldsymbol{g}}) + F(x, y) - b \\
& - (1 - \beta)\frac{m(\underline{\boldsymbol{g}})}{\mathcal{U}(\underline{\boldsymbol{g}})\mathcal{V}(\underline{\boldsymbol{g}})}\frac{1}{n_x}\sum_{i=1}^{n_x}\alpha(x_i, y, \underline{\boldsymbol{g}})S(x_i, y, g)\underline{\boldsymbol{g}}^u(x_i) \\
& - \beta\frac{m(\underline{\boldsymbol{g}})}{\mathcal{U}(\underline{\boldsymbol{g}})\mathcal{V}(\underline{\boldsymbol{g}})}\frac{1}{n_y}\sum_{j=1}^{n_y}\alpha(x, y_j, \underline{\boldsymbol{g}})S(x, y_j, \underline{\boldsymbol{g}})\underline{\boldsymbol{g}}^v(y_j) \\
& + \sum_{i=1}^{n_x}\sum_{i=1}^{n_y}\partial_{g_{ij}}S(x, y, \underline{\boldsymbol{g}})\mu^g(x_i, y_j, \underline{\boldsymbol{g}})
\end{aligned}
$$

[17] Data link https://www.bls.gov/productivity/tables/major-industry-total-factor-productivity-klems.xlsx. Last accessed on September 23, 2024.

in which

$$dg_t(x,y)/dt = \mu^g(x,y,\underline{\boldsymbol{g}}_t) = -\delta(x,y)g_t(x,y) + \frac{m(\underline{\boldsymbol{g}}_t)}{\mathcal{U}_t\mathcal{V}_t}\alpha(x,y,\underline{\boldsymbol{g}}_t) \times \left(g^w(x) - \frac{1}{n_y}\sum_{j=1}^{n_y}\underline{\boldsymbol{g}}_t(x,y_j)\right)\left(g^f(y) - \frac{1}{n_x}\sum_{i=1}^{n_x}\underline{\boldsymbol{g}}_t(x_i,y)\right)$$

and $\alpha(x,y,g)$ is given by:

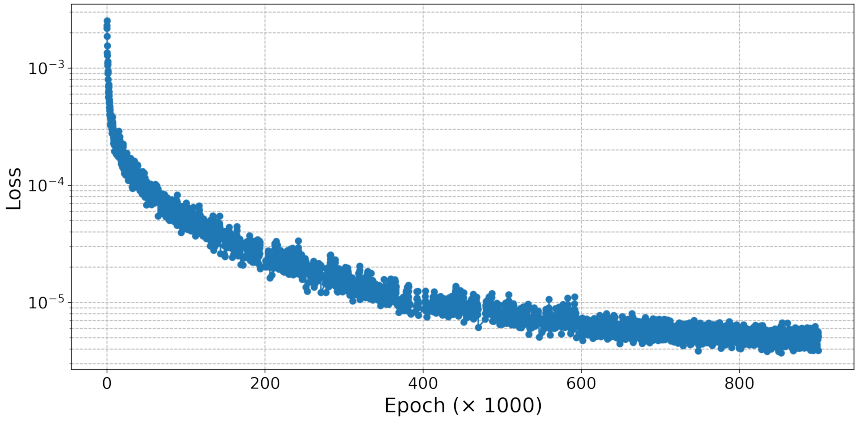
$$\alpha(x,y,\underline{\boldsymbol{g}}) = \left(1 + e^{-\xi S(x,y,\underline{\boldsymbol{g}})}\right)^{-1}.$$

C.3 Numerical method and performance

C.3.1 Hyperparameters for the neural networks

Training loss, learning rate, and sample size. Figure 11 presents the value of the loss function (2.11) along the training process. It takes 1.5 hours on an A100 GPU for the neural network to converge to a stable solution. The learning rate is 10^{-4} for the first 400,000 epoch, is 10^{-5} after that. Sample size: 256 in first 400k, 512 from after that. We use a cosine scheme to adjust the learning rate over time.

Figure 11: Loss function along training epochs



D Additional Details For The On-The-Job Search Model in Section 4

This Appendix completes the setup for the on-the-job search model in Section 4.

D.1 Equilibrium Description

Master Equation For Surplus: Under belief consistency, the differential equation for the surplus is the following:

$$
\begin{aligned}
&\rho S(x,y,z,g) = \rho(V^p(x,y,z,g) - V^v(y,z,g) + V^e(x,y,z,g) - V^u(x,z,g)) \\
&= F(x,y,z,g) - w(x,y,z,g) - (\delta + \eta\alpha^b(x,y,z,g))(1-\beta)S(x,y,z,g) \\
&+ \langle D_g V^p, \mu^g\rangle - \Big(\mathcal{M}^v\mathcal{C}^u \int \alpha(\tilde{x},y,z,g)(1-\beta)S(\tilde{x},y,z,g)\frac{g^u(\tilde{x})}{\mathcal{U}}d\tilde{x} + \langle D_g V^v, \mu^g\rangle \\
&+ \mathcal{M}^v\mathcal{C}^e \int \alpha^p(y,\tilde{x},\tilde{y},z,g)\frac{g(\tilde{x},\tilde{y})}{\mathcal{E}}(1-\beta)(S(\tilde{x},y,z,g) - S(\tilde{x},\tilde{y},z,g))d\tilde{x}d\tilde{y}\Big) \\
&+ w(x,y,z,g) + \mathcal{M}^e \int \alpha^e(x,y,\tilde{y},z,g)\beta(S(x,\tilde{y},z,g) - S(x,y))\frac{g^v(\tilde{y})}{\mathcal{V}}d\tilde{y} \\
&- \beta(\delta + \eta\alpha^b(x,y,z,g))S(x,y,z,g) + \langle D_g V^e, \mu^g\rangle \\
&- \Big(b + \mathcal{M}^u \int \alpha(x,\tilde{y},z,g)\beta S(x,\tilde{y},z,g)\frac{g^v(\tilde{y})}{\mathcal{V}}d\tilde{y}\Big) \\
&= F(x,y,z) - (\delta + \eta\alpha^b(x,y,z,g))S(x,y,z,g) - b \\
&- \mathcal{M}^v\mathcal{C}^u \int \alpha(\tilde{x},y,z,g)(1-\beta)S(\tilde{x},y,z,g)\frac{g^u(\tilde{x})}{\mathcal{U}}d\tilde{x} \\
&- \mathcal{M}^v\mathcal{C}^e \int \alpha^p(y,\tilde{x},\tilde{y},z,g)\frac{g(\tilde{x},\tilde{y},z,g)}{\mathcal{E}}(1-\beta)(S(\tilde{x},y,z,g) - S(\tilde{x},\tilde{y},z,g))d\tilde{x}d\tilde{y} \\
&+ \mathcal{M}^e \int \alpha^e(x,y,\tilde{y},z,g)\beta(S(x,\tilde{y},z,g) - S(x,y,z,g))\frac{g^v(\tilde{y})}{\mathcal{V}}d\tilde{y} \\
&- \mathcal{M}^u \int \alpha(x,\tilde{y},z,g)\beta S(x,\tilde{y},z,g)\frac{g^v(\tilde{y})}{\mathcal{V}}d\tilde{y} \\
&+ \lambda(z)(S(x,y,\tilde{z},z,g) - S(x,y,z,g)) + \langle D_g S, \mu^g\rangle
\end{aligned}
$$

where:

$$\alpha(x,\tilde{y},z,g) := \begin{cases} 1, & \text{if } S(x,\tilde{y},z,g) > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\alpha^b(x,\tilde{y}) := \begin{cases} 1, & \text{if } S(x,\tilde{y},z,g) < 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\alpha^e(x,y,\tilde{y},z,g) := \begin{cases} 1, & \text{if } S(x,\tilde{y},z,g) \geq S(x,y,z,g) \text{ and } S(x,\tilde{y},z,g) \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\alpha^p(y,\tilde{x},\tilde{y},z,g) := \begin{cases} 1, & \text{if } S(\tilde{x},y,z,g) \geq S(\tilde{x},\tilde{y},z,g) \text{ and } S(\tilde{x},y,z,g) \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Observe that:

$$\frac{\mathcal{M}^v\mathcal{C}^u}{\mathcal{U}} = \frac{m(\mathcal{W},\mathcal{V})}{\mathcal{W}\mathcal{V}}, \qquad \frac{\mathcal{M}^v\mathcal{C}^e}{\mathcal{E}} = \phi\frac{m(\mathcal{W},\mathcal{V})}{\mathcal{W}\mathcal{V}},$$
$$\frac{\mathcal{M}^e}{\mathcal{V}} = \phi\frac{m(\mathcal{W},\mathcal{V})}{\mathcal{W}\mathcal{V}}, \qquad \frac{\mathcal{M}^u}{\mathcal{V}} = \frac{m(\mathcal{W},\mathcal{V})}{\mathcal{W}\mathcal{V}}$$

and so the surplus equation becomes:

$$\begin{aligned}
\rho S(x,y,z,g) &= F(x,y,z,g) - (\delta + \eta\alpha^b(x,y,z,g))S(x,y,z,g) - b \\
&- (1-\beta)\frac{m(\mathcal{W},\mathcal{V})}{\mathcal{W}\mathcal{V}}\int \alpha(\tilde{x},y,z,g)S(\tilde{x},y,z,g)g^u(\tilde{x})d\tilde{x} \\
&- (1-\beta)\phi\frac{m(\mathcal{W},\mathcal{V})}{\mathcal{W}\mathcal{V}}\int \alpha^p(y,\tilde{x},\tilde{y},z,g)g(\tilde{x},\tilde{y})(S(\tilde{x},y,z,g) - S(\tilde{x},\tilde{y},z,g))d\tilde{x}d\tilde{y} \\
&+ \beta\phi\frac{m(\mathcal{W},\mathcal{V})}{\mathcal{W}\mathcal{V}}\int \alpha^e(x,y,\tilde{y},z,g)(S(x,\tilde{y},z,g) - S(x,y,z,g))g^v(\tilde{y})d\tilde{y} \\
&- \beta\frac{m(\mathcal{W},\mathcal{V})}{\mathcal{W}\mathcal{V}}\int \alpha(x,\tilde{y},z,g)S(x,\tilde{y},z,g)g^v(\tilde{y})d\tilde{y} \\
&+ \lambda(z)(S(x,y,\tilde{z},z,g) - S(x,y,z,g)) + \langle D_g S, \tilde{\mu}^g\rangle
\end{aligned}$$

Kolmogorov Forward Equation: The measure of matches evolves according to:

$$\begin{aligned} dg_t(x,y) = & -(\delta + \eta\alpha_t^b(x,y))g_t(x,y)dt \\ & - \underbrace{\mathcal{M}_t^e}_{\substack{\text{Rate. e-worker} \\ \text{meets}}} \underbrace{\int \underbrace{\alpha_t^e(x,y,\tilde{y})}_{\text{Prob. accept}} \underbrace{\frac{g_t^v(\tilde{y})}{\mathcal{V}_t}}_{\text{Prob. meet } \tilde{y}} d\tilde{y} \; \underbrace{g_t(x,y)}_{\text{Mass at } (x,y)}}_{\text{Mass accepting match}} dt \\ & + \underbrace{\mathcal{M}_t^u}_{\substack{\text{Rate. u-worker} \\ \text{meets}}} \alpha_t(x,y)\frac{g_t^v(y)}{\mathcal{V}_t}g_t^u(x)dt \\ & + \mathcal{M}_t^e \int \alpha_t^e(x,\tilde{y},y)\frac{g_t^v(y)}{\mathcal{V}_t}g_t(x,\tilde{y})d\tilde{y}dt \end{aligned}$$

where this KFE has been written from the perspective of the workers (it can be written equivalently from the point of view of the firms) and each term is written as:

$$\text{(Prob. worker meets)} \times \text{(Prob. acceptance)} \times \text{(prob. y)} \times \text{(mass of workers)}$$

The first term on the RHS is the exit rate due to exogenous separations, the second term is the exit rate due to workers finding better matches, the third term is new matches from unemployed workers finding jobs, and the final term is employed workers moving to (x,y). Observe that:

$$\frac{\mathcal{M}_t^e}{\mathcal{V}_t} = \phi\frac{m(\mathcal{W}_t,\mathcal{V}_t)}{\mathcal{W}_t\mathcal{V}_t} \qquad\qquad \frac{\mathcal{M}_t^u}{\mathcal{V}_t} = \frac{m(\mathcal{W}_t,\mathcal{V}_t)}{\mathcal{W}_t\mathcal{V}_t}$$

So, the KFE becomes:

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

If we know the measure of matches, then we can recover the other distribution:

$$g_t^e(x) = \int g_t(x,y)dy, \qquad g_t^u(x) = g_t^w(x) - \int g_t(x,y)dy,$$
$$g_t^p(y) = \int g_t(x,y)dx \qquad g_t^v(y) = g_t^f(y) - \int g_t(x,y)dx,$$
$$\mathcal{U}_t = \int g_t^u(x)dx, \qquad \mathcal{V}_t = \int g_t^v(y)dy$$
$$\mathcal{E}_t = \int g_t^e(x)dx, \qquad \mathcal{P}_t = \int g_t^p(y)dy$$

Discretization: For solving numerically, we discretize g and smooth α, α^b, α^e, and α^p in an analogous way to Subsection 2.3.

D.2 Relating to Block Recursivity as in Lise and Robin (2017)

In this section, we show what changes are required in our environment to get the block recursive results in Lise and Robin (2017).

D.2.1 Environment

We make the following changes to the environment from subsection 3.1.

Setting: The economy is populated by a continuum of infinitely lived workers indexed by ability x, and a continuum of firms indexed by technology y. The total measures of workers and firms are fixed and normalized to 1 and their densities are given by $g_t^w(x)$ and $g_t^f(y)$. However, now firms can post v job opportunities at exogenous cost $c(v)$. The aggregate state of the economy is indexed by z_t. At the beginning of each period, the aggregate state changes from z to z' at Poisson rate $\pi(z,z')$.

Meeting Technology: The total effective search effort is $\mathcal{W}_t = \mathcal{U}_t + \phi\mathcal{E}_t$. Let $v_t(y)$ denote the measure of type y job opportunities chosen by firm y. Let $V_t = \int v_t(y)dy$ denote the aggregate number of job opportunities. The total measure of meetings at time t is given by $M_t = m(\mathcal{W}_t, V_t)$. Define $\mathcal{M}_t^u := M_t/\mathcal{W}_t$ as the rate at which an unemployed searcher contacts a vacancy, and $\mathcal{M}_t^e = \phi\mathcal{M}_t^u$ is the rate at which an employed searcher contacts a vacancy in period t. Let $\mathcal{M}_t^v := M_t/V_t$ denote the rate per unit of recruiting effort $v_t(y)$ that a firm contacts any searching worker.

D.2.2 Value and Surplus Functions

Value of Unemployment: Let $V_t^u(x)$ denote the value of unemployment to a type x worker at t. Let $V_{0,t}^e(x,y)$ be the value to a type x worker who is hired from unemployment by a firm of type y. Lise and Robin (2017) assume the worker has no bargaining power so $V_{0,t}^e(x,y) = B_t(x)$ for all y. So, the HJB equation for the unemployed worker is:

$$\begin{aligned}\rho V_t^u(x) &= b(x) + \mathcal{M}_t^u \int \alpha_t(x,\tilde{y})(V_{0,t}^e(x,\tilde{y}) - V_t^u(x))\frac{v_t(\tilde{y})}{\mathcal{V}_t}d\tilde{y} + \partial_t V_t^u(x) \\ &= b(x) + \partial_t V_t^u(x)\end{aligned}$$

where $\alpha_t(x,y) = 0.5$. In recursive form, we have $V_t^u(x) = V^u(x,z_t)$ and so the HJB equation becomes:

$$\rho V_t^u(x,z_t) = b(x) + \sum_{z'} \pi(z,z')(V^u(x,z) - V^u(x,z))$$

Value and Surplus of a Match: Let $P_t(x,y)$ denote the present value of a match (x,y), including the continuation values to the worker and the firm upon separation (which in our notation would be $P_t(x,y) = V_t^e(x,y) + V_t^p(x,y)$.) Let $\tilde{P}_t(x,y,\tilde{y})$ denote the value to the incumbent firm and the worker after the worker moves to a new firm of type $\tilde{y}$. Then, $P_t(x,y)$ solves the HJB equation:

$$\begin{aligned}\rho P_t(x,y) = F(x,y,z_t) &+ \phi\mathcal{M}_t^u \int \alpha_t^e(x,y,\tilde{y})(\tilde{P}_t(x,y,\tilde{y}) - P_t(x,y))\frac{v_t(\tilde{y})}{V_t}d\tilde{y} \\ &+ (\delta + \eta\alpha^b(x,y))(V_t^u(x) - P_t(x,y)) + \partial_t P_t(x,y)\end{aligned}$$

If $P_t(x,\tilde{y}) > P_t(x,y)$, the worker moves and get the incumbent firm's value. So, after the move the incumbent firm has zero value and the incumbent worker has value $P_t(x,y)$, which implies

$$\tilde{P}_t(x,y,\tilde{y}) = 0 + P_t(x,y)$$

If $P_t(x,\tilde{y}) < P_t(x,y)$, the worker does not move and gets $P_t(x,\tilde{y})$. This redistributes surplus towards the worker but does not change the overall value to workers and firms combined, which implies that $\tilde{P}_t(x,y,\tilde{y}) = P_t(x,y)$.

In summary, the second term in the HJB equation is always zero, so we get:

$$\rho P_t(x,y) = F(x,y,z_t) + (\delta + \eta\alpha^b(x,y))(V_t^u(x) - P_t(x,y)) + \partial_t P_t(x,y)$$

Consider the surplus, which is defined as:

$$S_t(x,y) := P_t(x,y) - V_t^v(y) - V_t^u(x)$$

In Lise and Robin (2017), their free entry condition implies that $V_t^v(y) = 0$ for all y, so

$$S_t(x,y) = P_t(x,y) - V_t^u(x)$$

Putting the HJBs together, we have

$$\begin{aligned}\rho S_t(x,y) &= \rho(P_t(x,y) - V_t^u(x)) \\ &= F(x,y,z_t) + (\delta + \eta\alpha^b(x,y))(V_t^u(x) - P_t(x,y)) + \partial_t P_t(x,y) \\ &\quad - b(x) - \partial_t V_t^u(x) \\ &= F(x,y,z_t) - b(x) - (\delta + \eta\alpha^b(x,y))S_t(x,y) + \partial_t S_t(x,y)\end{aligned} \tag{D.1}$$

Equation (D.1) does not depend upon g and so the surplus is "block recursive"—it can be solved without knowing the distribution. This means that, in recursive form, we have the surplus $S(x,y,z)$ satisfies:

$$\begin{aligned}\rho S(x,y,z) = F(x,y,z) - b(x) - (\delta + \eta\alpha^b(x,y))S(x,y,z) \\ + \sum_{\tilde{z}} \pi(z,\tilde{z})(S(x,y,\tilde{z}) - S(x,y,z))\end{aligned}$$

KFE: The distribution of matches evolves according to:

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

where the $\alpha_t(x,y)$ are calculated from the surplus terms as in the main text.

Vacancies: Vacancies are pinned down by the cost of creation via:

$$c'[g_t^v(y)] = \mathcal{M}_t^v J(y,z)$$

and where $\mathcal{M}_t^v = \frac{m(\mathcal{W}_t,\mathcal{V}_t)}{\mathcal{V}_t}$ and $\mathcal{V}_t = \int g_t^v(y)dy$ and:

$$\begin{aligned} J(y,z) = & \int \frac{g_t^u(\tilde{x})}{\mathcal{W}_t}\alpha^b(\tilde{x},y,z)S(\tilde{x},y,z)d\tilde{x} \\ & + \phi\int\int \frac{g(\tilde{x},\tilde{y})}{\mathcal{W}_t}\alpha^e(\tilde{x},\tilde{y},y,z)(S(\tilde{x},y,z) - S(\tilde{x},\tilde{y},z))d\tilde{x}d\tilde{y} \end{aligned}$$

E Additional Details For The OTC Model in Section 5

E.1 Numerical Illustration

We now consider a calibration of the model that draws on Weill (2008), Chen et al. (2017), Payne and Szőke (2024), and incorporates our agent and asset specification.

Economic parameters: We consider an environment with four types of agents: $\{D, C, U, P\}$, where type D are interpreted as dealers in the primary bond market, type C are interpreted as liquidity-constrained hedge funds, type U are unconstrained hedge funds, and type P are pension/insurance funds with a long investment horizon. Formally, the matrices for holding costs, $\psi(i,\tau)$, switching rates, $\lambda_{ij}(z)$, and primary

market participation, $\xi(i,\tau)$, are given in Tables 9, 10, and 11 respectively. The dealer agents (type D) are the only agents who are assigned assets in the primary market. They do not get a net benefit from holding the asset but instead only from trading the asset. The hedge funds randomly switch between getting net benefit from holding any asset (type U) and getting net loss from holding all assets (type C). In this sense, they face the risk of becoming "liquidity constrained" and highly incentivized to sell assets. The pension/insurance funds face a penalty for holding short-maturity assets, interpreted as a regulatory constraint on short-asset exposure.

		Maturity (τ)			
		$\tau_1 = 0.25$	$\tau_2 = 1.0$	$\tau_3 = 5$	$\tau_4 = 10$
Agent Type (i)	D	$\delta\phi(1,z)$	$\delta\phi(2,z)$	$\delta\phi(3,z)$	$\delta\phi(4,z)$
	C	0.02	0.02	0.02	0.02
	U	0.0	0.0	0.0	0.0
	P	0.02	0.02	0.01	0.00

Table 9: Holding costs: $\psi(i,\tau)$.

Agent Type (j)

	D	C	U	P
D				
C			0.1	
U		0.7		
P				

(a) $\lambda(i,j)$ for $z = z_L$.

Agent Type (j)

	D	C	U	P
D				
C			0.1	
U		0.5		
P				

(b) $\lambda(i,j)$ for $z = z_M$.

Agent Type (j)

	D	C	U	P
D				
C			0.1	
U		0.3		
P				

(c) $\lambda(i,j)$ for $z = z_H$.

Table 10: Switching rates $\lambda(i,j)$ across different aggregate states.

		Maturity (τ)			
		$\tau_1 = 0.25$	$\tau_2 = 1.0$	$\tau_3 = 5$	$\tau_4 = 10$
Agent Type (i)	D	ξ_1	ξ_2	ξ_3	ξ_4
	C	–	–	–	–
	U	–	–	–	–
	P	–	–	–	–

Table 11: Primary market participation: $\xi(i, \tau)$.

We consider the following matching rates, which specify that agents can trade more quickly with the dealers than with each other (following Chen et al. (2017)):

$$\kappa_{a,b} = \begin{cases} 50, & \text{if } (a,b) = (in, jok) \text{ and } i, j \neq A, \\ 50, & \text{if } (a,b) = (iok, jok) \text{ and } i, j \neq A, \\ 75, & \text{if } (a,b) = (in, Aok) \text{ and } i \neq A, \\ 0, & \text{if } (a,b) = (iok, Aol) \text{ and } \forall i, \\ 0, & \text{if } (a,b) = (in, jn) \text{ and } \forall i, j, \end{cases}$$

We impose that agents have equal bargaining power unless they match with a dealer, in which case they have bargaining power 0.05 (following Weill (2008) and Chen et al. (2017)):

$$\beta_{a,b} = \begin{cases} 0.5, & \text{if } (a,b) = (in, jok) \text{ and } i, j \neq A, \\ 0.5, & \text{if } (a,b) = (iok, jol) \text{ and } i, j \neq A, \\ 0.05, & \text{if } (a,b) = (in, Aok) \text{ and } i, j \neq A, \end{cases}$$

The other economic parameters are listed in Table 12. We calibrate the model at the annual frequency. Where possible, we take standard parameters from the literature.

Neural network parameters: We describe the details of the neural network approximation and sampling in Table 13. We use a fully connected feed-forward network with 8 layers, 100 neurons per layer, and a combination of GELU($\cdot$) activation functions. The training loss is shown in Figure 12.

Parameter	Interpretation	Value	Target/Source
ρ	Discount rate	0.05	Average short rate
δ	Bond Coupon Rate	0.01	
Aggregate State: $z \in \{z_B, z_N, z_G\}$			
$\phi(z)$	Coupon haircut	$(0.986, 0.991, 0.997)$	Chen et al. (2017)
$\pi(z)$	Principal haircut	$(0.986, 0.991, 0.997)$	Chen et al. (2017)
$\zeta_{N,B}$, $\zeta_{N,G}$	Transition rate: normal to bad/good	0.1	Crisis every 10 years
$\zeta_{B,N}$, $\zeta_{G,N}$	Transition rate: bad/good to normal	0.5	Average crisis duration 2 years

Table 12: Economic Parameters.

Parameter	Value
Number of layers	8
Neurons per layer	100
Activation function	GELU$(\cdot)$
Initial learning rate	10^{-4}
Final learning rate	10^{-6}
Initial sample size per epoch	256
Sample size per epoch	1024
Convergence threshold for target calibration	10^{-7}

Table 13: Neural network parameters

Figure 12: Loss function along training epochs