ISSN 22799362

Inclusive Monetary Policy in a Model with Heterogeneous Workers

Federico Ravenna and Carl E. Walshy

No. 734
December 2024

Carlo Alberto Notebooks

www.carloalberto.org/research/workingpapers

Inclusive Monetary Policy in a Model with Heterogeneous Workers

Federico Ravenna* and Carl E. Walsh†

December 22, 2024

Abstract

Central banks are increasingly debating monetary policies aimed at reducing inequality and ensuring employment gains are spread widely across all parts of the labor market. What are the trade-offs faced by 'inclusive policies', and the implications for the efficiency of the aggregate economy? We address this question within a model that allows for workers with different levels of productivity competing in the same job market. We compare traditional and 'inclusive' policies in terms of their impact on earnings and employment inequality, on the labor market outcomes of lower-productivity, lower-income workers, and in terms of their inflation outcomes. Inclusive policies come at a high cost in terms of inflation, but they can substantially reduce the uneven burden of a recessionary shock on the lowest-productivity workers. We provide a normative assessment, and show that while making monetary policy more inclusive is beneficial for the overall economy, making monetary policy *much* more inclusive results in sizeable deviations from the first best allocation.

Keywords: Unemployment, heterogeneity, selection, COVID-19, monetary policy
JEL classification: E24, E32, E52.

1 Introduction

In recent years many central banks have debated how much their policies generate distributional effects across different deciles of the households' income distribution, and how this should affect their actions. The Federal Reserve has lead the way in describing the maximum employment component of its dual mandate as a 'broad-based and inclusive goal'. In this view, monetary policy should ensure employment gains are spread broadly across all segments of the labor market (Board of Governors of the Federal Reserve (2020)). Optimal monetary policy results derived within Heterogeneous Agents New Keynesian (HANK) models provide a rationale

*Collegio Carlo Alberto, University of Turin, HEC Montreal and CEPR. Email: federico.ravenna@carloalberto.org
†University of California, Santa Cruz. Email: walshc@ucsc.edu.

for monetary policy to trade-off aggregate output and price stabilization against reducing uninsurable consumption risk and supporting lower-income, high-marginal utility households (Bhandari, Evans, Golosov, and Sargent (2021)).

While providing consumption insurance may be a desirable goal for monetary policy, it involves distributional issues that are often considered outside the realm of central banks' mandates (Bernanke (2015)).

This paper provides an assessment of monetary policies in an economy where heterogeneity across agents' labor productivity generates an incentive for 'inclusive policies' even with complete consumption insurance. Within a model with a distribution of labor earnings and a labor quality ladder we can neatly separate the incentives for an inclusive monetary policy related to the efficiency of the aggregate economy from the welfare issues related to unequal distribution of consumption. We study the unemployment, inflation and aggregate efficiency implications of policies supporting workers who on average have the worst labor market outcomes.

We address these issues within a new Keynesian model with search and matching in the labor market that allows for workers with different levels of productivity, or efficiency, to compete in the same job market. The model introduces a dimension of heterogeneity - unobserved productivity - that is well documented in the labor literature. We show this results in an important externality in the market allocation; firms cannot discriminate based on unobserved characteristic expending resources to post vacancies, but they can discriminate after screening workers during the hiring process. The externality translates into an 'inefficiency wedge' because firms become overly selective *against* low-efficiency workers, who are then over-represented in the unemployment pool during recessions. This model is used in Ravenna and Walsh (2022) to assess the impact of the COVID-19 shock on the distribution of productivity across employed and unemployed workers. The model predicts low-productivity workers bear the brunt of a COVID-19-type recession. In the present paper, we adopt this framework to analyze the trade-offs and efficiency implications for traditional and inclusive monetary policies in a COVID-19 recovery scenario that involves both demand and supply disruptions and that provides a strong rationale for an inclusive policy.

We show that monetary policy can be 'inclusive' and address inequality among different groups of workers ranked by their productivity level only at a substantial cost in terms of inflation. Monetary policy does alleviate the inefficiency wedge arising from labor market distortions but is an inadequate instrument to target the employment outcomes of specific wage-groups. In a recession, the low-efficiency workers are the first to lose their jobs, and end up being a larger and larger share of the unemployed as the recession peaks. Firms become very selective in a recession, hiring only very productive workers when demand is low. This outcome is suboptimal for the aggregate economy, and policy tools providing incentives to firms to target hiring of lower-productivity workers would be the most effective for addressing the inefficiency. However, monetary policy can only support low-efficiency workers by improving labor market outcomes for *all* worker types.

We compare traditional and 'inclusive' policies in terms of their impact on earnings and employment inequality, on the labor market outcomes of low-productivity workers, and in terms of their inflation outcomes. Inclusive policies can substantially reduce the uneven burden of a recessionary shock for the lowest-productivity workers. To reduce unemployment among the low-wage, low-productivity workers by 50%, inflation would double, compared to a policy reducing low-productivity workers unemployment by only 33%.

We then compare traditional and 'inclusive' policies in terms of their impact on the efficiency of the economy - that is, in terms of how much a given policy reduces the inefficient distortions in a market allocation relative to the first best when workers of heterogeneous productivity populate the labor market. We provide a normative assessment, and show that while making monetary policy more inclusive is beneficial for the overall economy, making monetary policy *much* more inclusive results in sizeable deviations from the first best allocation.

When responding to a recessionary shock, we find that a policy of price stability achieves economic outcomes very close to the first best. An average inflation targeting (AIT) policy such as the one currently adopted by the Federal Reserve performs nearly as well as a price stability policy.

Our approach isolates the impact on welfare that comes exclusively through improving the way in which the labor market functions (to this end, it is sufficient to optimally change the incentives of firms in creating vacancies) from the impact that would come from reallocating consumption across high and low-productivity workers. The latter impact identifies the effects of what is more akin to a redistributive fiscal policy. To focus on labor market distortions, we assume consumption is perfectly insured across all workers. Therefore, the inefficiency wedges we identify abstract from income distribution considerations across households. We show that even if one did not care about relative consumption across households, there would still be an argument for making policy more inclusive - albeit by a small amount.

If one instead assumed that consumption cannot be insured (as in the HANK literature) the planner would want to relieve hardships of low income households by using policies that trade-off aggregate inefficiencies against the incentives to compensate for the differential impact of aggregate and idiosyncratic shocks across workers who are heterogeneous with respect to wage income and asset holdings.

In our work we focus on labor earnings heterogeneity, as opposed to other forms of heterogeneity, such as financial market access or wealth distribution, for two reasons. First, it is well documented that workers at the bottom of the wage distribution enjoy much more volatile unemployment rates over the business cycle, and this concern is often mentioned in the policy-makers' discussion about the benefit of inclusive monetary policy (Federal Reserve Bank ofNew York (2021)). Second, the empirical evidence across the literature on the impact of monetary policy shocks across the income distribution is univocal about expansionary policy supporting low-income households by raising their labor earnings.

The closest paper to our work is Dolado, Motyovszki, and Pappa (2021), which examines the impact of a monetary easing in a search and matching (SAM) model with high and low-skill workers and capital-skill complementarity. Because they focus on capital-skill complementarity and on the transmission mechanism of monetary policy, rather than aggregate economy efficiency, they use a segmented market model for each labor-skill. In our model, competition across workers for the same vacancies implies the addition of a composition externality - which would be absent if labor markets were segmented - on top of the well-known congestion externality found in SAM models. The composition externality is not internalized by firms, and it results in endogenous and inefficient cyclicality in the incentives for firms to post vacancies. This also leads to inefficient fluctuations in the average labor productivity among those workers who are successfully matched with vacancies.[1]

Consistent with our results that the inefficiency wedge is nearly eliminated by a policy of price stability, Dolado, Motyovszki, and Pappa (2021) find that strict inflation targeting dominates standard Taylor rules in terms of stabilizing the wage skill-premium.

Instead of comparing labor market outcomes in response to hypothetical demand and supply shocks, we study policies along the path of a simulated COVID-19 recovery obtained by fitting a general equilibrium model to forecasts for the U.S. economy in the first quarter of 2020, net of the massive fiscal policy interventions that were implemented after the first month of the pandemic. This scenario, following Ravenna and Walsh (2022), is designed to capture a large shock to the U.S. economy that combines an exogenous rise in job destruction, a shock that captures the mass layoffs at the onset of the pandemic in 2020, with a negative shock to aggregate demand. Importantly, this scenario highlights the role of heterogeneity, as initially the labor supply shock hits high- and low-productivity workers equally, but over time the dynamics across worker-types diverge as firms in the downturn endogenously separate from low-productivity workers and become especially selective in favor of hiring high-productivity workers during the recovery. Thus the COVID-19 scenario provides one of the best cases for an inclusive monetary policy, and is a very illustrative example of the trade-offs faced by the monetary authority in a model with heterogeneous workers.

The paper is structured as follows. Section 2 discusses related literature. Section 3 presents the key elements of the model. Section 4 discusses the parameterized recession scenario while section 5 presents the dynamics of unemployment, earnings inequality and inefficiency wedge. Section 6 compares inflation targeting and inclusive monetary policies. Section 7 concludes. The Appendix provides details on the model, the social planner allocation, and the parameterization.

[1] A segmented-market model similar to Dolado, Motyovszki, and Pappa (2021) is used in Bergman, Matsa, and Weber (2022) in a paper mostly focused on empirical estimates of the impact of monetary policy across labor market groups, and in the HANK model in Alves and Violante (2023), which assesses inclusive policies and also include human capital depreciation when unemployed and sticky wages. Our model provides the simplest possible setup for the composition externality to affect allocations in combination with nominal price ridity, assuming efficient Nash-bargained wages.

2 Related Literature

The policy debate on whether the impact of monetary actions on inequality should be taken into account when setting policy mandates for central banks has become very active over the last few years.[2] The possibility that pursuing a monetary policy leading to a stronger recovery would disproportionately benefit low-wage workers was already raised by Okun in the 1970s, as pointed out in Alves and Violante (2023). Whether distributional issues should weigh in central bank decision-making remains a controversial issue. A 2021 survey across UK economists (Ilzetky (2021)) found 90% of the respondents in favor of inequality considerations playing no or minimal role in Bank of England policy decisions - in part because our understanding of the impact of monetary policy on inequality through several channels is still incomplete.

While some work finds that including a distributional mandate for monetary policy is unwarranted because monetary policy has fairly evenly distributed impact across the income distribution (McKay and Wolf (2023)), our work relates to a large body of evidence supporting the finding that expansionary monetary policy shocks increase earnings at the bottom of the wage distribution.

Some studies finds that the fall in income inequality resulting from expansionary monetary policy through the labor earnings channel can be more than compensated by the increase in income inequality through the asset-valuation channel, favoring high net worth deciles of the population (Andersen, Johannesen, Jørgensen, and Peydró (2023)). The empirical support for the relevance of the asset-valuation channel is mixed, and appears to be dependent on the specific asset distribution within a country (Mäki-Fränti, Silvo, Gulan, and Kilponen (2022)). There is instead unequivocal empirical evidence supporting the hypothesis that expansionary monetary policy compresses the income distribution through the extensive margin of employment: a relatively larger share of households from the lower income deciles become employed, and labor earnings inequality falls at the bottom of the wage distribution (Amberg, Jansson, Klein, and Rogantini-Picco (2022), Coibion, Gorodnichenko, Kueng, and Silvia (2017), Lenza and Slacalek (2024), Coglianese, Olsson, and Patterson (2023), Samarina and Nguyen (2024)). Our paper also relates to an older literature on trend-increase in earnings inequality, which finds that in the U.S. earnings inequality for bottom percentiles of annual earnings increases during recessions and is highly cyclical, driven by the extensive employment margin (Heathcote, Perri, and Violante (2020)). This empirical regularity is consistent with the mechanism at work in our model, where in recessions unemployment increases mostly at the bottom of

[2] For example, the Federal Reserve Bank of New York 2021 Symposium on Heterogeneity invited contributors to specifically address - among others - the questions of whether running the economy hot for an extended time reduces inequality and improves outcomes disproportionately for disadvantaged groups or for low-income households, and of how should heterogeneity affect the design of optimal policy, in particular the trade-off between price stability and full employment objectives (Federal Reserve Bank ofNew York (2021)). Yves Mersch, member of the Executive Board of the ECB, in a 2014 speech discussed the impact of monetary policy on ineqaulity through the earnings channel, noting that "evidence does suggest that labour earnings at the bottom of the distribution are most affected by business cycle fluctuations. So if monetary policy reduces unemployment, it will also reduce inequality" (European Central Bank (2014)).

the earnings distribution and predominantly through the extensive margin, while at the top of the earnings distribution there is little employment selection.

The employment selection in the SAM model we adopt leading to composition effects among the employed and unemployed pool of workers is consistent with evidence on cyclicality in *unobserved* worker quality.[3] Abowd, Lengermann, and McKinney (2003) using LEHD matched employer-employee data estimates a worker-skill index, accounting for worker-firm sorting. The skill index contains all the usual measurable worker characteristics, and also factors unobserved by the econometrician: innate ability, educational quality, social capital, and effort. The paper finds that the unobservable part of the skill index is more highly correlated with annualized wages than he observable part, with education, race, and sex only explaining between 8-11% of the total variation in the skill index. In a related empirical study, Bonhomme, Lamadon, and Manresa (2019) decomposes the variance of U.S. workers log-earnings - accounting for employer characteristics - net of observed covariates, and finds that worker heterogeneity explains 60% of earnings variation. Firm-variation explains less than 3% of earnings variation, though the estimates show strong sorting effects between workers and firms. Grigsby (2022) uses individual-worker data to explain the decline in employment in the 2008-2009 recession and the contemporaneous rise in real wages. To this end, the paper classifies occupations into groups with similar skill requirement, ranked among others by educational attainment, and estimates the relative human capital of worker-types *within* these groups. The results show that low-wage workers were especially likely to separate to non-employment relative to high-wage workers in the same occupation. Our theoretical model can be interpreted as explaining employment dynamics within a given sector-specific or education-specific labor market.

The results in this paper are also related to the recent literature on optimal monetary policy in HANK models. Dávila and Schaab (2023) establish that a utilitarian planner optimally chooses to raise output above the natural level, since 'overheating' the economy allows redistribution towards low-income, indebted households. This redistribution channel is consistent with the adoption of an inclusive monetary policy, and plays a large role also in the work of Acharya, Challe, and Dogra (2023) and Bhandari, Evans, Golosov, and Sargent (2021). Both papers find that the goal of stabilizing consumption risk across workers who face different idiosyncratic shocks to income and different ability to insure against aggregate shock plays a quantitatively large role in the optimal policy, at the cost of sacrificing traditional goals such as price stability. In our work we abstract from distributional considerations, and find that the aggregate inefficiencies generated by workers heterogeneity in productivity warrants deviations from price stability. Importantly, it is the inefficient choice in firms' hiring policies that drives the result, thus optimal policy should strive for distorting firm's behaviour relative to

[3] Human capital as measured by observable worker characteristics is also cyclical, with employment rates among low-skill workers being more cyclical in the U.S. data compared to high-skill worker. Bosler, Daly, Fernald, and Hobijn (2016) estimate a much faster increase in observable labor-quality growth during the 2008-2010 recessionary period relative to the early 2000s.

the market allocation, while workers consumption is perfectly insured.

3 The model of productivity heterogeneity

In this section, we describe the basic model of worker heterogeneity and explain how selectivity in hiring and retention decisions affects employment dynamics. The model is fully analyzed in Ravenna and Walsh (2022), and deviates from a standard NK model with search and matching model with endogenous separations such as Walsh (2005) by assuming workers' productivity is given by the combination of a random transitory component realized at the time of production and a permanent component which is worker-specific. A summary of the complete model can be found in the Appendix. The model consists of households, wholesale and retail firms, and a monetary policy authority. The representative household purchases consumption goods, holds bonds, and supplies labor to wholesale firms. Wholesale firms hire labor in a market characterized by search and matching frictions and produce a homogeneous good that is sold in a competitive market to retail firms. Retail firms transform the wholesale good into differentiated final goods which are sold to households for consumption and to wholesale firms to use in posting job vacancies. Prices of retail goods are sticky a la Calvo.

3.1 Worker types and productivity

Workers are of two types that differ in their average productivity and in its variability: we refer to these types as low-efficiency workers and high-efficiency workers. While an unemployed worker's type is unobserved ex ante, we assume a firm that is hiring engages in a process of interviewing, or screening, during which the firm is able to observe the productivity of a job applicant.

A fraction $\bar{\gamma}$ of workers are of low (l) average efficiency, while the remaining $1-\bar{\gamma}$ are of high (h) average efficiency. Total employment is $N_t = N_t^l + N_t^h$, where N^j is the number of type j workers who are employed. Define $\xi_t \equiv N_t^l/N_t$ as the fraction of employed workers who are of type l. The productivity of a type h worker is constant and equal to ϕ^h, while the productivity for a type l worker is stochastic and equal to $a_{i,t}^l\phi^l$, where $a_{i,t}^l$ is an idiosyncratic, stochastic productivity shock to worker i of type l. A low-efficiency worker is less productive than an high-efficiency worker on average, but not always. For high realizations of the idiosyncratic productivity shock, $a_{i,t}^l\phi^l$ can exceed ϕ^h. The productivity volatility of low-efficiency workers could arise because they experience highs and lows, extremely productive at times, unproductive other times, or they may be workers with unstable or chaotic lives outside of work or health issues that get reflected in variation in their job performance. Whatever the source, employers may have difficulty discerning these characteristics of workers without interviewing them or actually observing them as an employee.

3.2 Households, labor flows, and vacancies

The household consists of a continuum of workers. The representative household maximizes

$$\mathrm{E}_t \sum_{i=0}^{\infty} \beta^i \left\{ D_t \frac{\mathcal{C}_{t+i}^{1-\sigma}}{1-\sigma} - \left[v(h_{t+i}^h)(1-\xi_{t+i})N_{t+i} + \xi_{t+i}N_{t+i} \int_{\bar{a}_t}^{1} v(h_{i,t+i}^l) f(a)da \right] \right\}, \tag{1}$$

where $\sigma > 0$ is the coefficient of relative risk aversion, D_t is an aggregate preference shock, $\mathcal{C}_t$ is the sum of a market-purchased composite consumption good C_t and home-produced consumption by unemployed workers $C_t^u = (1-N_t)w^u$. The utility term in square brackets is the disutility to the household of having N_t members working. Hours worked by a type l will depends on the worker's idiosyncratic productivity shock, while because all type h workers are equally productive, they all supply the same number of hours.

The aggregate number of job interviews per period is determined through random matching, and all job seekers have identical interview-finding probability. If the efficiency type is revealed to be h, the worker is hired and produces with probability equal to one. If an interview reveals the job seeker is a type l, firms hire only type l workers whose productivity is sufficient to generate a positive surplus. Because currently employed type l workers also receive new idiosyncratic productivity realizations, only those who continue to generate a positive surplus are retained.

At the start of each period, there is an exogenous, stochastic separation realization ρ_t^x that affects all employed workers, regardless of type. In addition, with probability $\rho_t^n \equiv F(\bar{a}_t^l)$, a low-efficiency worker's productivity draw will be less than an endogenous $\bar{a}_t^l$ value guaranteeing a non-negative surplus from either an existing match or an interview.

Endogenous selection of the marginal worker

Labor is used by wholesale firms to produce a homogenous output that is sold in a competitive market at price P_t^w. Let P_t be the final goods price index and define $\mu_t = P_t/P_t^w$ as the retail-price markup. Wholesale firms post vacancies V_t, interview and screen applicants, and make retention decisions. The optimization problem of the firm implies a cutoff value $\bar{a}_t^l$ of a worker's idiosyncratic productivity realization at which the match surplus $s_{i,t}^l$ is equal to 0:

$$\bar{a}_t^l = \left(\frac{\mu_t}{\phi^l \bar{h}_{i,t}^l} \right) \left(\frac{v(\bar{h}_{i,t}^l)}{\lambda_t} - q_t^l + w_t^{u,l} \right), \tag{2}$$

where the number of hours $\bar{h}_t^l$ maximizes the joint surplus for a worker with $a_{i,t}^l = \bar{a}_t^l$. Given household preferences (9), optimal hours satisfies $v'(h_{i,t}^l)/\lambda_t = a_{i,t}^l \phi^h/\mu_t$. Equation (2) implies that an increase in the retail price markup μ_t or an exogenous fall in demand D_t reducing the marginal utility of consumption λ_t, lowers the value of intermediate firms' output, and the minimum worker-productivity level $\bar{a}_t^l$ necessary to generate a positive match surplus rises. This increase in $\bar{a}_t^l$, reduces the fraction of low-efficiency job seekers who receive job offers, and increases the endogenous separation rate of already employed low-efficiency workers. This leads to an increase in the share of low-efficiency workers in the unemployed pool (i.e., γ_t rises).

Vacancy posting when high and low efficiency workers compete for jobs

The number of vacancies posted by wholesale firms V_t, together with the number of job seekers S_t, determines the number of interviews I_t via a standard CRS matching function, which in turns determines the probability k_t^f for a posted vacancy to result in an interview. We assume Nash bargaining with firms receiving a share $1-\eta$ of the joint surplus from a match. The job posting condition takes the form

$$\frac{\kappa}{k_t^f} = (1-\eta)\left[(1-\gamma_t)s_t^h + \gamma_t(1-\rho_t^n)\mathrm{E}_t(s_{i,t}^l|hiring)\right] \tag{3}$$

where κ is the cost of posting a vacancy, expressed in terms of final goods. The left side of (3) is the expected cost of filling a position, given by κ divided by the probability k_t^f of the firm conducting an interview. The right side of the equation is the expected benefit from the match, that is, the firm's share of the expected surplus. With probability $(1-\gamma_t)$ the firm interviews (and hires) a high-efficiency worker and with probability γ_t, it interviews a low-efficiency worker which results in a hire with probability $1-\rho_t^n$. Because the expected surplus from a high-efficiency worker is greater than the expected surplus obtained from entering into an interview with a low-efficiency worker, the incentive to post vacancies falls when a rise in γ_t reduces the average quality of the unemployment pool. In turn, γ_t rises in a recession. Thus recessions are times when firms have an incentive to post fewer vacancies - reducing the job finding probability of both h and l workers, *and* when firms become more picky, since the productivity $\bar{a}_t^l$ of the marginal worker hired has to increase

3.3 Monetary policy

Consider a negative demand shock that causes a rise in the retail price markup as wholesale prices, which are flexible, fall relative to sticky retail prices. The resulting rise in $\overline{a}_t^l$ increases the endogenous separation rate and generates an inflow into unemployment of type l workers. It also reduces the outflow of type l workers from unemployment as more are screened out in interviews. With a higher share γ_t of low-efficiency workers in the pool of unemployed workers, firms posting vacancies are more likely to interview a type l workers and less likely to make a successful hire, and from (3) the incentive to create vacancies falls. Unemployment duration for low-efficiency workers rises *both* because the probability of getting interviewed has fallen *and* because the probability of being hired, conditional on being interviewed, has fallen.

Similarly, a persistent shock to the exogenous separation rate ρ_t^x increases unemployment of both worker types. However, by reducing the expected duration of matches, the continuation value of a match falls, and, from (2), $\bar{a}_t^l$ rises, making re-hiring for low efficiency workers in the unemployment pool harder.

Because the outflow from unemployment of l workers is slower during the recovery following a recessionary shock, expansionary monetary policy will benefit low-efficiency workers disproportionately. Notice however that monetary policy can support employment of low efficiency workers only by providing firms an incentive to increase vacancies V_t, which will create additional employment opportunities for both h and l workers.

The baseline scenario used to assess alternative policies assumes the central bank follows a Taylor rule targeting inflation:

$$\ln(1 + i_t) = -\ln \beta + \omega_\pi \pi_t \tag{4}$$

where $\omega_\pi = 1.5$.

4 A severe recession scenario in a parameterized model with worker heterogeneity

We use the model of worker heterogeneity to assess the impact for several measures of inequality among workers across alternative monetary policy rules, and the implications for efficiency of the allocations in the aggregate economy. Rather than examining average outcomes, we study a specific episode of recovery from a deep recession - the recovery after the initial COVID-19 lockdown in the first quarter of 2020. This experiment can shed light on the consequences of 'inclusive monetary policies' which aim specifically at ensuring that a recovery spreads to all segments of the labor market, including those that experience larger rise in unemployment in a recession, and thus tend to be more expansionary than traditional monetary policies accounting only for aggregate outcomes.

The COVID-19 recovery path provides a useful scenario since it gives the best chance for an inclusive policy to be effective. The initial recession can be modeled as a combination of a fall in demand and an increase in firm-employee separations affecting all workers (net of sectorial effects). As we discuss in the following, within our model an exogenous increase in the separation rate can be interpreted as a supply or productivity shock. Note that while this is an aggregate shock affecting all worker types, low efficiency workers are screened against in the recovery hiring, so that the burden of the recessionary shock is mostly on their shoulders. In this scenario, inclusive policies would indeed be called upon to redistribute the recessionary shock across workers of different productivity levels.

Ravenna and Walsh (2022) use the model of worker heterogeneity to investigate the impact of a COVID-19 recession and discuss the key propagation mechanisms that cause the pandemic to have differential effects across worker-types. We follow their baseline parameterization to build a recovery path consistent with the U.S. data from 2020. We summarize here the parameterization and the shocks employed to capture the COVID-19 recovery path.

4.1 Worker Heterogeneity Parameterization

Table 1 reports the key parameters and moments of the model. Details on the model parameterization are in the Appendix.

Table 1: Parameters and Steady State Values		
Targeted Steady State Values		
Unemployment rate	u_{ss}	5.6%
Unemployment rate: $l - efficiency$ labor	u_{ss}^{l}	9.87%
Unemployment rate: $h - efficiency$ labor	u_{ss}^{h}	2.97%
Average hours per worker	h_{ss}^{av}	0.33
Vacancy posting cost share of output	$\frac{\kappa V_{ss}}{Y_{ss}}$	0.015
Probability of vacancy matched with applicant	k_{ss}^{f}	0.9
Implied Parameters and Steady State Values		
Labor force $l - efficiency$ workers share	$\overline{\gamma}$	0.38
Unemployment share $l - efficiency$	γ_{ss}	0.52
Employment share $l - efficiency$	ξ_{ss}	0.36
Relative productivity of high-/low-efficiency workers		1.16
Relative productivity of employed/unemployed workers		1.04
Interview matching function efficiency	ψ	0.743
Disutility of labor hours	ℓ	8.5
Value of home production	w^{u}	0.057
Vacancy posting cost	κ	0.036
Inverse of labor hours supply elasticity	χ	1
Relative risk aversion	σ	1
Unemployment elasticity of matching function	α	0.6
Workers' share of surplus	η	0.6
Steady-state separation rate	ρ_{ss}	7.45%
Exogenous separation rate	ρ^{x}	5.8%
Endogenous separation/screening rate	ρ_{ss}^{n}	4.9%
AR(1) parameter for exogenous shocks	ρ_{z}	0.7
Price elasticity of retail goods demand	ε	11
Discount factor	β	0.99
Average retail price duration (quarters)	$\frac{1}{1-\omega}$	4

Table 1: Baseline Parameterization. Average productivity of high- and low-efficiency worker-hours is given by ϕ^{h} and $\phi^{l} \int_{0}^{1} a_{i}^{l} dF(a_{i}^{l})$. U.S. unemployment rate for low- and high-efficiency workers from the classification of workers based on employment spell duration from LEHD data taken from Gregory, et al. (2021). See the Appendix for details.

In our parameterization, the share of type l workers in the total labor force is 38%. Because the separation rate for these workers is about 40% larger than the average separation rate, their share in the steady-state pool of job seekers γ_{ss} is 52%, while their share ξ_{ss} in steady-state employment is only 36%. Thus, low-efficiency workers are over-represented in the pool of unemployed, and this pool has a lower average productivity than the pool of employed workers, as reported in table 1. When matched for an interview with a firm, high-efficiency workers are expected to have an hourly productivity 16% higher than low-efficiency workers. The productivity ratio between employed and unemployed workers is smaller and equal to 1.04, since only relatively highly productivity type l workers are retained in employment. The extent of selection at hiring is small; firms screen out only about 2.55% of the workers they interview and just 4.9% of the type l workers who are interviewed. However the screening-out rate increases in a recession as γ_t, the share of low-efficiency workers among the unemployed increases and $\bar{a}_t^l$, the minimum productivity for a match to generate a positive value, increases.

4.1.1 Shocks

While sectorial redistributional effects were important, the rise in the saving rate and the subsequent fall in inflation in the first quarter of 2020 suggest the pandemic also featured an overall drop in spending, as would be generated by a negative aggregate demand shock.

To describe the COVID-19 initial drop in aggregate output, therefore, we assume that the economy is hit by both a positive shock to exogenous separations and a preference shock that reduces demand. The public heath responses to COVID-19 resulted in the destruction of job matches and had features common to a negative supply shock in that it reduced the economy's ability to produce without generating a corresponding drop in demand. The separation shock captures some of the supply side aspects of COVID-19 in a manner that adversely affects all workers, regardless of type. Based on CPS data, Cortes and Forsythe (2020) report that over 80% of the pandemic-induced fall in employment in April 2020 resulted from individuals exiting employment, while reduced hiring accounted for the rest. We then capture the accompanying fall in aggregate demand through a negative preference shock, representing a shift in households' preferences away from current consumption and towards home-production and leisure.

To produce the COVID-19 scenario, we condition on forecast paths for output and total separation produced at the beginning of the COVID-19 recession. The model uses the forecasts to endogenously allocate the path of total separations and output across an exogenous, unexpected shock to ρ_t^x and an exogenous preference shock D_t. We then produce conditional forecasts for labor market variables and other aggregate variables which are not supplied to the model. The Appendix describes the data used in simulating the recovery path.

Both shocks are essential to discuss a pandemic-induced recession. The exogenous separation shocks act as a supply shock; the value of each existing or potential match falls, ceteris paribus, since the cost per vacancy is fixed but its return in terms of match-lifetime production

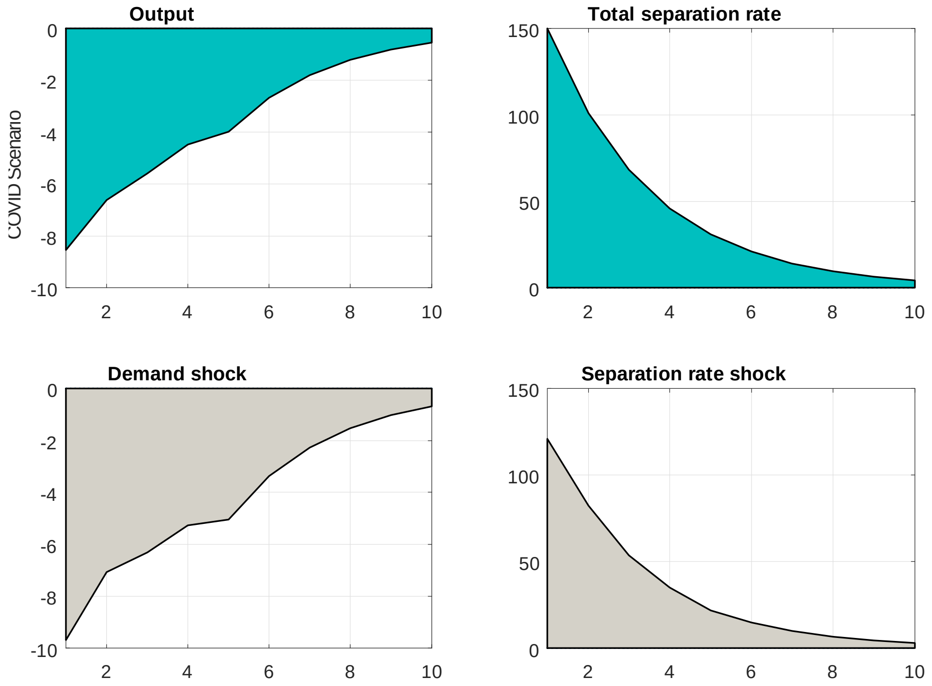

Figure 1: Parameterized recovery path after COVID-19 initial recession. Targeted values (first row) and filtered shock (second row). The path is conditional on the baseline inflation targeting policy rule

is lower, given that the match is expected to have a shorter duration. By itself, the separation shock is not sufficient to lower demand in line with the SPF forecast for the U.S. economy as it pushes firms to quickly rehire a large share of the separating workers. The preference shock acts as a demand shock that allows the model to match the output loss for a given separation path. The demand shock lowers output and the demand for labor but cannot, by itself, reasonably account for all of the surge in total separations. Together, the two shocks capture important aspects of the 2020 recession. Figure 1 reports the targeted values for output y_t and total separation rate ρ_t, and the implied shocks filtered through the model.

5 Inequality and efficiency along the recovery path

5.1 Unemployment dynamics across worker types

Consider the outcome of a COVID-19 recovery after the initial 2020Q1 drop in aggregate activity. Ravenna and Walsh (2022) find that conditional on the observed and forecast path of output and total separations, the parameterized model estimates that more than two-thirds of excess unemployment is shouldered by low efficiency workers even though the initial spike in job loss affects both worker types. On the firms' side, the wedge between the interview probability and the job-filling probability increases - firms become more 'picky'. An econometrician not accounting for time-varying selection would measure a fall in the efficiency of matching during the pandemic recession and a negative TFP shock affecting the pool of unemployed workers. In effect, average hourly productivity is falling by about 1.5% within the pool of unemployed exclusively because of the shift in its composition.

The behavior of unemployment implied by the model is shown in Figure 2, which plots the aggregate unemployment rate and the unemployment rates of high-efficiency and low-efficiency workers in the left panel. Because the steady-state value of U^l is much higher than that of U^h, each series on the left-hand side panel is shown as percentage points deviation from its own steady-state value. The right panel translates these unemployment movements for each worker-group into the share of the *aggregate* unemployment rate response that consists of each worker type, in excess of the aggregate steady state unemployment rate. The outer envelope of the shared regions in the right panel is the total unemployment rate and corresponds to the bold line in the left panel. For type h workers, their unemployment rate rises by less than 3 percentage point above its steady-state value. In contrast, that of type l workers jumps by almost 15 percentage points, from its steady-state value of just under 10% to almost 25%. Thus, the unemployment rate of type l workers rises more and from a higher steady-state base than the total unemployment rate or the rate for type h workers. And after 6 quarters, the unemployment rate among high-efficiency workers is less than 1 percentage point above its steady-state value, while for low-efficiency workers it is still 4.6 percentage points above steady state. The right panel of the figure shows the composition of total unemployment among worker types. Even though type l workers are less than 40% of the labor force, they account for the bulk of the pandemic-induced higher aggregate unemployment rate.

In a pandemic scenario social distancing and lockdowns leads to mass layoffs that affect both worker types. Because most workers are high-efficiency types, more type h workers initially flow into the unemployment pool and the share of low-efficiency workers in the unemployment pool γ_t initially falls, as shown in the top left panel of Figure 3. But the fall in γ_t is quickly reversed and then remains persistently above its steady-state value. This reversal reflects greater selectivity by firms; the productivity cutoff value $\bar{a}_t^l$ rises as shown in the top right panel of the figure. Firms become more selective for two reasons. First, the drop in demand for the wholesale good results in a fall in its price relative to the index of sticky retail goods prices; that is, the retail price markup μ_t rises, as the bottom left panel of the figure shows. The marginal revenue product of a type l worker with idiosyncratic productivity $a_{i,t}^l$ is $a_{i,t}^l \phi^l / \mu_t$. When μ_t rises, the value of $a_{i,t}^l$ necessary to generate a positive match surplus rises. Second, the value of $\bar{a}_t^l$ also depends on the continuation value of a match. With a persistent rise in ρ_t^x the expected duration of a match and its continuation value falls, implying only very productive type l workers generate a positive surplus. The rise in $\bar{a}_t^l$ increases the endogenous separation rate, and this further reduces the expected duration of a match. The rate at which type l workers enter the pool of unemployment rises and their exit rate falls, causing γ_t to rise above its steady-state value. As both γ_t and $\bar{a}_t^l$ increase, the efficiency of the aggregate

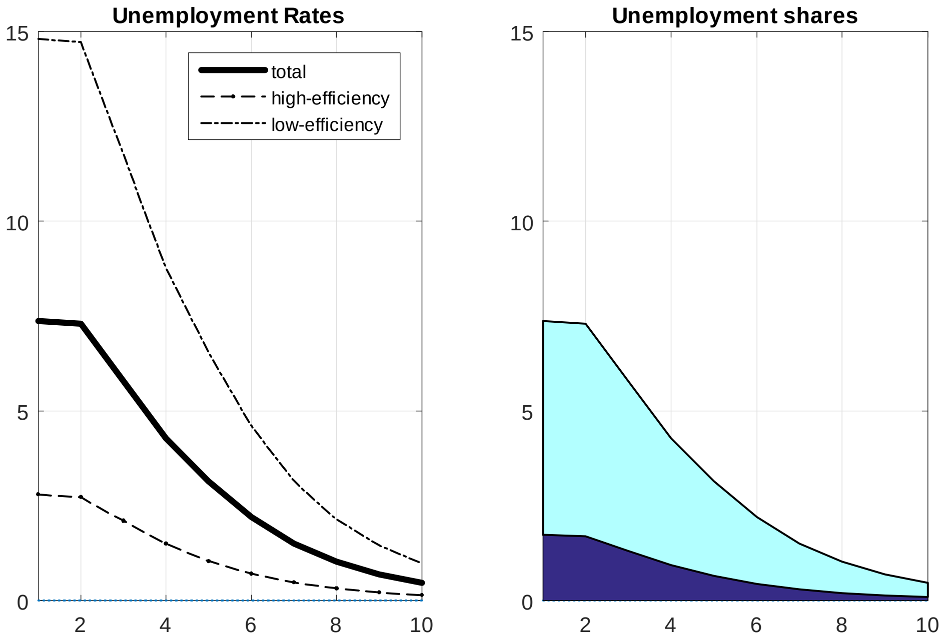

Figure 2: Recovery path following COVID-19 shock unemployment responses. The left panel shows the responses of U, U^h, and U^l, with each expressed in terms of percentage point deviation from their steady-state values. The right panel shows the aggregate unemployment rate response and the share accounted for by each worker type: type l (in light) and type h (in dark). Unemployment rate is measured in percentage points of excess unemployment above the aggregate economy steady state unemployment rate.

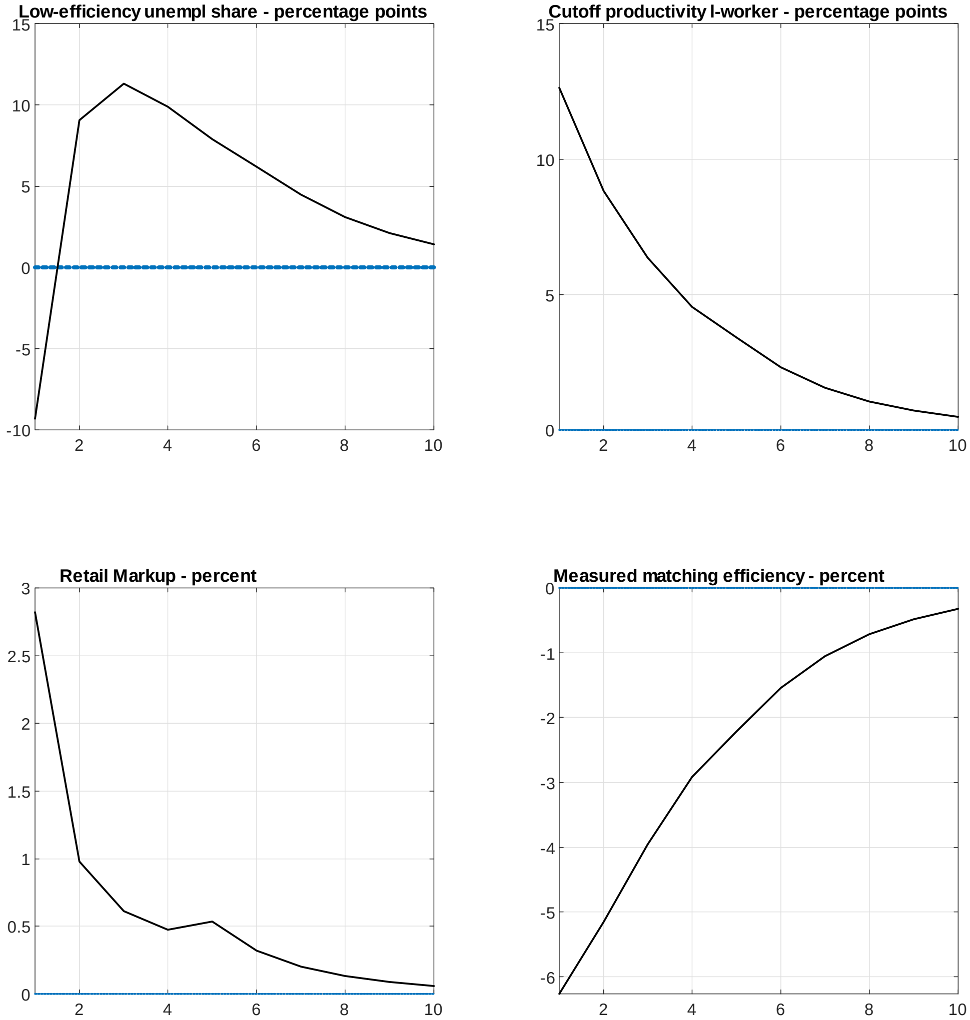

Figure 3: Recovery path following COVID-19 shock: Responses of γ_t (upper left), $\bar{a}_t^l$ (upper right), μ_t (lower left) and $1 - \gamma_t \rho_t^n$ (lower right) to the COVID-19 shocks. Upper panels are percentage point deviations from steady state; lower panels are percent deviations from steady state.

matching function as defined by $1 - \gamma_t \rho_t^n$ falls (see bottom right panel).

5.2 Inequality in earnings: aggregate and worker-specific outcomes

One key question policymakers face when assessing the impact of inclusive policies is how to measure the impact of alternative policies on labor market outcomes. We showed that the burden of the recession in the COVID-19 scenario falls squarely on low efficiency workers when measured by the unemployment rate[4]. However measures of wage inequality may show tightening in the wage distribution across employed workers. We use the model to study the dynamics of the wage distribution over the recovery path.

[4] Since in the model there is no out-of-labor-force margin, changes in the unemployment rate mirror changes in the employment rate.

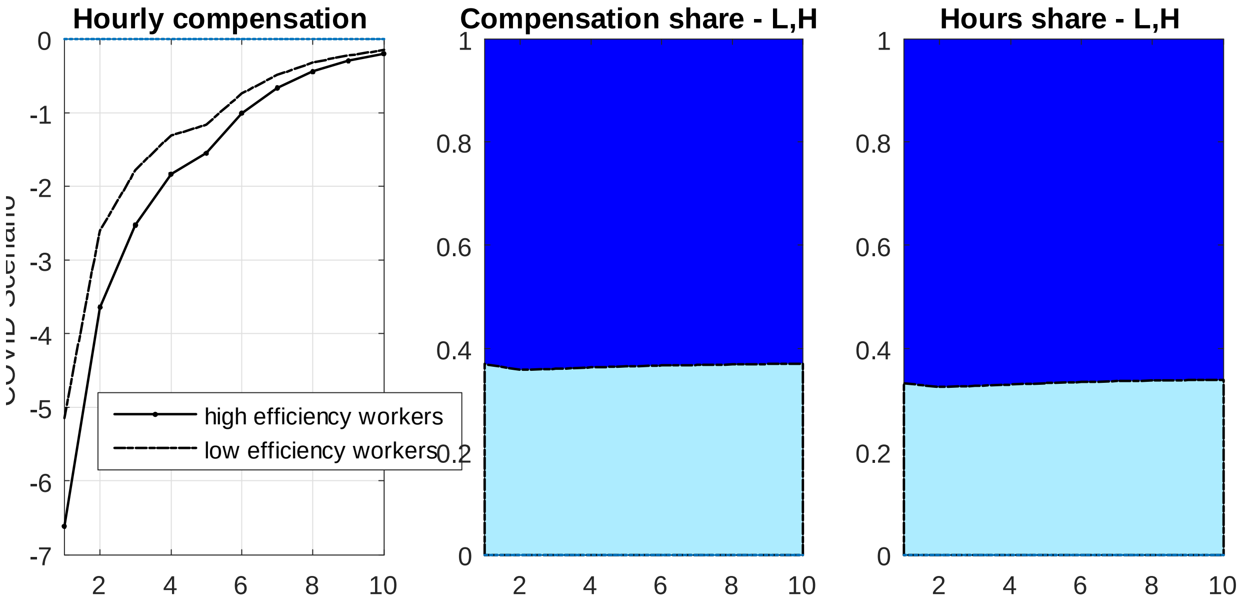

Figure 4: First panel: Recovery path. Impulse response function of hourly compensation for $h-$workers and for the average $l-$worker. Second and Third panels: Share of total compensation and total hours in the economy accruing to each group of workers along the recovery path. Dark areas indicate share accruing to $h-$workers.

First, we examine aggregate outcomes. The first panel of figure 4 shows that in the recovery path scenario over the 10 quarters following the initial shock, hourly compensation falls by about 5% for low-efficiency worker, and by a larger amount for high-productivity workers, equal to about 6.5%. The dynamics of hourly compensation for both groups of workers is similar along the recovery path. The overall share of economy-wide labor compensation accruing to the two worker-types is about constant over the recovery. The share of total hours worked in the economy displays a similar behaviour over the recovery. Overall the aggregate statistics point towards a similar dynamics for the hourly wage of l and h workers, and for little change in how hours or earnings are distributed across the two types of workers.

Next, we investigate the sources of the aggregate earnings dynamics. Along the COVID-19 recovery the average number of hours per employee in the model does not vary significantly - a pattern observed also in the US data. The center panel of Figure 4 shows that on average the earnings dynamics of high and low productivity workers ares similar, and overall the aggregate allocation across workers of the total hours worked in the economy (right panel) does not deviate from the pre-pandemic allocation.

This aggregate result hides deep changes along the recovery path in the distribution of average hours and in the distribution of total earnings within the labor force of each worker type. First, employment of high-efficiency workers is only moderately affected in the recession, however they experience a sharp fall in individual labor hours. Second, employment of low-efficiency worker decreases severely in the recession, and simultaneously the least productive

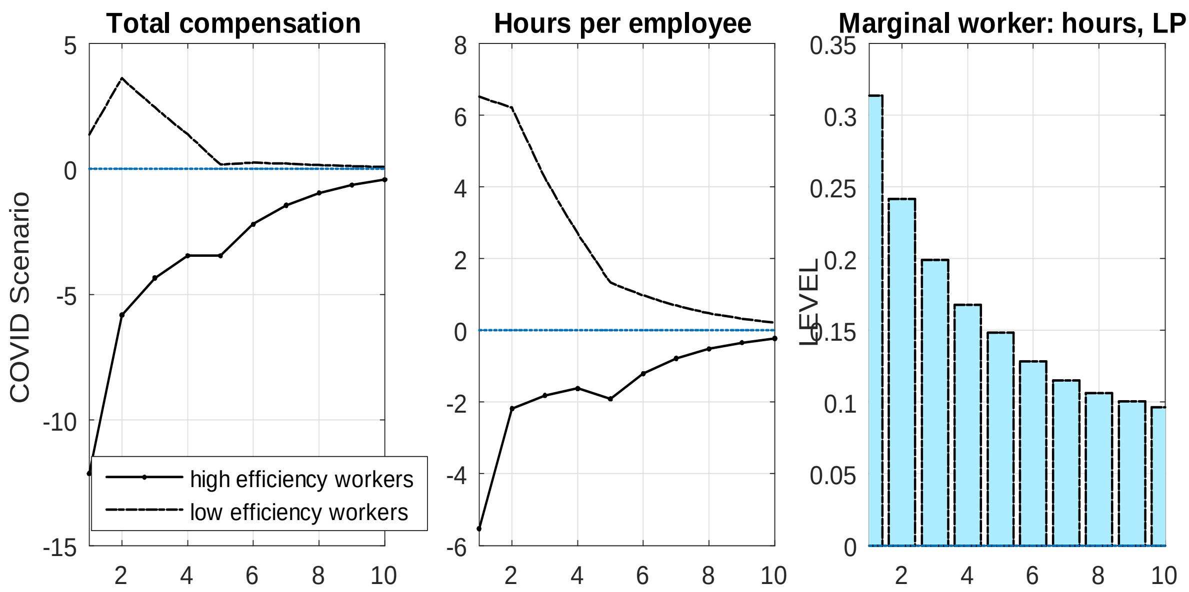

Figure 5: First two panels show total compensation per employee and total hours worked per employee for $h-$ and $l-$workers. Last panel shows absolute number of hours as share of available time supplied by the employed worker with the lowest level of productivity (the marginal worker).

workers within this pool are heavily selected against in firms' hiring.

Figure 5, second panel, shows that hours per employee *increase* for low-productivity workers. Because hours per employee (the intensive margin) increase for the workers' group with the largest fall in the total number of employees (the extensive margin), the aggregate hours share in the economy for each worker group, shown in the third panel of figure 4, is approximately constant, even if the average worker within each efficiency group experienced a large change in the intensive margin of work.

As a result, the increase in average hours for each low-productivity worker employed is large enough to lead to a *rise* in total compensation for this group of workers, as shown in the first panel of figure 5, which shows compensation per worker rather than the hourly wage. This in turn implies that inequality in total earnings across those high and low productivity workers who are employed *falls* during the recovery.

This translates into improved earning outcomes for low-productivity workers who remain employed. However, the rise in earnings and the fall in earnings inequality only occurs among the employed workers - those workers for which we can measure earnings in the data - and is *entirely* driven by firms becoming more selective in hiring. The rise in earnings does not translate into the *average* low-productivity worker benefitting from the increase in measured hours and earnings within the employed pool of workers that occurs after the COVID-19 shock. In fact, the low-productivity group is overall markedly worse off, since the fall in inequality is entirely driven by stronger selection at hiring. The number of hours for the marginal $l-$worker

- that is for the least productive employed worker in the economy that will still generate a positive surplus in a match and will not be screened out by firms - increases by more than three-fold, as shown in the third panel of figure 5. The increase in hours is approximately proportional to the increase in the minimum level of productivity $\bar{a}_t^l$ that is the lower bound for a match to be formed between a firm and an l worker. Overall a large share of marginal workers at the bottom of the productivity distribution are driven out of employment.

A policy aimed at reducing inequality among all workers - not only among the employed ones - should target the extensive margin of employment, by lowering the level of unemployment among the most disadvantaged group of workers instead of increasing the measured amount of total hours this group provides to firms. In equilibrium, lowering $l-$unemployment implies lowering the minimum level of productivity for the marginal worker that will generate a positive surplus. This will result in lower hours and earnings for the marginal and for the average $l-$worker, and an increase in earnings inequality across *employed* workers - however this is an equilibrium implication of reducing the inequality of labor market outcomes across *all* workers.

5.3 Inequality and the inefficiency wedge

Selection across workers of different productivity levels leads to a new source of inefficiency relative to NK models with random search in the labor market and homogeneous workers. In the competitive equilibrium, individual firms ignore the effects their vacancy posting and separation decisions have on the average quality-composition of the pool of unemployed. Relative to the efficient equilibrium, low-efficiency workers experience spells of unemployment that are inefficiently frequent and average unemployment duration that is inefficiently long.[5]

This implies that there exist an incentive for monetary policymakers to stimulate the economy to facilitate hiring of low efficiency workers. The efficient equilibrium provides a benchmark for interpreting the aggregate consequences of the distortion in the distribution of labor hours across workers. As in Ravenna and Walsh (2022), one can measure the deviation of the market allocation from the first best by adding up the deviations of group-specific unemployment rates and of output from the efficient outcomes over the 10 quarters of the COVID-19 recovery predicted by the model. The result is shown in figure 6.

Excess unemployment for each worker type are shown in the left panel of Figure 6. Both unemployment rates increase more in the market equilibrium than in the social planner's allocation, but the impact is particularly pronounced for type l workers. After 5 quarters, type l workers have suffered a cumulative efficiency wedge equal to 100% of their steady-state unemployment rate. Since the latter is roughly 10%, they experience an additional 10 percentage points of unemployment over this period that is socially inefficient. In contrast, type h workers, while also experiencing inefficiently high unemployment, suffer cumulative excess unemployment of less than 20% of their steady-state unemployment rate of 3%. With

[5] For a proof see Ravenna and Walsh (2022). The Appendix provides the solution to the planner's problem, and a discussion of the difference between the market and optimal allocations.

Figure 6: Recovery path following COVID-19 shock cumulative efficiency losses. The cumulative gap between the market outcome and the efficient allocation in response to the COVID-19 shocks for unemployment rates by worker type (left panel) and for output (right panel) measured as a percent of steady-state values.

unemployment rates higher and employment lower in the market equilibrium, aggregate output is also inefficiently low in response to the shocks. The cumulative loss in output due to the inefficiency wedge is shown in the right panel of figure 6. After 10 quarters, this efficiency loss totals 4.75% of steady-state output.

The result suggests that a policy that reduces inequality in unemployment outcomes across worker groups and raises employment for $l-$workers is desirable from a welfare perspective.

6 Assessing cost and benefits of inclusive monetary policies

In this section we assess the implications of inclusive monetary policies, and the desirability of such policies from the perspective of closing the gap with the efficient allocation for the aggregate economy.

The previous section showed that any inclusive policy should reduce inequality across workers by improving employment levels and making employment accessible to a larger share of low-productivity workers. Reducing measured wage inequality would instead be a misleading objective.

6.1 Policy rules

We now consider modifications to the policy rule (4) that are designed to more directly target labor market developments. Specifically, we consider an unemployment rate version of a basic Taylor rule of the form

$$\ln(1 + i_t) = -\ln \beta + \omega_\pi \pi_t + \omega_U U_t, \quad (5)$$

where U_t is aggregate unemployment. We focus on how the value of ω_U, measuring the aggressiveness with which policy responds to unemployment, affects the economy's response in our COVID-19 scenario. We illustrate the role of ω_U by discussing the results for three different values: $\omega_U = 0$ corresponding to our benchmark rule (4); a rule with $\omega_U = -0.4$, which succeeds in reducing the impact of the pandemic on aggregate unemployment relative to the benchmark rule by 33%; and a rule with $\omega_U = -0.8$ to capture a stronger response to unemployment[6]. Finally we consider an Average Inflation Targeting (AIT) rule where the interest rates responds to an average of inflation over the past 12 months:

$$\ln(1 + i_t) = -\ln \beta + \frac{\omega_\pi}{4}[\pi_t + \pi_{t-1} + \pi_{t-2} + \pi_{t-3}] \quad (6)$$

For all three rules, we maintain $\omega_\pi = 1.5$ as in the benchmark rule. We also compare the performance of the Taylor rules to a price stability policy.

6.2 Inflation and unemployment outcomes of inclusive policies

We have established that an 'inclusive policy' should reduce inequality by achieving a reduction in the low-efficiency workers unemployment, and thus making employment accessible to the more disadvantaged among the workers in the unemployment pool. Figure 7 illustrates the implications for inflation and unemployment of an inclusive policy.

Responding to unemployment following policy (5) with a coefficient $\omega_U = -0.4$ reduces low productivity workers excess unemployment from 15 to 10 percentage points above the steady state. This improvement comes at a steep cost in terms of inflation, which rises to 2% on impact. This is a relatively modest level in absolute terms, however consider that the COVID-19 shock is deflationary under the baseline policy, with inflation falling about 1.5% below its pre-recession level. Alternative models for inflation dynamics - for example, adding a share of backward-looking firms implying a hybrid NK Phillips curve - would reduce the absolute level of the inflation increase, however would not substantially modify the *relative* increase in inflation. That is, even with a more inertial model of price-setting, an inclusive policy would lead to an increase in inflation of the same size relative to the inflation level at the time of the

[6] For reference, the Taylor rule discussed in the 2021 Monetary Policy Report (MRP) of the Federal Reserve (p. 44) puts a coefficient of one on the unemployment rate gap. Since the MRP rule expresses interest rates and inflation at annual rates, while our framework is based on a quarterly frequency, the MRP value of one would imply a value $\omega_U = -0.25$ in (5), while the MRP's balanced-approached rule implies a value of $\omega_U = -0.5$. Thus our choice of $\omega_U = -0.4$ is bracketed by the values in the MRP, while $\omega_U = -0.8$ represents a much stronger response to unemployment.

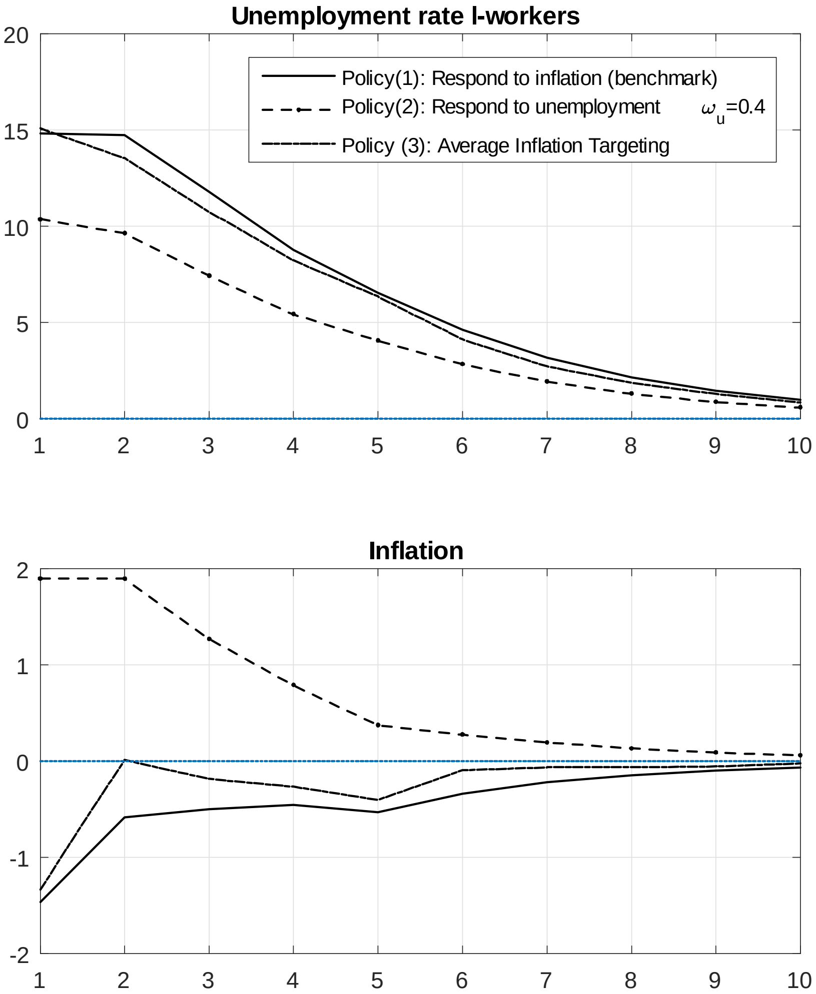

Figure 7: Excess unemployment rate for $l-$workers above the steady state value, measured in percentage points, and retail inflation rate along the recovery path under three different monetary policies.

recessionary shock, for a given change in the unemployment rate.

The AIT policy only marginally benefits the employment rate of low-productivity workers, reducing their unemployment rate by about 1% in the first two quarters. However, it achieves an inflation path closer to price stability, minimizing the distortion from markup fluctuations which occurs whenever shocks drive the inflation rate away from 0. Note that along the recovery path, which leads the economy to lower the inflation rate below the steady state value, an AIT policy is a more accommodative policy compared to the benchmark inflation targeting rule, as it results in a higher inflation rate.

6.3 Distributional outcomes and the inflation trade-off of inclusive policies

We use five metrics to evaluate the distributional effects of alternative monetary policies, and the related trade-off with inflation. The first is the output loss, measured as the cumulative loss in output, expressed as a percentage of steady-state output, over the first two years of the recovery. The second metric we call the 'type l unemployment loss'. This is defined as the cumulative excess percentage points of total unemployment U_t accounted for by l-workers over the same two-year horizon, relative to the steady state.[7] We also calculate the corresponding type h unemployment loss. Our fourth metric is a measure of the distribution of excess unemployment across the two worker types. We label this the *inequality ratio* and define it simply as the ratio of the type l unemployment loss to the type h unemployment loss. A smaller value of this ratio implies that the burden of aggregate unemployment is more equally distributed across the two groups. 'Cumulative inflation' is a measure of inflation volatility over the first 8 quarters, measuring the cumulative absolute value of inflation deviation from price stability. 'Sacrifice ratio' over the horizon T is defined as the cumulative volatility of inflation (measured in absolute values), divided by the cumulative increase in unemployment relative to the initial level conditional on the benchmark policy:

$$Sacrifice\ Ratio_T \equiv -\frac{\sum_{t=1}^{T=8} abs(\pi_t)}{\sum_{t=1}^{T=8} (U_t - U_0)}.$$

This definition of the sacrifice ratio measures how large inflation volatility needs to be for every point of excess unemployment gained relative to its initial level when the COVID-19recession starts. A higher ratio indicates that each 1% reduction in unemployment has a larger cost in terms of deviations from price stability. The implications of the alternative rules are reported in Table 2.

[7] At each t the excess $l-$unemployment rate is computed as $\alpha U_t^l - U_{ss}$ where $\alpha = (N_t^l - N_{ss}^l)/(N_t - N_{ss})$.Scaling the number of unemployment workers by the total number of unemployed allows us to compare the unemployment losses across workers types.

Table 2: Outcome for Alternative Policies				
	Alternative policies			
	(1)	(2)	(3)	(4)
Variables in policy rule	π	π, U	π, U	**AIT**
Response coefficients ω_π and ω_U	**1.5, 0**	**1.5, −0.4**	**1.5, −0.8**	
1) Output loss	**34.89%**	**26.85%**	**22.42%**	**34.35%**
2) l**-unemployment loss**	**25.34%**	**16.38%**	**11.54%**	**23.89%**
3) h**-unemployment loss**	**7.24%**	**6.60%**	**6.25%**	**7.21%**
4) Inequality ratio	**3.49**	**2.48**	**1.84**	**3.31**
5) Cumulative inflation (volatility)	**4.24**	**6.81**	**12.62**	**2.42**
6) Sacrifice ratio (volatility)	**0.16**	**0.19**	**0.31**	**0.08**

Table 2: Policies:1 benchmark policy given by (4); column 2 given by (5) with $\omega_U = -0.4$, parameterized to reduce the rise in unemployment in the first period of the COVID-19shock by 33% relative to column 1; column 3 given by (5) with $\omega_U = -0.8$; column 4 given by average inflation targeting policy (6). Output and unemployment losses by worker-types, the unemployment inequality ratio, are defined in the text. Cumulative inflation: cumulative absolute value of inflation deviation from price stability. Sacrifice ratio: Cumulative volatility of inflation (absolute values), divided by cumulative fall in unemployment relative to initial level at start of the recession.

First, consider the performance of the l-unemployment loss and the inequality ratio under inclusive policies in columns (2) and (4), in Table 2 relative to the benchmark policy in column (1) given by eq. (4). Inclusive policies have a large impact in reducing inequality in unemployment across worker groups, and in reducing the absolute level of unemployment of low productivity workers. Such workers benefit disproportionately from these policies.

Second, consider the trade-off with inflation. Inclusive policies come at a large cost in terms of overall inflation volatility (see row 5). Unfortunately, inclusive policies also suffer from an increasing average cost of inflation volatility per point-reduction of unemployment (see row 6). This means that the marginal cost of making policies more inclusive along the recovery path is rapidly increasing.

Third, an AIT policy (column 4 in Table 2) returns only a modest gain over the benchmark policy in terms of reducing inequality and improving outcomes for low productivity workers. However, since it stabilizes inflation, it comes at a very low sacrifice ratio. In other words, inflation stabilization comes at a low cost (in fact, an improvement) in l-unemployment under AIT. On the other hand, the impact of AIT on the inequality ratio (see row 4) is negligible.

The general explanation for these results is that monetary policy affects the aggregate economy but cannot ensure that expansionary policy designed to dampen the adverse impacts of the pandemic will propagate in a manner that helps those workers most affected by the recession. Monetary policy is able to reduce unemployment by reducing interest rates and increasing aggregate demand. This increased demand for retail goods also boosts the demand for wholesale goods and the demand for labor. Under the benchmark policy, the retail price markup rises as the recessionary drop in demand causes wholesale prices to fall. Inclusive policies are more expansionary, support aggregate demand and output, and thus limit the rise in the markup and in $\bar{a}_t^l$, helping to reduce the inflow of type l workers into unemployment and increase their exit rate from unemployment. However, the effect on the inequality ratio is limited because the markup affects the marginal revenue product of *both* worker types; the value of the output produced by a type h worker is ϕ^h/μ_t and that by a type l worker is $a_{i,t}^l \phi^l/\mu_t$. To the extent μ_t rises less, the match surpluses for low-efficiency workers *and* high-efficiency workers fall less. Since both worker types benefit from the inclusive policy, the ability of monetary policy to have a significant and differential impact on the different worker groups is limited.

6.4 Efficiency outcomes for inclusive policies

Reducing inequality in labor market outcomes can be a desirable goal for monetary policy for two reasons. First, it may be welfare-improving whenever inequality in labor market outcomes translates into inequality in consumption outcomes. Divergence in marginal utility of consumption across households implies that increasing income for low-productivity workers at the expense of high-productivity workers is optimal. This source of welfare gain is what the existing literature incorporating heterogeneity has focused on; we exclude this gain by assuming perfect consumption insurance. Second, even if consumption is perfectly insured across workers, welfare can be improved by enacting policies that reduce the market inefficiency from firms' hiring too few low-productivity workers. This aggregate inefficiency, on which our paper focuses, is independent of consumption distributional considerations, and aligns with traditional goals of monetary policy aiming at minimizing aggregate distortions in the economy.

To assess the implications of inclusive policies, we measure the cumulative unemployment wedge for low productivity workers, that is the distance from the unemployment outcome that would be obtained in the first best. The comparison is illustrated in figure 8.

The benchmark policy generates a large inefficiency wedge, implying it would be optimal to have policies subsidizing firms' hiring of low-productivity workers. An inclusive monetary policy would achieve this goal by reducing the markup μ_t, thus reducing the level of productivity for the employed marginal $l-$worker necessary to generate a positive surplus.

An inclusive policy targeting unemployment results in an excessive fall in $l-$unemployment relative to the first best. hence the inefficiency wedge turns negative. The result hinges on the nature of the shock - the initial recession is the result of real shocks which require in the first best a substantial increase in unemployment of both worker types.

Average inflation targeting reduces significantly the $l-$unemployment inefficiency wedge. Instead, a policy of price stability results in an inefficiency wedge close to 0. This implies that eliminating the price rigidity distortion also goes a long way towards eliminating the search inefficiency. This is not surprising since the model assumes efficient Nash bargaining.

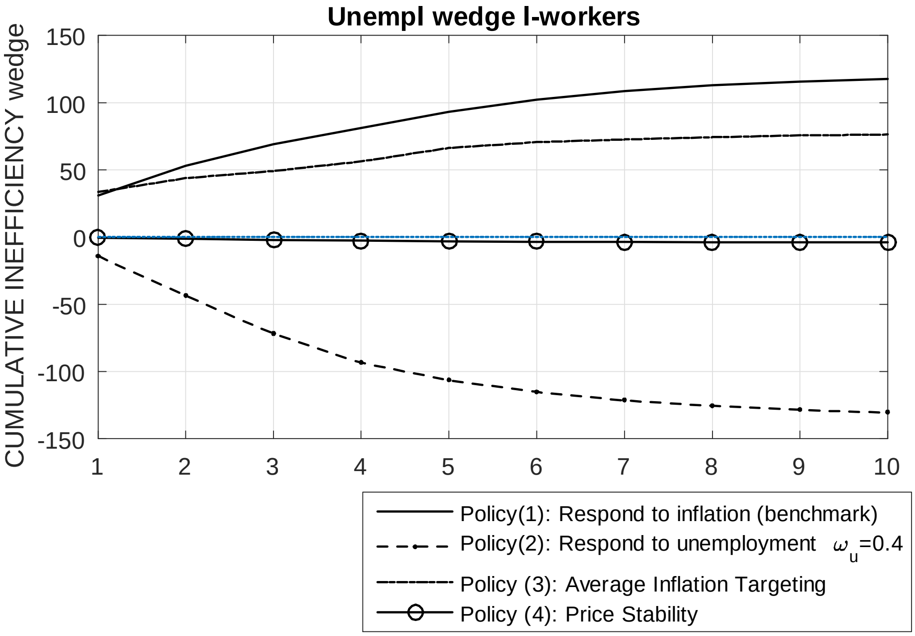

Figure 8: Cumulative $l-$workers unemployment inefficiency wedge in the recovery path under three different Taylor-rule policies, and under price stability (circles). Wedge measured in percentage of steady-state unemployment for $l-$workers.

Deviations from this wage setting benchmark would introduce an additional distortion in the overall market outcome for the aggregate employment level, which may easily compound with and worsen the distortion in the allocation of hours across workers of different productivity level.

7 Conclusions

We show that monetary policy can indeed reduce inequality and promote inclusive employment. However achieving this goal can be costly in terms of inflation and suboptimal in terms of welfare. We distinguish the positive impact of policy, from the normative aspect: what is the cost in terms of inflation, and what is the welfare improvement. Many shocks can increase inequality of earnings, and this may (in part) be efficient.

We compare traditional and 'inclusive' policies in terms of their impact on earnings and employment inequality, on the labor market outcomes of lower-productivity, lower-income workers, and in terms of their inflation outcomes. Inclusive policies come at a high cost in terms of inflation, but can substantially reduce the uneven burden of a recessionary shock for the lowest-productivity workers.

Traditional and 'inclusive' policies differ in their impact on the efficiency of the economy - that is, in terms of how much a given policy reduces the inefficient distortions in a market allocation relative to the first best when workers of heterogeneous productivity populate the labor market. We provide a normative assessment, and show that while making monetary policy more inclusive is beneficial for the overall economy, making monetary policy *much* more inclusive results in deviations from the first best allocation.

Our modeling approach allows us to isolate the impact on welfare that comes exclusively through improving the way in which the labor market functions (to this end, it is sufficient to optimally change the incentives of firms in creating vacancies) from the impact that would come from reallocating consumption across high and low-productivity worker. The latter impact is akin to a redistributive fiscal policy. Our results assume that consumption is perfectly insured across all workers. Thus, the inefficiency wedge we identify abstracts from income distribution considerations. Even if one did not care about relative consumption across households, there would still be an argument for making policy more inclusive - albeit by a small amount. Overall, we provide a framework which separates the welfare issues related to the efficiency of the aggregate economy, from the welfare issues related to unequal distribution of consumption.

References

Abowd, J. M., P. Lengermann, and K. L. McKinney (2003): "The measurement of human capital in the US economy," *U.S. Census Bureau - LEHD Program Technical Paper*, TP-2002-09.

Acharya, S., E. Challe, and K. Dogra (2023): "Optimal monetary policy according to HANK," *American Economic Review*, 113(7), 1741–1782.

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

Amberg, N., T. Jansson, M. Klein, and A. Rogantini-Picco (2022): "Five facts about the distributional income effects of monetary policy shocks," *American Economic Review: Insights*, 4(3), 289–304.

Andersen, A. L., N. Johannesen, M. Jørgensen, and J.-L. Peydró (2023): "Monetary policy and inequality," *The Journal of Finance*, 78(5), 2945–2989.

Bergman, N., D. A. Matsa, and M. Weber (2022): "Inclusive monetary policy: How tight labor markets facilitate broad-based employment growth," *National Bureau of Economic Research.*

Bernanke, B. S. (2015): "Monetary Policy and Inequality," *Brookings Commentary.*

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2021): "Inequality, Business Cycles, and Monetary-Fiscal Policy," *Econometrica*, 89(6), 2559–2599.

Board of Governors of the Federal Reserve (2020): "2020 Statement on Longer-Run Goals and Monetary Policy Strategy," *Federal Reserve Statement.*

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A distributional framework for matched employer employee data," *Econometrica*, 87(3), 699–739.

Bosler, C., M. C. Daly, J. G. Fernald, and B. Hobijn (2016): "The outlook for us labor-quality growth," *National Bureau of Economic Research.*

Coglianese, J. M., M. Olsson, and C. Patterson (2023): "Monetary policy and the labor market: a quasi-experiment in sweden," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2023-123).

Coibion, O., Y. Gorodnichenko, L. Kueng, and J. Silvia (2017): "Innocent Bystanders? Monetary policy and inequality," *Journal of Monetary Economics*, 88, 70–89.

Cortes, G. M., and E. Forsythe (2020): "The Heterogeneous Labor Market Impacts of the Covid-19 Pandemic," *Upjohn Institute Working Paper 20-327.*

Dávila, E., and A. Schaab (2023): "Optimal monetary policy with heterogeneous agents: Discretion, commitment, and timeless policy," .

Dolado, J. J., G. Motyovszki, and E. Pappa (2021): "Monetary policy and inequality under labor market frictions and capital-skill complementarity," *American economic journal: macroeconomics*, 13(2), 292–332.

European Central Bank (2014): "Monetary policy and economic inequality," *Keynote speech by Yves Mersch.*

Federal Reserve Bank ofNew York (2021): "Conference: Heterogeneity in Macroeconomics: Implications for Policy," .

Gregory, V., G. Menzio, and D. G. Wiczer (2021): "The Alpha Beta Gamma of the Labor Market," *NBER Working Paper No. 28663.*

Grigsby, J. (2022): "Skill Heterogeneity and Aggregate Labor Market," *NBER Working Paper No. 30052*, pp. 0–83.

Heathcote, J., F. Perri, and G. L. Violante (2020): "The rise of US earnings inequality: Does the cycle drive the trend?," *Review of economic dynamics*, 37, S181–S204.

Ilzetky, E. (2021): "Monetary policy and inequality," *VOX EU Column.*

Lenza, M., and J. Slacalek (2024): "How does monetary policy affect income and wealth inequality? Evidence from quantitative easing in the euro area," *Journal of Applied Econometrics*, 39(5), 746–765.

Mäki-Fränti, P., A. Silvo, A. Gulan, and J. Kilponen (2022): "Monetary policy and inequality: The Finnish case," *Bank of Finland Research Discussion Papers 3/2022.*

McKay, A., and C. K. Wolf (2023): "Monetary policy and inequality," *Journal of Economic Perspectives*, 37(1), 121–144.

Ravenna, F., and C. E. Walsh (2008): "Vacancies, Unemployment, and the Phillips Curve," *European Economic Review*, 52(8), 1494–1521.

Ravenna, F., and C. E. Walsh (2022): "Worker Heterogeneity , Selection , and Unemployment Dynamics in a Pandemic," *Journal of Money, Credit and Banking*, 54, S1(0), 1–43.

Samarina, A., and A. D. Nguyen (2024): "Does monetary policy affect income inequality in the euro area?," *Journal of Money, Credit and Banking*, 56(1), 35–80.

Walsh, C. E. (2005): "Comment on A. Levin, A. Onatski, J. Williams and N. Williams 'Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models'," in *NBER Macroeconomics Annual*, ed. by M. Gertler, and K. Rogoff, vol. 20, pp. 297–308. The MIT Press, Cambridge, MA.

Appendix

8 Model

8.1 Labor flows

The endogenous fraction of existing matches filled by low-productivity workers is ξ_t. Those that survived the exogenous separation hazard ρ_t^x receive a new productivity shock and are retained if and only if $a_{i,t}^l > \bar{a}_t^l$. This occurs with probability $1 - F(\bar{a}_t^l) = 1 - \rho_t^n$. All high-productivity workers surviving the exogenous separation hazard are retained. Thus, actual employment in period t is equal to

$$
\begin{aligned}
N_t &= (1-\rho_t^x)\left[\left(1-\xi_{t-1}\right)+\xi_{t-1}(1-\rho_t^n)\right]N_{t-1}+H_t \\
&= (1-\rho_t^x)\left(1-\xi_{t-1}\rho_t^n\right)N_{t-1}+H_t
\end{aligned}
$$

where H_t equals new hires. ξ_t evolves according to

$$\xi_t = (1-\rho_t^n)\left[\frac{(1-\rho_t^x)\xi_{t-1}N_{t-1}+\gamma_t k_t^w S_t}{N_t}\right], \tag{7}$$

where S_t is the number of job seekers, k_t^w is the fraction who are interviewed, and γ_t is the share of low-productivity workers among S_t. Thus, $\gamma_t k_t^w S_t$ type l workers are interviewed and the fraction $1-\rho_t^n$ have productivity realizations that exceed $\bar{a}_t^l$ and so are hired.

Job seekers at t who are of quality l equal the total number of low-efficiency workers minus the number of matches of quality l that survive the exogenous separation hazard:

$$\gamma_t = \frac{L^l-(1-\rho_t^x)\xi_{t-1}N_{t-1}}{S_t}. \tag{8}$$

The efficiency-weighted average productivity of both employed workers and the pool of job seekers will change over time because γ_t is endogenous and persistent, even though $a_{i,t}^l$ is *i.i.d.*. During recessions, the outflow from employment rises and the inflow into employment falls, resulting in an increase in the average productivity among those still employed and a fall in the average efficiency level of those who are unemployed.

8.2 Households and retail firms

8.2.1 Households

The representative household maximizes

$$\mathrm{E}_t\sum_{i=0}^{\infty}\beta^i\left\{D_t\frac{C_{t+i}^{1-\sigma}}{1-\sigma}-\left[v(h_{t+i}^h)(1-\xi_{t+i})N_{t+i}+\xi_{t+i}N_{t+i}\int_{\bar{a}_t}^1 v(h_{i,t+i}^l)f(a)da\right]\right\}, \tag{9}$$

where $\sigma > 0$ is the coefficient of relative risk aversion, D_t is an aggregate preference shock, $\mathcal{C}_t$ is the sum of a market-purchased composite consumption good C_t and home-produced consumption by unemployed workers $C_t^u = (1 - N_t)w^u$. In (9), the term

$$v(h^h_{t+i})(1 - \xi_{t+i})N_{t+i} + \xi_{t+i}N_{t+i} \int_{\bar{a}_t}^1 v(h^l_{i,t+i})f(a^l)da^l$$

is the disutility to the household of having N_t members working, where hours worked depends on type and the idiosyncratic productivity shocks. We assume $v(h_{t+i}) = \ell h_{t+i}^{1+\chi}/(1 + \chi)$.

Market consumption C_t is a Dixit-Stiglitz composite good consisting of the differentiated products produced by retail firms and is defined as

$$C_t = \left[\int_0^1 c_{k,t}^{\frac{\theta-1}{\theta}} dk \right]^{\frac{\theta}{\theta-1}} \qquad \theta > 0.$$

Given prices $p_{k,t}$ for the final goods, this preference specification implies the household's demand for good k is

$$c_{k,t} = \left(\frac{p_{k,t}}{P_t} \right)^{-\theta} C_t, \tag{10}$$

where the aggregate retail price index P_t is defined as

$$P_t = \left[\int_0^1 p_{k,t}^{1-\theta} dj \right]^{\frac{1}{1-\theta}}.$$

If i_t is the nominal rate of interest, the representative household's first order conditions imply the following must hold in equilibrium:

$$\lambda_t = \beta(1 + i_t)\mathrm{E}_t \left(\frac{P_t}{P_{t+1}} \right) \lambda_{t+1}, \tag{11}$$

where $\lambda_t = D_t \mathcal{C}_t^{-\sigma}$.

8.2.2 Retail firms

There are a continuum of retail firms, indexed by j, who purchase the wholesale good and convert it into differentiated final goods that are sold to households and wholesale firms. Retail firms maximize profits subject to a CRS technology for converting wholesale goods into final goods, the demand functions (10), and a restriction on the frequency with which they can adjust their price. Each period a firm can adjust its price with probability $1 - \omega$. For retail firms, real marginal cost is the price of the wholesale good relative to the price of final output, P_t^w/P_t.

A retail firm k that can adjust its price in period t chooses $p_t(k)$ to maximize

$$\sum_{s=0}^{\infty} (\omega\beta)^s \mathrm{E}_t \left[\left(\frac{\lambda_{t+s}}{\lambda_t} \right) \left(\frac{p_t(k) - P^w_{t+s}}{P_{t+s}} \right) Y_{t+s}(k) \right],$$

subject to

$$Y_{t+s}(k) = Y^d_{t+s}(k) = \left(\frac{p_t(k)}{P_{t+s}} \right)^{-\theta} Y^d_{t+s}, \tag{12}$$

where Y^d_t is aggregate demand for the basket of final goods. The first order condition for those firms adjusting their price in period t is

$$p_t(k)\mathrm{E}_t \sum_{s=0}^{\infty} (\omega\beta)^s \left(\frac{\lambda_{t+s}}{\lambda_t} \right) \left(\frac{p_t(k)}{P_{t+s}} \right)^{1-\theta} Y_{t+s} = \left(\frac{\theta}{\theta-1} \right) \mathrm{E}_t \sum_{s=0}^{\infty} (\omega\beta)^s \left(\frac{\lambda_{t+s}}{\lambda_t} \right) \left(\frac{1}{\mu_{t+s}} \right) \left(\frac{p_t(k)}{P_{t+s}} \right)^{1-\theta} Y_{t+s}.$$

All adjusting firms choose the same reset price p^*_t, and the aggregate price level is then

$$P_t^{1-\theta} = (1-\omega) \left(p^*_t\right)^{1-\theta} + \omega P_{t-1}^{1-\theta}.$$

When linearized around a zero-inflation steady state, these conditions yield a new Keynesian Phillips curve in which the retail price markup

$$\mu_t \equiv \frac{P_t}{P^w_t}$$

is the driving force for inflation; a rise in the markup implies a fall in real marginal costs for retail firms. As in a standard Phillips curve, the elasticity of inflation with respect to real marginal costs will be $\delta \equiv (1-\omega)(1-\beta\omega)/\omega$.

8.3 Match surplus in the competitive equilibrium

We first discuss the derivation of the surplus generated by a producing match of type h and then indicate the modifications needed when considering a type l worker.

8.3.1 Type h workers

A type h worker in a match produces $\phi^h h^h_t$, where ϕ^h is the worker's fixed hourly productivity and h^h_t is hours worked. For such a worker who produces in period t, the surplus is the value of output expressed in terms of retail goods prices $\phi^h h^h_t/\mu_t$ plus the continuation value q^h_t value of the match net of the utility cost of hours $v(h^h_t)/\lambda_t$ and the worker's outside opportunity $w^{u,h}_t$:

$$s^h_t = \left(\frac{\phi^h h^h_t}{\mu_t} \right) + q^h_t - \frac{v(h^h_t)}{\lambda_t} - w^{u,h}_t, \tag{13}$$

where h_t^h is chosen to maximize

$$\left(\frac{\phi^h h_t^h}{\mu_t}\right) - \frac{v(h_t^h)}{\lambda_t}.$$

Equation (13) highlights that the surplus s_t^h is the value in excess of the outside opportunity to the workers and firm. For the firm, the job posting conditions implies the value of a vacancy is zero in equilibrium and so it is the worker's outside opportunity of searching $w_t^{u,h}$ that is relevant. The continuation value q_t^h equals the probability the match survives the exogenous separation hazard times the expected future value of a match plus the probability the match does not survive times the expected future outside opportunity cost:

$$\begin{aligned} q_t^h &= \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) \left[(1-\rho_{t+1}^x)\left(s_{t+1}^h + w_{t+1}^{u,h}\right) + \rho_{t+1}^x w_{t+1}^{u,h}\right] \\ &= \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) \left[(1-\rho_{t+1}^x) s_{t+1}^h + w_{t+1}^{u,h}\right]. \end{aligned} \tag{14}$$

The worker's outside opportunity is

$$w_t^{u,h} = w^u + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) \left[(1-\rho_{t+1}^x) k_{t+1}^w \eta s_{t+1}^h + w_{t+1}^{u,h}\right], \tag{15}$$

where η is the worker's share of the surplus. Subtracting (15) from (14) yields

$$q_t^h - w_t^{u,h} = -w^u + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) (1-\rho_{t+1}^x)\left(1-\eta k_{t+1}^w\right) s_{t+1}^h.$$

Using this in (13),

$$s_t^h = \left(\frac{\phi^h h_t^h}{\mu_t}\right) - \frac{v(h_t^h)}{\lambda_t} - w^u + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) (1-\rho_{t+1}^x)\left(1-\eta k_{t+1}^w\right) s_{t+1}^h. \tag{16}$$

It may be useful to provide an alternative derivation of s_t^h, one that serves to make the timing assumptions of the model more transparent and highlights the separate match valuations of the worker and the firm.

N_t^h workers of type h are employed and producing in period t and $1 - N_t$ are unmatched. The value of being employed is denoted by e_t^h (for employed), and the valuation equation takes the form

$$e_t^h = w_t^h(h_t^h) - \frac{v(h_t^h)}{\lambda_t} + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) \left[(1-\rho_{t+1}^x) e_{t+1}^h + \rho_{t+1}^x k_{t+1}^w e_{t+1}^h + \rho_{t+1}^x \left(1-k_{t+1}^w\right) u_{t+1}^h\right],$$

where $w_t^h(h_t^h)$ is the wage of a type h worker as a function of hours worked, while $v(h_t^h)/\lambda_t$ is the disutility of supplying h_t^h hours, expressed in terms of retail goods. With probability $1-\rho_t^x$ the worker survives the exogenous separation process and produces in period $t+1$. The term $\rho_{t+1}^x k_{t+1}^h e_{t+1}^h$ reflects the assumption that workers exogenously separated can search.

With probability ρ^x_{t+1}, a worker separates, enters the labor market and, with probability k^w_{t+1} obtains an interview and is hired (as no type h workers are screened out). With probability $1-k^w_{t+1}$, the exogenously separated workers do not make a match.

Denote by u^h_t the value of being unemployed. The valuation equation for u^h_t, is

$$u^h_t = w^u + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left[k^w_{t+1} e^h_{t+1} + \left(1 - k^w_{t+1}\right) u^h_{t+1} \right],$$

where w^h is any unemployment benefit or value of home production. The unemployed worker finds a match with probability k^w_{t+1} and receives e^h_{t+1}, and with probability $1-k^w_{t+1}$ the worker remains unmatched.

The surplus to a type h worker of being in a match is

$$\begin{aligned} e^h_t - u^h_t &= \left[w^h_t(h^h_t) - w^h - \frac{v(h^h_t)}{\lambda_t} + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left(1 - \rho^x_{t+1} + \rho^x_{t+1} k^w_{t+1}\right) e^h_{t+1} + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \rho^x_{t+1} \left(1 - k^w_{t+1}\right) \right. \\ &\quad - \left[\beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) k^w_{t+1} e^h_{t+1} + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left(1 - k^w_{t+1}\right) u^h_{t+1} \right] \\ &= w^h_t(h^h_t) - \frac{v(h^h_t)}{\lambda_t} - w^h + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left(1 - \rho^x_{t+1}\right) \left(1 - k^w_{t+1}\right) \left(e^h_{t+1} - u^h_{t+1}\right). \end{aligned}$$

From the job-posting condition, the value of a vacancy is driven to zero, so the value J^h_t to a firm of a filled job with a type h worker is

$$J^h_t = \left(\frac{\phi^h h^h_t}{\mu_t} \right) - w^h_t(h^h_t) + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left(1 - \rho^x_{t+1}\right) J^h_{t+1}.$$

The joint surplus of a match is therefore

$$\begin{aligned} s^h_t &= J^h_t + e^h_t - w^u_t = \left(\frac{\phi^h h^h_t}{\mu_t} \right) - \frac{v(h^h_t)}{\lambda_t} - w^h + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left(1 - \rho^x_{t+1}\right) J^h_{t+1} \\ &\quad + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left(1 - \rho^x_{t+1}\right) \left(1 - k^w_{t+1}\right) \left(e^h_{t+1} - u^h_{t+1}\right), \end{aligned}$$

where, again, h^h_t is chosen to maximize output net of the disutility of supplying hours. With Nash bargaining, the worker's share is equal $e^h_{t+1} - u^h_{t+1} = \eta s^h_{t+1}$, and the firm's share is $J^h_{t+1} = (1-\eta) s^h_{t+1}$. Hence,

$$\begin{aligned} s^h_t &= \left(\frac{\phi^h h^h_t}{\mu_t} \right) - \frac{v(h^h_t)}{\lambda_t} - w^h + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left(1 - \rho^x_{t+1}\right) (1-\eta) s^h_{t+1} \\ &\quad + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left(1 - \rho^x_{t+1}\right) \left(1 - k^w_{t+1}\right) \eta s^h_{t+1}, \end{aligned}$$

or

$$s_t^h = \left(\frac{\phi^h h_t^h}{\mu_t}\right) - \frac{v(h_t^h)}{\lambda_t} - w^h + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) \left(1-\rho_{t+1}^x\right) \left(1-\eta k_{t+1}^w\right) s_{t+1}^h,$$

which is the same as (16).

8.3.2 Type l workers

For type l workers, the endogenous separation hazard and the stochastic productivity shock must be taken into account. Because workers differ, it is now also necessary to start with the joint surplus of a match with worker i of type l with current productivity $a_{i,t}^l \phi^l$:

$$s_{i,t}^l = \left(\frac{a_{i,t}^l \phi^l h_{i,t}^l}{\mu_t}\right) - \frac{v(h_{i,t}^l)}{\lambda_t} + q_t^l - w_{i,t}^{u,l},$$

where h_t^l will vary with $a_{i,t}^l$ and is chosen to maximize

$$\left(\frac{a_{i,t}^l \phi^l h_{i,t}^l}{\mu_t}\right) - \frac{v(h_{i,t}^l)}{\lambda_t}.$$

In place of (14) and (15), one now has

$$\begin{aligned} q_t^l &= \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) \left[\left(1-\rho_{t+1}^x\right)\left(1-\rho_{t+1}^n\right)\left(s_{t+1}^i + w_{t+1}^{u,l}\right) + \left[\rho_{t+1}^x + \left(1-\rho_{t+1}^x\right)\rho_{t+1}^n\right] w_{t+1}^{u,l}\right] \\ &= \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) \left[\left(1-\rho_{t+1}^x\right)\left(1-\rho_{t+1}^n\right) s_{t+1}^l + w_{t+1}^{u,l}\right]. \end{aligned}$$

where $\rho_{t+1}^{n,l} = F(\bar{a}_t^l)$ is the probability the worker's productivity realization falls below the threshold level $\bar{a}_t^l$, where $F(.)$ is the cumulative distribution of the idiosyncratic productivity shock. Notice that the expectation incorporates the probability of surviving the exogenous separation but then receiving a low productivity draw and experiencing endogenous separation. This occurs with probability $\left(1-\rho_{t+1}^x\right)\rho_{t+1}^n$. Because idiosyncratic productivity shocks are i.i.d., the expected future value of s_{t+1}^l is independent of i. Similarly, the worker's outside opportunity is

$$w_t^{u,l} = w^u + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) \left[\left(1-\rho_{t+1}^x\right)\left(1-\rho_{t+1}^n\right) k_{t+1}^w \eta s_{t+1}^l + w_{t+1}^{u,l}\right],$$

where the assumption of Nash bargaining as been used. Combining these equations for $s_{i,t}^l$, q_t^l and $w_t^{u,l}$ yields

$$s_{i,t}^l = \left(\frac{a_{i,t}^l \phi^l h_{i,t}^l}{\mu_t}\right) - \frac{v(h_{i,t}^l)}{\lambda_t} - w^u + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) \left(1-\rho_{t+1}^x\right)\left(1-\rho_{t+1}^n\right)\left(1-\eta k_{t+1}^w\right) s_{t+1}^l. \quad (17)$$

Equation (17) can also be derived by using the valuations $e_{i,t}^{l}$, u_{t+1}^{l} and $J_{i,t+1}^{h}$ as previously done for a type h worker.

8.3.3 Optimal hours

Since hours will be chosen to maximize the surplus of a match, $v'(h_t^h)/\lambda_t = \phi^h/\mu_t$ for a type h worker, which, using the functional form for the disutility of leisure implies

$$\frac{v'(h_t^h)}{\lambda_t} = \frac{\ell\left(h_t^h\right)^{\chi}}{\lambda_t} = \frac{\phi^h}{\mu_t} \Rightarrow h_t = \left(\frac{\lambda_t \phi^h}{\ell \mu_t}\right)^{\frac{1}{\chi}}. \tag{18}$$

Hence,

$$\frac{v(h_t)}{\lambda_t} = \frac{1}{\lambda_t}\frac{\ell\left(h_t^h\right)^{1+\chi}}{1+\chi} = \frac{1}{\lambda_t}\frac{\ell\left(h_t^h\right)^{\chi}}{1+\chi}h_t^h = \frac{1}{1+\chi}\frac{\phi^h h_t^h}{\mu_t},$$

and

$$\frac{\phi^h h_t^h}{\mu_t} - \frac{v(h_t^h)}{\lambda_t} = \frac{\phi^h h_t^h}{\mu_t} - \frac{1}{1+\chi}\frac{\phi^h h_t^h}{\mu_t} = \left(\frac{\chi}{1+\chi}\right)\frac{\phi^h h_t^h}{\mu_t},$$

or,

$$\frac{\phi^h h_t^h}{\mu_t} - \frac{v(h_t^h)}{\lambda_t} = \left(\frac{\chi}{1+\chi}\right)\frac{\phi^h h_t^h}{\mu_t} = \left(\frac{\chi}{1+\chi}\right)\left(\frac{\lambda_t}{\ell}\right)^{\frac{1}{\chi}}\left(\frac{\phi^h}{\mu_t}\right)^{\frac{1+\chi}{\chi}} \tag{19}$$

For type l workers,

$$\frac{a_{i,t}^l \phi^l h_{i,t}^l}{\mu_t} - \frac{v(h_{i,t}^l)}{\lambda_t} = \left(\frac{\chi}{1+\chi}\right)\left(\frac{\lambda_t}{\ell}\right)^{\frac{1}{\chi}}\left(\frac{a_{i,t}^l \phi^l}{\mu_t}\right)^{\frac{1+\chi}{\chi}}. \tag{20}$$

8.3.4 Cutoff productivity

The match with a type l worker ends, of an interview with a type l worker fails to result in a match if the joint surplus is non-positive. This occurs for a worker with idiosyncratic productivity $a_{i,t}^l$ if $s_{i,t}^l \leq 0$, or when $a_{i,t}^l \leq \bar{a}_t^l$, where

$$\bar{a}_t^l = \left(\frac{\mu_t}{\phi^l}\right)\left[\left(\frac{1+\chi}{\chi}\right)\left(\frac{\ell}{\lambda_t}\right)^{\frac{1}{\chi}}\left(w_t^{u,l} - q_t^l\right)\right]^{\frac{\chi}{1+\chi}}.$$

The RHS does not depend on the firms, implying all firms employ the same cutoff productivity level for hiring and retention decision.

9 The social planner's problem

The planner's problem is to maximize

$$\mathrm{E}_t \sum_{i=0}^{\infty} \beta^i \delta_t^i \left[D_t \frac{\mathcal{C}_{t+i}^{1-\sigma}}{1-\sigma} - (1-\xi_{t+i}) N_{t+i} v(h_{t+i}^h) - \xi_{t+i} N_{t+i} \int_{\bar{a}_t^l}^1 v(h_{i,t+i}^l) f(a) da \right]$$

subject to

1. Labor market flows:

$$N_t = (1-\rho_t^x)\left(1-\xi_{t-1}\rho_t^n\right) N_{t-1} + H_t,$$

   where

$$\xi_t = (1-\rho_t^n)\left[\frac{(1-\rho_t^x)\xi_{t-1}N_{t-1} + \gamma_t k_t^w S_t}{N_t}\right],$$

$$S_t = 1-(1-\rho_t^x)\, N_{t-1},$$

$$I_t = \psi V_t^{1-a} S_t^a, \qquad 0 < \alpha < 1,\, \psi > 0,$$

$$H_t = (1-\gamma_t) k_t^w S_t + \gamma_t (1-\rho_t^n)\, k_t^w S_t = (1-\gamma_t \rho_t^n)\, k_t^w S_t,$$

$$\gamma_t \equiv \frac{S_t^l}{S_t} = \frac{L^l - (1-\rho_t^x)\xi_{t-1}N_{t-1}}{S_t}.$$

2. Production technology:

$$Y_t = \left\{ (1-\xi_t)\phi_t^h h_t^h + \xi_t \phi^l \left[\frac{\int_{\bar{a}_t^l}^1 a_{i,t}^l h_{i,t}^l dF_l(a_i)}{1-F_l(\overline{a}_t^l)} \right] \right\} N_t.$$

3. Goods clearing:

$$\phi^h h_t^h N_t^h + \phi^l \int_{\bar{a}_t^l}^1 a_{i,t}^l h_{i,t}^l di + w^u \left(L - N_t^h - N_t^l\right) - \kappa V_t - \mathcal{C}_t L \geq 0$$

$$\mathcal{C}_t = C_t + w^u \left(L - N_t^h - N_t^l\right)$$

$$Y_t = C_t + \kappa V_t,$$

In deriving the solution, let A_t^j be the set of employed type j workers and let $\#N_t^j$ be their number. In addition, let $\bar{a}_t^{sl}$ denote the threshold productivity level for type l workers in the social planners solution.

The social planner's problem is described by

$$W_t(N_{t-1}^h, N_{t-1}^l) = \max_{\mathcal{C}_t, C_t, V_t, I_t, \bar{a}_t^l, N_t^h, \#N_t^l, h_t^h, h_{i,t}^l} \left[\begin{array}{c} U(\mathcal{C}_t)L - \int_{i\in A_t^h} v\left(h_{i,t}^h\right) di \\ - \int_{i\in A_t^l} v\left(h_t^l\right) di + \beta \mathrm{E}_t W_{t+1}(N_t^h, N_t^l) \end{array} \right]$$

subject to

$$\phi^h \int_{i\in A_t^h} h_t^h di + \phi^l \int_{i\in A_t^l} a_{i,t}^l h_{i,t}^l di - \kappa V_t - C_t L \geq 0$$

$$C_t L + w^u \left(L - N_t^h - N_t^l\right) - \mathcal{C}_t L \geq 0$$

$$\left[\left(1-\rho_t^x\right) N_{t-1}^h + \left(1-\gamma_t\right) I_t\right] - N_t^h = 0$$

$$\left[1 - F(\bar{a}_t^{sl})\right]\left[\left(1-\rho_t^x\right) N_{t-1}^l + \gamma_t I_t\right] - N_t^l = 0$$

$$\psi V_t^{1-\alpha}\left[1 - \left(1-\rho_t^x\right)\left(N_{t-1}^h + N_{t-1}^l\right)\right]^\alpha - I_t = 0$$

where

$$\gamma_t \equiv \frac{L^l - (1-\rho_t^x) N_{t-1}^l}{\left[1-\left(1-\rho_t^x\right)\left(N_{t-1}^h + N_{t-1}^l\right)\right]} = \frac{S_t^l}{S_t^h + S_t^l}.$$

Define $\lambda_{1,t}$, $\lambda_{2,t}$ τ_t^h, $\tau_{i,t}^l$, and ζ_t as the Lagrangian multipliers on these five constraints. The first-order conditions take the form:

$$\mathcal{C}_t\text{: } U'(\mathcal{C}_t)L - \lambda_{2,t}L = 0$$

$$C_t\text{: } -\lambda_{1,t}L + \lambda_{2,t}L = 0 \Rightarrow \lambda_{1,t} = \lambda_{2,t} = U'(\mathcal{C}_t)$$

$$V_t\text{: } -\kappa\lambda_{1,t} + \psi\left(1-\alpha\right)\left(\frac{I_t}{V_t}\right)\zeta_t = 0$$

$$I_t\text{: } \left(1-\gamma_t\right)\tau_t^h + \gamma_t\left[1-F(\bar{a}_t^{sl})\right]\tau_t^l - \zeta_t = 0$$

$$h_t^h\text{: } -v'(h_t^h) + \lambda_{1,t}\phi^h = 0 \Rightarrow \frac{v'(h_t^h)}{U'(\mathcal{C}_t)} = \phi^h$$

$$h_{i,t}^l, \text{ for } a_{i,t}^l \geq \bar{a}_t^{sl}\text{: } -v'\left(h_{i,t}^l\right) + \lambda_{1,t}\phi^l a_{i,t}^l = 0 \Rightarrow \frac{v'(h_{i,t}^l)}{U'(\mathcal{C}_t)} = a_{i,t}^l\phi^l$$

$$N_t^h\text{: of type } h\text{: } -v(h_t^h) + \lambda_{1,t}\phi^h h_t^h - \lambda_{2,t}w^u - \tau_t^h + \beta \mathrm{E}_t \frac{\partial W_{t+1}(N_t^h, N_t^l)}{\partial N_t^h} = 0$$

$$N_t^l \text{ of type } l\text{: } -v(h_{i,t}^l) + \lambda_{1,t}a_{i,t}^l\phi^l h_{i,t}^l - \lambda_{2,t}w^u - \tau_{i,t}^l + \beta \mathrm{E}_t \frac{\partial W_{t+1}(N_t^h, N_t^l)}{\partial N_t^l} = 0$$

Let $x_t \equiv (1-\rho_t^x)\left\{\tau_t^h - \left[1-F(\bar{a}_t^{sl})\right]\tau_t^l\right\}$.Define $\bar{s}_t^j \equiv \tau_t^j/\lambda_{1,t} = \tau_t^j/U'(\mathcal{C}_t)$ and $X_t \equiv x_t/\lambda_{1,t}$. Then Ravenna and Walsh (2021) shows that:

$$\begin{aligned}\bar{s}_t^h \equiv & \frac{\tau_t^h}{\lambda_{1,t}} = \left[\phi^h a_{i,t}^h h_{i,t}^h - \frac{v(h_{i,t}^h)}{U'(\mathcal{C}_t)} - w^u\right] + \beta\left(1-\rho_{t+1}^x\right)\mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(1-\alpha k_{t+1}^w\right)\bar{s}_{t+1}^h \\ & -\beta(1-\alpha)\mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\gamma_{t+1}k_{t+1}^w X_{t+1}, \end{aligned} \tag{21}$$

and

$$\begin{aligned}\bar{s}_{i,t}^l \equiv & \frac{\tau_{i,t}^l}{\lambda_{1,t}} = \left[\phi^l a_{i,t}^l h_{i,t}^l - \frac{v(h_{i,t}^l)}{U'(\mathcal{C}_t)} - w^u\right] + \beta\left(1-\rho_{t+1}^x\right)\mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(1-\alpha k_{t+1}^w\right)\left[1-F(\bar{a}_{t+1}^{sl})\right]\bar{s}_{t+1}^l \\ & +\beta(1-\alpha)\mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(1-\gamma_{t+1}\right)k_{t+1}^w X_{t+1}. \end{aligned} \tag{22}$$

9.1 Efficiency

To assess efficiency, it is necessary to compare equations (16) and (17) from the competitive market equilibrium with equations (21) and (22) from the social planner's allocation. Imposing the conditions $\mu_t = 1$, $\eta = \alpha$ that ensure the competitive market equilibrium is efficient in a new Keynesian model with homogeneous labor in a search and matching model, (16) and (17) become

$$s_t^h = \left[\phi^h h_t^h - \frac{v(h_t^h)}{\lambda_t} - w^u\right] + \beta\mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(1-\rho_{t+1}^x\right)\left(1-\alpha k_{t+1}^w\right)s_{t+1}^h. \tag{23}$$

$$s_{i,t}^l = \left[a_{i,t}^l \phi^l h_{i,t}^l - \frac{v(h_{i,t}^l)}{\lambda_t} - w^u\right] + \beta\mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(1-\rho_{t+1}^x\right)\left(1-F(\bar{a}_{t+1}^l)\right)\left(1-\alpha k_{t+1}^w\right)s_{t+1}^l. \tag{24}$$

In the presence of labor heterogeneity, the social planner's valuation of matches given by (21) and (22) differ from (23) and (24) due to the addition terms in (21) and (22) that involve X_{t+1}. This term measures the difference in the expected future value of a type h worker over a type l worker. It captures the composition externality that arises in the presence of heterogeneous labor; employment and separation decisions by firms ignore the impact their decisions have on the composition of the pool of unemployed workers. The social planner takes this externality into account. That is, evaluated at the same allocation,

$$\begin{aligned} s_t^h - \bar{s}_t^h = & \beta\left(1-\rho_{t+1}^x\right)\mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(1-\alpha k_{t+1}^w\right)\left(s_{t+1}^h - \bar{s}_{t+1}^h\right) \\ & +\beta(1-\alpha)\mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\gamma_{t+1}k_{t+1}^w X_{t+1} \end{aligned} \tag{25}$$

and

$$s_{i,t}^{l}-\bar{s}_{i,t}^{l} = \beta\left(1-\rho_{t+1}^{x}\right)\mathrm{E}_{t}\left(\frac{\lambda_{t+1}}{\lambda_{t}}\right)\left(1-\alpha k_{t+1}^{w}\right)\left[1-F(\bar{a}_{t+1}^{sl})\right]\left(s_{t+1}^{l}-\bar{s}_{t+1}^{l}\right) -\beta\left(1-\alpha\right)\mathrm{E}_{t}\left(\frac{\lambda_{t+1}}{\lambda_{t}}\right)\left(1-\gamma_{t+1}\right)k_{t+1}^{w}X_{t+1}, \tag{26}$$

where

$$X_{t+1}\equiv\left(1-\alpha\right)\left(1-\rho_{t+1}^{x}\right)k_{t+1}^{w}\left[\bar{s}_{t+1}^{h}-\left(1-\rho_{t+1}^{n}\right)\bar{s}_{i,t+1}^{l}(a_{i,t+1}^{l})\right]. \tag{27}$$

Because $X_{t+i}\geq 0$, all else equal, type h workers are overvalued and type l workers are undervalued in the competitive equilibrium.

Both $s_{i,t}^{l}$ and $\bar{s}_{i,t}^{l}$ are increasing functions of the idiosyncratic productivity of worker i. Since private matches involving type l workers end whenever $a_{i,t}^{l}<\bar{a}_{t}^{l}$, where $\bar{a}_{t}^{l}$ is defined such that $s_{i,t}^{l}(\bar{a}_{t}^{l})=0$, the fact that $\bar{s}_{i,t}^{l}\left(a_{i,t}\right)>s_{i,t}^{l}\left(a_{i,t}\right)$ in the face of labor heterogeneity (i.e., $\mathrm{E}_{t}\left(\lambda_{t+1}/\lambda_{t}\right)\left(1-\gamma_{t+1}\right)X_{t+1}>0$ in (26) when $0<\gamma_{t+1}<1$) implies

$$\bar{s}_{i,t}^{l}\left(\bar{a}_{t}^{l}\right)>s_{i,t}^{l}\left(\bar{a}_{t}^{l}\right)=0. \tag{28}$$

That is, a type l worker who generates a zero surplus in the competitive equilibrium would, from the perspective of the social planner, still generate a positive surplus. Therefore, the cutoff productivity level in the efficient allocation (i.e., $\bar{a}_{t}^{sp}$ such that $\bar{s}_{t}^{l}(\bar{a}_{t}^{sp})=0$) is less that $\bar{a}_{t}^{l}$. Some type l workers who experience endogenous separation and become unemployed in the competitive equilibrium would remain employed by the social planner. Similarly, some unemployed type l workers who obtain interviews but are screened out in the competitive equilibrium would be hired by the social planner. This also translates into a higher share of low-efficiency workers among the unemployed and a lower expected benefit to posting vacancies in the competitive equilibrium.

10 Model Parameterization

10.1 Steady state Parameters

Six parameters are key to determining the impact of labor-productivity heterogeneity on the model economy. These parameter are the value of home production w^{u}, the coefficient ℓ scaling the disutility of labor hours, the cost of vacancy posting κ, the productivity of the matching technology ψ, the relative steady-state productivity of high- to low-efficiency workers $\phi^{h}/\left(\phi^{l}\int_{0}^{1}a_{i}^{l}dF(a_{i}^{l})\right)$ and the labor force share of low-efficiency workers $\bar{\gamma}$. Values for these parameters are selected by jointly targeting six steady-state values: the steady-state aggregate unemployment rate U_{ss} and the unemployment rates U_{ss}^{l} and U_{ss}^{h} for each worker type, average hours per worker h_{ss}^{av}, vacancy posting costs κV_{ss} as a share of output, and the probability k_{ss}^{f} of a vacancy match with a job application. The target steady state values and the implied parameters are reported in Table 1.

U_{ss} is set to the average U.S. civilian unemployment rate over 1948:Q4 to 2019:Q4. Neither U^l_{ss} nor U^h_{ss} are directly observable, so our baseline parameterization follows Gregory, Menzio, and Wiczer (2021) who use the Longitudinal Employer and Household Dynamics (LEHD) data from 1997 to 2014 to estimate the labor market shares and unemployment rates for three separate worker types that differ by employment duration. We map these estimates into our model with two types of workers, implying a share of low-efficiency workers in the labor force of 38% with an average unemployment rate of 9.87%, and a share of high-efficiency workers of 62%, with an average unemployment rate of 2.97%.

This baseline parameterization implies an unemployment rate ratio U^l_{ss}/U^h_{ss} equal to 3.3. Given that U^h_{ss} and U^l_{ss} are not observable, Ravenna and Walsh (2021) also considered alternatives that resulted in a higher value of 4.2 and a lower value of 2.5 for this ratio, keeping U_{ss} constant at 5.6%. The higher value of U^l_{ss}/U^h_{ss} comes from an alternative mapping of the three labor types in Gregory, Menzio, and Wiczer (2021) into our model with two labor types that implies $\bar{\gamma} = 20\%$, $U^h_{ss} = 2.1\%$ and $U^l_{ss} = 14.4\%$. The lower value of U^l_{ss}/U^h_{ss} is based on the observed difference in average unemployment rate for workers aged 16 to 24 (4.4%) and over-24 (11.6%). Our general conclusions are robust to these alternative parameterizations.

The other targeted moments reported in Table 1 include steady-state hours per worker h^{av}_{ss}, the steady-state aggregate separation rate ρ_{ss} and the probability of a match between an applicant and a vacancy k^f_{ss}. These values are parameterized to standard values in U.S. business cycle literature. The share of output devoted to hiring activities is in line with empirical evidence reported in Ravenna and Walsh (2008). For parameters standard to new Keynesian models, we adopt values common in the literature, and for the benchmark monetary policy rule, we use a standard value of 1.5 for the response coefficient on inflation. We assume the i.i.d. productivity shock $a^l_{i,t}$ has a uniform distribution with support $(0, 1]$.

In the simulations, we follow the standard approach in NK models of assuming the existence of a tax/subsidy system that ensures the steady-state allocations in the planner's problem and the market allocation coincide. This requires that the Hosios condition holds and that the study-state distortions due to imperfect competition and the selection distortion are eliminated.

10.2 Shocks

Our simulated paths are hypothetical paths in the absence of the effects of the vast income support measures (the CARES act) and expansionary monetary policies (discretionary cuts in the federal funds rate, forward guidance, and quantitative easing measures) which are outside our model and that affected the actual path of the economy within the first months of the pandemic. Our target output path is based on the 2020Q2 vintage of the Survey of Professional Forecasters (SPF) for GDP growth, relative to a potential growth path of 2% per year. The SPF predicts GDP growth only up to 5 quarters ahead. For the subsequent 5 quarters, we assume the economy recovers at the same rate as over the forecast horizon, with the shortfall in output dissipating at an estimated quarter-over-quarter rate of 67.4%. The SPF does not provide a forecast for total separations, so our target path for separations is obtained by measuring the 2020Q2 increase in the cumulated CPS transition rate from employment to non-employment over each month, as a share of the previous month employment stock, relative to the transition rate averaged over the 2018-2019 period.[8] The path for total separations then assumes the initial rise reverts back to steady state at the same rate as the SPF output forecast.[9]

---

[8] Considering the ratio of total transitions over a quarter relative to the average number employed over the quarter returns a similar increase in transition rates for 2020Q2.

[9] The SPF provides a forecast for unemployment. Using as target variables the 2020Q2 vintage of unemployment and output, the model-implied paths for the ρ_t^x and D_t shocks that are similar to the our baseline and has limited impact on the results. We prefer our specifications, which allows the model to produce a conditional forecast for unemployment – which turns out to be very close to the SPF forecast.