Competitive Search with Repeated Moral Hazard*

Fei Zhou

Hong Kong Baptist University

Abstract

This paper studies the effects of moral hazard on employment and wage dynamics in a continuous-time competitive search model. Unobservable idiosyncratic factors require employers to design dynamic contracts to incentivize workers to exert effort. A higher level of public aggregate productivity reduces the relative importance of unobservable idiosyncratic factors, which facilitate the detection of shirking and relaxes firms' incentive constraint. This channel amplifies the elasticity of vacancy to changes in aggregate productivity and induces time-varying schemes in wage compensation. When the informational friction is disciplined by the moment on individual workers' performance pay in PSID, the model produces significantly higher unemployment volatility than that without the moral hazard problem. In addition, the model generates endogenous counter-cyclical wage distributions.

Keywords: Moral Hazard, Dynamic Contract, Search and Matching Friction, Unemployment Volatility, Wage Dispersion

JEL classifications: E240, E250, E320, C730, J330

*I would like to thank Pengfei Wang, Yuliy Sannikov and Yang Lu for their invaluable advice. I am particular grateful for the insightful comments from Ji Huang and Yuzhe Zhang. I am thankful to Zhen Huo, Yan Ji, Shenzhe Jiang, Xuan Li, Zheng Liu, Jonathan Payne, José-Víctor Ríos-Rull, Janghoon Shon, and audiences in HKUST macro group, HKBU, Jinan University, SUFE for their feedback. I also acknowledge the fellowship support from HKUST, HKPFS and RGC ECS project # 22611821.

1. Introduction

Two types of frictions are pervasive in the labor market: search frictions that prevent full employment and moral hazard that prevents efficient production. The literature has made significant progress in understanding the effects of each of these frictions in isolation, while much less is known about how their joint presence shapes labor market outcomes. In this paper, we provide a quantitative evaluation of this question.

We build a competitive search model with aggregate productivity shocks. In contrast to a canonical search model, the output produced by a firm-worker pair depends on three components: a public observed productivity shock, unobserved efforts by workers that complement with the productivity shock, and an unobservable idiosyncratic factor. The difficulty in separating the latter two components is the root of the moral hazard problem. Instead of offering a constant wage plan which results in shirking, firms are better off by providing a dynamic wage contract to incentivize workers to exert effort. Solving for the optimal contract can be technically demanding, as the history of firms' outputs and workers' distribution amount to potential state variables. To make this model tractable, we combine the continuous-time approach to address the principle-agent problem (Sannikov, 2008) with block recursive equilibrium (Menzio and Shi, 2010). The former permits a recursive formulation of the optimal contract that can be characterized by a convenient ordinary differential equation, and the latter allows us to simply trace the aggregate productivity without worrying about other aspects of the economy. Thanks to this tractability, we can explore the model's implications for employment volatility and wage dynamics.

Qualitatively, three main findings stand out when the moral hazard issue is taken into account, all of which help to bring the model's predictions closer to the data. First, the employment rate is more responsive to the aggregate productivity shock. To see why this is the case, consider the effects of a positive productivity shock. Ceteris paribus, higher productivity increases the production efficiency. More importantly, it also enlarges the role of workers' effort in determining firms' output relative to idiosyncratic factors. In other words, any deviation of effort from its desired level becomes easier to detect. As a result, firms' incentive constraint is relaxed, which leads to higher profit and more vacancy posting. The counter-cyclical information rent amplifies the effects of productivity shocks on employment. In contrast, in an environment where workers' effort is observable, higher productivity still improves production efficiency, but the additional incentive channel is muted.

Second, the wage dispersion is endogenously counter-cyclical. In the absence of moral hazard, a risk-neutral firm finds it optimal to offer a constant compensation plan such that perfect risk sharing is achieved for risk-averse workers. This plan becomes suboptimal when effort is unobservable as workers will always choose to shirk. To provide incentive, it is necessary to vary compensation with output which exposes workers to a certain level of risk. Related to the previous discussion, with a higher productivity, the more relaxed incentive constraint also reduces firms' need to expose workers to risk. Consequently, compensation is less volatile over time and less cross-sectionally dispersed.

Third, the individual wage dynamics feature a scarring effect of unemployment, a slow recovery from the low compensation trap and strong persistence in the labor income. Similar to DeMarzo, Fishman, He, and Wang (2012) and Bolton, Chen, and Wang (2011), the optimal contract requires the average compensation to be back-loaded to provide dynamic incentives. A typical worker's wage path grows with her job tenure, as if reputation is accumulated gradually. When separated from a job, the existing reputation disappears, and a worker has to rebuild her reputation from scratch after re-employment. This property helps rationalize the documented scarring effect. In addition, the dynamic contract induces endogenous persistence of compensation in response to transitory idiosyncratic factors due to the fact that a workers' outside option is a reflective boundary of the promised utility. When the promised utility drifts towards this bound, it immediately leaves this bound but will return to it frequently whenever a bad idiosyncratic factor realizes again (Grochulski and Zhang, 2019). Therefore, a worker may be trapped at a low compensation regime for a relatively long time.

To quantify these effects, it is crucial to discipline the underlying information friction faced by firms. In our setup, this friction is parameterized by the variance of unobserved idiosyncratic factors. Our calibration strategy is to match the moments related to the performance-pay residual in the Panel Study of Income Dynamics (PSID) after controlling for observed workers' characteristics. We focus on the performance pay instead of the regular wage as the former is more directly related to firms' needs to provide incentives and it mitigates the concern that firms may have information advantage on firm-worker pairs over econometricians when setting wages. The variance of the idiosyncratic factor is then recovered by the method of indirect inference: we choose the parameters such that when running the same regression using model-generated data, the moments related to the performance-pay residual is the same as that obtained using PSID data.

We first conduct comparative statics analysis by comparing unemployment rates across different steady-state productivity levels. Under our baseline calibration, the unemployment rate elasticity with respect to aggregate productivity is -0.27, which is 3 times higher than that in the environment without the moral hazard issue. With the moral hazard problem, the response of unemployment can be decomposed to an efficiency component and an incentive component. It turns out that the quantitative importance of these two channels are comparable to each other. With recurrent productivity shocks, the model can account for about 50% the observed unemployment rate volatility, which is significantly higher than that in the model without informational frictions.

The baseline model also generates strong counter-cyclical wage dispersion: the correlation between aggregate output and wage dispersion is -0.34. This is consistent with the empirical findings in Storesletten, Telmer, and Yaron (2004). Notably, the distribution of a primitive idiosyncratic factor across business cycles remains invariant, whereas the interaction between aggregate economic conditions and the optimal contract endogenously shifts the wage distribution.

In terms of individual wage dynamics, the model predicts a significant wage scarring effect. The first-year wage payment in the new job spell is 54% less than the most recent payment from the last job. This scarring effect of unemployment is also broadly consistent with Jacobson et al. (1993), Couch and Placzek (2010), and Barnette and Michaud (2011), who document that wages decline by 15% – 40% after job displacement. Second, the model features a lingering effect for low-income workers. Conditional on hitting the lower bound of wage payment in the contract, it takes a worker 8.7 quarters to recover to her economy average wage level. This helps to explain why the model-implied income Gini coefficient is 0.22 despite that workers are homogeneous ex ante.

The presence of moral hazard also has important implications for government policies. We conduct two counterfactual exercises based on our calibrated model: an increase of unemployment benefits and an increase of the minimum wage. With a higher level of unemployment benefit, unemployed workers are better off, which in turn improves the outside option of employed workers. Not surprisingly, this change limits firms' ability to punish their employees and tightens their incentive constraint, which results in a higher unemployment rate in the long run. The dynamic contract setup allows us to further inspect the heterogeneous impacts on workers' compensation. We find that the compensation for junior workers with a short tenure reduces most, while workers whose tenure are longer than three years are affected much more mildly.

In the second policy experiment, minimum wage is imposed. As noted above, when moral hazard is absent, firms would like to pay a constant compensation that depends only on the average output during the job spell. The minimum wage requirement is therefore irrelevant, provided that the constant compensation is large enough. With moral hazard, the history of output matters, and the minimum wage constraint will be binding for a nonnegligible fraction of workers. We show that the unemployment rate is highly sensitive to the minimum wage, and the incentive channel plays an important role in driving this result. Furthermore, two competing forces shape the compensation schedule: higher minimum wage tend to increase the compensation mechanically, while it reduces back-loaded reward for reputation. These two forces leave the compensation for workers with short tenure increasing, while the compensation for workers with long tenure dampened.

Related literature. This paper contributes to the growing literature that analyzes how micro-level moral hazard effects transmit to macro-level business cycle fluctuations. It is most related to studies on dynamic contract and labor market frictions.

Our work builds directly on the continuous-time optimal contract literature. We extend the principal-agent problem in Sannikov (2008) to a general equilibrium framework, where workers' outside option and the starting promised utility are determined endogenously. Moreover, we allow for interaction between aggregate productivity shocks and workers' incentive constraint, which is necessary to understand the problem's business cycle implications. From a technical perspective, we incorporate the advance in Grochulski and Zhang (2019), which allows for a more flexible arrangement when the promised utility reaches the worker's outside option. It leads to more realistic income inequality due to a slow reflection at the lower end. Compared with the discrete-time dynamic contract model pioneered by Spear and Srivastava (1987),Holmstrom and Milgrom (1987) and Phelan and Townsend (1991), the continuous-time method is more tractable, for example DeMarzo and Sannikov (2006) and DeMarzo, Fishman, He, and Wang (2012) are some of its applications in macro finance.

Our work is closely related to Moen and Rosén (2011), who also study how incentive contracts affect labor search outcomes. We share the similarity that counter-cyclical information rents help to amplify unemployment volatility. However, our paper differs from theirs in several important ways: first, in Moen and Rosén (2011), the unobservable idiosyncratic factor remains constant during a job spell. As a result, the optimal wage contract is constant over time. Our model complements their work by allowing the unobservable idiosyncratic factor to be stochastic, and our optimal contract is a long-term contract that provides dynamic incentives. The recurrent idiosyncratic factor makes our model easier to connect to micro-level data to quantify the underlying information frictions. Second, while Moen and Rosén (2011) focus on steady-state analysis, our model permits aggregate productivity shocks, which makes it possible to directly examine unemployment volatility. Third, the dynamic contract with recurrent idiosyncratic factors and productivity shocks also predicts counter-cyclical wage dispersion and rich individual wage dynamics that are consistent with micro-level evidence.

Our paper also belongs to a growing literature that studies the moral hazard problem in a general equilibrium environment. Payne (2018) and Phelan (2017) consider similar dynamic contract problems to that in our paper, but their applications focus on the effects of moral hazard issues in financial intermediary and entrepreneur business income, respectively. Doligalski, Ndiaye, and Werquin (2020) studies nonlinear income taxation in a model where job features pay-for-performance because of moral hazard. Li and Wang (2018) theoretically demonstrates that the existence of moral hazard can generate endogenous labor market fluctuations. Lamadon (2016) investigates how match-specific shocks transmit to worker compensation using matched employer-employee data, where the effort is assumed to be avoid-to-terminate effort that does not directly enter the production function as in Tsuyuhara (2016), and there is no aggregate productivity shock. On that front, Souchier (2022) documents the pass-through of firm-level and sectoral productivity shocks to wages using French employer-employee data.

Clearly, our paper builds on the search and matching framework developed by Diamond (1982) and Mortensen and Pissarides (1994). We assume competitive search (Moen, 1997) and adopt block recursive equilibrium (Menzio and Shi, 2010, 2011). We contribute to the literature by incorporating the dynamic contract problem into the labor search framework and studying the interaction between the two types of frictions. As documented in an influential paper by Shimer (2005), it is difficult for the standard labor search model to account for the magnitude of unemployment fluctuations. Our paper provides a different mechanism that amplifies the effects of aggregate productivity shocks on labor market outcomes, which does not rely on nominal rigidities as in Hall (2005) and Gertler and Trigari (2009) or on a high replacement ratio as in Hagedorn and Manovskii (2008). Our paper also complements the literature which studies the firm's hiring dynamics over business cycles: Bils, Chang, and Kim (2022) consider a matching model with sticky wages within matches and variable effort from workers. They find that effort's response can greatly increase wage inertia. But the effort in their framework is publicly observable to all. Kennan (2010) extends the Mortensen-Pissarides model to allow the employer to have private information on idiosyncratic factors. The procyclicality in information rents can generate higher unemployment fluctuation.

Outline. This paper is organized as follows. Section 2 describes the environment, the contracting problem, and the search and matching problem, then solves them in a general equilibrium framework and obtains the equilibrium. Section 3 calibrates the model and estimates individual risk using micro-level data. Section 4 analyzes the steady-state model and elaborates the key channel that generates the result through comparative statics analysis. Section 5 presents the quantitative results in the dynamic model. Section 6 discusses policy implications. Section 7 concludes the paper.

2. Model

In this section we build a competitive search model that incorporates a repeated moral hazard problem in a general equilibrium environment. We first focus on the analysis in steady state, and later extend the model with aggregate productivity shocks in Section 5.

2.1 Model environment

Consider a small open economy where time is continuous. There is a single perishable consumption good produced by pairwise firm-worker matches. The economy is populated with a large number of risk-neutral firms and a continuum of infinitely-lived workers. Based on their statuses in the labor market, workers are either employed or unemployed, and firms are either operating or staying vacant.

We start with the employed workers and operating firms. Operating firms are assumed to have full claim over all the residuals, offer a stream of consumption and advise workers to exert effort to maximize their expected profits. We further assume that operating firms have unlimited access to an international credit market – they can borrow or lend at a fixed interest rate r. Over time, firms make a profit or incur a loss from the employment relationship and repay the borrowed amount or withdraw savings from the international credit market.[1] Workers, on the other hand, are assumed to have access to neither saving technologies nor asset markets. Essentially, contracting with firms is the only vehicle available for workers to make intertemporal transfers.

Preferences. Employed workers evaluate the contract according to

$$\mathbb{E}\left[r\int_0^\infty e^{-rt}\left(u\left(C_t\right)-\phi\left(A_t\right)\right)dt\right],$$

where C_t is a nonnegative flow of consumption, and $A_t \in \mathcal{A}$ is the effort level. We impose the following assumptions on the preference that are common in the literature on dynamic contract to facilitate the analysis. The set of feasible effort levels $\mathcal{A}$ is compact with the smallest element being 0. The worker's utility is bounded form below ($u(0) = 0$) and that $u : [0, \infty) \to [0, \infty)$ is increasing, concave and of class C^2, with $\lim_{C\to\infty} u'(C) = 0$. The worker's disutility $\phi : \mathcal{A} \to \mathcal{R}$ is continuous,

[1] We do not need to worry about the firm's liability issue here. We can either assume that firms that eventually incur net negative profit can file for bankruptcy or assume that one firm can hire a continuum of workers and that by law of large numbers, the firm will always obtain a positive net profit.

increasing and convex, with $\phi(0) = 0$. In addition, there is a $\gamma_0 > 0$ such that $\phi(A) \geq \gamma_0 A$ for all $A \in \mathcal{A}$. Finally, we assume that the discount rates for workers and firms are set at the common rate r, implying that they are equally patient and that workers will eventually reach permanent retirement if not exogenously separated from firms.[2]

Production. The total output Y_t is the cumulative output produced by an operating firm starting from the moment it matches with a new worker and up to time t. dY_t is the current output flow, which evolves according to

$$dY_t = (zA_t dt + \sigma dB_t) \cdot \mathbb{1}_{t \in [0,\tau)}.$$

The amount of output is determined by three factors: the productivity z, the worker's effort level A_t, and an idiosyncratic factor captured by the Brownian motion dB_t with volatility σ. The productivity z and the effort A_t complements each other. Importantly, the operating firm can only observe the total output and cannot directly separate the effort level from the idiosyncratic factor. This creates room for workers to shirk. For now, we impose that the productivity z is fixed, but we allow it to be stochastic in Section 5.

Remark. In the baseline specification, we impose that the aggregate productivity z only scales the effect of effort A_t on the output, but is independent of the effect of idiosyncratic factor dB_t. This assumption holds naturally when we interpret the idiosyncratic factor as random disturbances that are not directly related to aggregate economic conditions. One may conjecture in certain environments the aggregate productivity could potentially amplify or dampen the effects of dB_t. To accommodate this possibility, we also consider the following extension

$$dY_t = (zA_t dt + z^{\varrho} \sigma dB_t) \cdot \mathbb{1}_{t \in [0,\tau)}, \tag{1}$$

where the parameter ϱ controls the extent to which the productivity shifts the relative importance of effort and noise in shaping the output. If $\varrho > 0$, the aggregate productivity amplifies the effects of dB_t, otherwise, it dampens. We will discuss in more details how our model predictions are modified under different values of ϱ in section 4.1.

Separation. The contract may be terminated under two circumstances: one is that workers quit and switch to their outside option (endogenous separation); the other is that the contract is hit by an

[2]If the worker is more patient than the firm, we can reasonably expect that the payment to the worker to be more "back-loaded" (firms tend to postpone the payment to the future). Vice versa, if the worker is more impatient than the firm, the payment to the worker will be less "back-loaded".

exogenous separation shock $N_{x,t}$ with the Poisson arrival rate λ_x (exogenous separation).[3] We use τ to denote the random termination time of an employment relationship. Upon separation, employed workers immediately join the unemployment pool and lose all their past employment history and output record; remaining operating firms are worthless and are therefore destroyed.[4]

Search market. The search process is directed a lá Moen (1997). For simplicity, we abstract from on-the-job search, so at any time, only the unemployed workers are searching for jobs, and only vacant firms are posting jobs. A jobless worker enjoys a flow value of benefit b, while an idle firm obtains zero flow return. Firms must pay a flow cost k to keep a vacancy open to applications. There are potentially many submarkets indexed by the worker's expected lifetime promised utility W_0, and firms and workers can choose which market they want to enter.

Within a submarket, new jobs are formed according to a constant return to scale matching technology $m(v, u)$, where $v\,(W_0)$ and $u\,(W_0)$ are the measure of vacancies created by firms and the measure of workers searching for jobs in submarket W_0, respectively. Denote $\theta(W_0) = v(W_0)/u(W_0)$ as the tightness of submarket W_0. The firm's job filling rate in a particular submarket is $q\,(\theta) = m(v, u)/v$, and the worker's job finding rate in a particular submarket is $p\,(\theta) = m(v, u)/u$. We further assume that $q'(\theta) < 0$ and $p'(\theta) > 0$ to guarantee that in a tighter market (with a lower θ), it is easier for firms to fill vacancies but more difficult for workers to find a job.

2.2 Contracting problem

Now we turn to the contract problem between a pair of matched firm and worker. In the dynamic setting, the firm has to specify how the compensation is related to the current and past outcomes—a repeated moral hazard problem.

Perfect-information benchmark. Before exploring the properties of constrained optimal contract, we first discuss the optimal contract under perfect information, where all idiosyncratic factors are publicly observable such that any deviation from the optimal effort can be accurately punished. This first-best case serves as a useful benchmark for the repeated moral hazard problem.

[3]We will show that by allowing for temporary suspension of workers, the optimal contract will eliminate all endogenous separations. When workers are temporarily suspended, they devote zero effort, and the output remains zero. Furthermore, retirement is a special stage in an employment relationship and is thus not considered separation.

[4]One may assume that after separation, instead of vanishing, operating firms become vacant again and start to search for new workers, but the free entry condition guarantees that all vacant firms obtain zero expected profit, and the measure of operating firms is determined by this free entry condition. Thus, it does not matter whether firms still exist and become vacant for a next worker or simply completely vanish after separations.

To proceed, it is useful to define the notion of continuation value. For a stream of contingent compensation and advised effort (C_t, A_t), the continuation value is the worker's expected discounted utility by following the recommendation,

$$W_t = \mathbb{E}_t \left[r \int_t^{\tau} e^{-r(s-t)} \left[u\left(C_s\right) - \phi\left(A_s\right) \right] ds + e^{-r(\tau - t)} W_u \right], \tag{2}$$

where W_u is the the expected utility for an unemployed worker as she will directly enter into the unemployment pool once separated from a firm. Note that this is an endogenous object that will be determined in the equilibrium.

If firms can perfectly observe the idiosyncratic factor, any deviation on the workers' side can then be precisely punished. For any initially promised utility W chosen in the submarket, firms will provide full insurance to workers and workers have no incentives to shirk.

Proposition 2.1. *With perfect information, workers exert constant effort $A_{FB}(W)$ and are compensated with a constant stream of consumption $C_{FB}(W)$:*

$$\begin{aligned} A_{FB}(W) &= \operatorname*{argmax}_{A \in \mathcal{A}} \frac{zA - u^{-1}\left[rW + \phi(A) - (W_u - W)\lambda_x \right]}{r + \lambda_x}, \\ C_{FB}(W) &= u^{-1}\left[rW + \phi(A_{FB}) - (W_u - W)\lambda_x \right]. \end{aligned}$$

Proof. See Appendix B.1. □

It is also useful to note that with perfect information, the idiosyncratic factor is irrelevant for workers' welfare, and it follows that it is irrelevant for the aggregate equilibrium outcomes. In contrast, with asymmetric information, this contract is clearly not incentive compatible, and we will see that the nature of the optimal contract interacts with the equilibrium condition.

With asymmetric information, the key issue is how to provide incentive for workers in the most efficient way. When the compensation flow does not respond to outputs, workers will exert minimum effort and hide behind the excuse of receiving adverse idiosyncratic factors. To motivate workers, firms can instead sign contingent contracts with workers, which deliver workers a stream of consumption that is correlated with their output.

The operating firm's problem is to choose a stream of consumption $C \equiv \{C_t, 0 \le t < \infty\}$ and recommendation of effort $A \equiv \{A_t, 0 \leq t < \infty\}$ to maximize their profit

$$\max_{A,C} \mathbb{E}\left[\int_0^\tau e^{-rt}\left(dY_t - C_t dt\right)\right], \quad \text{subject to}$$

1. the contract delivers the worker the initial promised utility W_0

$$\mathbb{E}\left[r\int_0^\tau e^{-rt}\left(u\left(C_t\right) - \phi\left(A_t\right)\right)dt + e^{-r\tau}W_u\right] \geq W_0,$$

2. the advice of effort is incentive compatible.

Such contract should ensure that workers are paid back the utility as initially promised, and workers are provided with proper incentives to exert effort as time evolves. The first constraint requires the firm to fulfill the promise they made in the submarket indexed by W_0. The exact form of the second constraint lies in the center of the contract problem and is quite involved. We proceed by first formulating the IC constraint and then characterizing the firm's optimal contract.

2.2.1 IC constraint

In principle, the firm could and should rely on the entire history of performances to infer and incentivize the worker's effort. However, it is technically difficult to keep track of such an infinite-dimension object. To circumvent this difficulty, we follow Spear and Srivastava (1987) and Sannikov (2008) to use the promised utility (2) as a state variable that summarizes the worker's past performance. This strategy is well justified because of the recursive nature of the worker's and the firm's problems. Intuitively, the worker's incentives remain unchanged even if we replace the continuation contract that follows a given history with a different contract, as long as this new contract provides the same promised utility. Thus, to characterize the optimal dynamic contract, we can instead ensure that following any history, the continuation contract is optimal given the worker's promised utility.

With the promised utility W_t as the state variable, its evolution given a contract that specifies functions $C(W)$ for compensation flow and $A(W)$ for recommended effort level can be expressed as[5]

$$dW_t = r\left[W_t - u\left(C(W_t)\right) + \phi\left(A(W_t)\right)\right]dt + r\Psi_b(W_t)\underbrace{\left(dY_t - zA(W_t)dt\right)}_{\sigma dB_t} + \left(W_u - W_t\right)\left(dN_{x,t} - \lambda_x dt\right). \tag{3}$$

This differential equation specifies how the promised utility responds to the underlying stochastic shocks. The process $\Psi_b(W_t)$ is the key component of the contract that measures its exposure to the idiosyncratic factor. To better understand this evolution, imagine that the promised utility is represented by a bank account at the firm where the worker deposits her utility. The utility in this account requires a return rate of r. If the firm pays out compensation C, then $u(C)$ amount of utility will be withdrawn from this account. If the worker works at effort intensity A, this account will receive a deposit of $\phi(A)$. In the presence of separation shocks, the worker also requires additional return λ_x for the potential utility loss.

Crucially, to incentivize the worker, this account should be set risky and correlated with outcomes. This imposes conditions on the exposure to the idiosyncratic factor, $\Psi_b(W_t)$, as specified in the following lemma.

Lemma 2.1. *In an incentive-compatible contract, the exposure to the idiosyncratic factor satisfies*

$$\Psi_{b,t} = \frac{\phi'\left(A_t\right)}{z}, \tag{4}$$

and the promised utility evolves according to

$$dW_t = \left[r\left(W_t - u\left(C_t\right) + \phi\left(A_t\right)\right) - \left(W_u - W_t\right)\lambda_x\right]dt + r\frac{\phi'\left(A_t\right)}{z}\sigma dB_t. \tag{5}$$

Proof. See Appendix B.3 and Appendix B.4. □

According to (3), the marginal cost and marginal benefit of shirking for the worker are $z\Psi_{b,t}$ and $\phi'(A)$, respectively. To prevent the worker from shirking, the exposure to the idiosyncratic factor should be at least equating the marginal benefit of shirking with its cost, which leads to condition (4). With the incentive-compatible exposure, the evolution of the promised utility is simplified to (5), which will be leveraged momentarily as we formulate the optimal contract problem.

[5]The differential equation is derived in Appendix B.3.

2.2.2 Recursive contract problem

Consider the recursive version of the firm's problem. Denote $F(W)$ as the firm's expected maximum present value when promising W to the worker and is subject to the IC constraint. When first matched with the worker in submarket W_0, firm's expected profit is simply $F(W_0)$. The optimal contract problem can be represented in the following form of Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{aligned}(r+\lambda_x)\,F(W_t) = \max_{A(W),C(W)} \quad & zA(W_t) - C(W_t) \\ & + F'(W_t)\left\{r\left[W_t - u\left(C(W_t)\right) + \phi\left(A(W_t)\right)\right] - (W_u - W_t)\,\lambda_x\right\} \\ & + \frac{F''(W_t)}{2}\left[r\frac{\phi'\left(A(W_t)\right)}{z}\sigma\right]^2 .\end{aligned} \tag{6}$$

The firm is maximizing the expected flow of profit plus expected change of future profit due to the drift and volatility of the worker's continuation value. Note that the drift term and the volatility term of $F(W_t)$ correspond to those in condition (5), which has already incorporated the incentive-compatibility constraint.

To completely characterize the optimal contract, we still need to impose proper boundary conditions for the HJB equation. The lower bound is simply the outside option for a worker, W_u, as the worker can always leave the contract for free. The upper bound W_r is the expected utility for a retired worker and it is finite because firms cannot promise infinite utility to workers. When the promised utility is too high, it will cost a firm too much to motivate the worker due to the income effect. At such point, the firm is better off by retiring the worker. It turns out that the upper bound is absorbing and the lower bound is reflective. We now elaborate how the boundary conditions are determined.

Absorbing upper bound. The optimal contract features an absorbing upper bound where workers get retired and can publicly exert zero effort. Retired workers are still considered employed since they are still being paid by firms but the relationship will end when the exogenous separation shock hits the pair. Under the retirement plan, retired workers will obtain a constant consumption stream $C_R(W)$ over time and firms will obtain negative expected payoff $F_R(W)$, where[6]

$$C_R(W) = u^{-1}\left[W - \frac{\lambda_x}{r}(W_u - W)\right], \qquad \text{and} \qquad F_R(W) = -\frac{C_R}{r+\lambda_x}.$$

[6] The derivation is in Appendix B.2.

Therefore, at the upper bound W_r, the value matching and smooth pasting conditions must be satisfied to make the employment plan consistent with the retirement plan,

$$F(W_r) = F_R(W_r), \qquad \text{and} \qquad F'(W_r) = F_R'(W_r). \tag{7}$$

Reflective lower bound. We follow Grochulski and Zhang (2019) to allow a flexible temporary suspension, which implies that the optimal contract features a reflective lower bound.[7] Under a temporary suspension plan, workers are allowed to temporarily put in zero effort and to return back to work when it is a favorable option for both parties.

We delegate the formal analysis to Appendix B.5, and briefly describe the dynamics here. The state space can be divided into two regions: one is low-effort region where workers are temporarily suspended ($A = 0$); the other is high-effort region where workers put positive effort ($A > 0$). Denote the splicing point where the low-effort region switches to high-effort region as W_s. Once W_t leaves $[W_u, W_s)$, it reflects off W_s and remains in $[W_s, W_r]$. The optimal contract is obtained when W_s is set as low as possible because a larger support $[W_s, W_r]$ could sustain positive effort for a longer period. With a larger high-effort support, the volatility in the continuation value will not drive W_t to collide with W_s as quickly. Therefore, the low-effort region will degenerate into one point at W_u. To guarantee the consistency of the contract over the entire contracting space, the smooth pasting condition must be satisfied at point W_u

$$F'(W_u) = \frac{F(W_u) - F_R(\frac{\lambda_x W_u}{r+\lambda_x})}{W_u - \frac{\lambda_x W_u}{r+\lambda_x}} \tag{8}$$

where the right-hand side represents the slope of the profit function in the temporary suspension region.[8]

Optimal contract. The optimal contract then solves the differential equation (6), together with the boundary conditions (7) and (8). The first-order conditions with respect to the compensation and effort schedule lead to

$$z_t + rF'(W_t)\,\phi'(A_t) + r^2F''(W_t)\left(\frac{\sigma}{z}\right)^2 \phi'(A_t)\,\phi''(A_t) = 0,$$

$$-1 - rF'(W_t)\,u'(C_t) = 0.$$

[7]This is in contrast with Sannikov (2008) where the contract is terminated at W_u and the lower bound becomes an absorbing state.

[8]See Figure 12 in Appendix B.5 for a virtual explanation.

The results are reminiscences of those obtained in Sannikov (2008), except that in our model, W_u is an endogenous object in equilibrium and firms are allowed to temporarily suspend a worker. Figure 1 illustrates that the optimal contract features a "back-loading" consumption plan and hard-to-motivate senior workers due to the income effect.

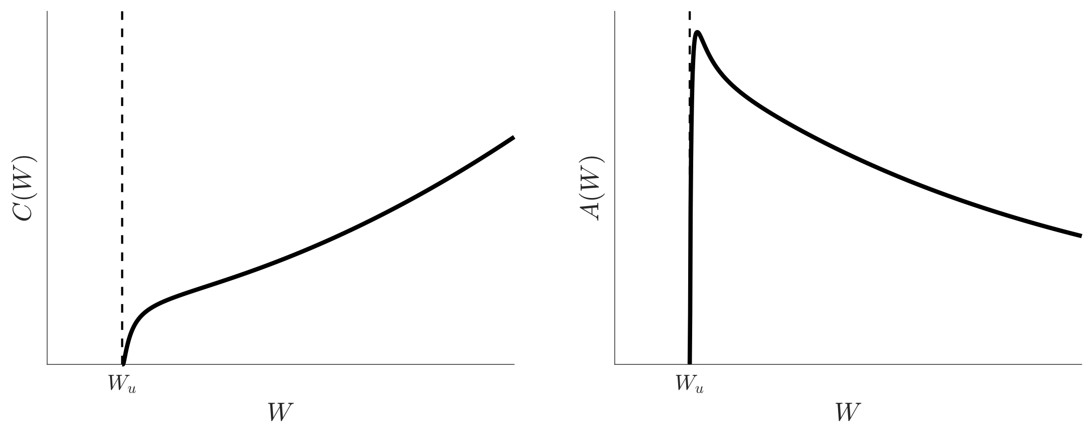


Figure 1: Optimal Compensation and Effort Schedules

Notes: The left panel shows that the compensation schedule is backward-loading. The right panel shows that the effort schedule increases with W near the lower bound and then decreases with W, suggesting that senior workers are difficult to motivate.

At individual level, the compensation profile over tenure directly depends on the evolution of the promised utility; at the aggregate level, the total output is shaped by the distribution of workers over the utility space, which indirectly depends on the evolution of the promised utility. To pin down W_t, we now move to the general equilibrium that determines which submarket W_0 is chosen and the value of being unemployed W_u.

2.3 Competitive search market

In this section, we embed the contract problem into a general equilibrium framework. We assume labor search is directed as in Moen (1997) and characterize the optimal decisions for vacant firms and unemployed workers. The outside option and the initial promised utility that were taken as given in Section 2.2 will now be endogenously determined.

Vacant firms. Observing a menu of submarkets labeled by W_0, firms simultaneously choose whether to create vacancies and where to locate them. In equilibrium, only a subset of submarkets open. For those submarkets that do open, the firm's free entry condition is satisfied in the sense that the firm's expected profit of posting a job equals its posting cost k,

$$k = q\left(\theta\left(W_0\right)\right) F\left(W_0\right). \tag{9}$$

The firm's expected profit of creating a vacancy is higher if the job filling rate $q\left(\theta\left(W_0\right)\right)$ is larger or the expected value of hiring a worker $F\left(W_0\right)$ is higher. We can rewrite (9) as

$$\theta\left(W_0\right) = q^{-1}\left(\frac{k}{F(W_0)}\right). \tag{10}$$

This equation defines a unique relationship between the initial promised utility and the labor market tightness. If a firm's expected value of hiring decreases, it will be less willing to post a job, and collectively this will tighten the labor market and make the hiring process slower. This connects the properties of the optimal contract with the labor market outcomes.

Unemployed workers. We assume that all unemployed workers participate in job searching. Unemployed workers take the menu of submarket as given and optimally submit their applications. If an unemployed worker successfully finds a job, she will sign a new contract that delivers her the lifetime utility W_0. Otherwise, she remains unemployed and obtains an unemployment benefit flow b. Anticipating the relationship between the market tightness and promised utility (10), the value of being unemployed is given by the following lemma.

Lemma 2.2. *For an unemployed worker, the value function W_u and the directed search decision W_0 solve the following recursive problem:*

$$rW_u = \max_{W_0 \in \mathcal{W}} ru\left(b\right) + p\left[q^{-1}\left(\frac{k}{F(W_0)}\right)\right]\left(W_0 - W_u\right). \tag{11}$$

Proof. See Appendix B.7. □

Equilibrium. We adopt the recursive equilibrium concept à la Menzio and Shi (2010). Crucially, where the unemployment rate in an equilibrium does not depend on the distribution of workers over their employment status and promised utility level. This property is particularly useful when there is aggregate uncertainty as in Section 5.

Definition 2.1. *A* ***block recursive equilibrium*** *consists of a market tightness function $\theta : \mathcal{W} \to \mathbb{R}_+$, a utility function of unemployed workers W_u, a promised utility function of employed workers $W \in [W_u, W_r]$, an initially promised utility function for newly hired workers $W_0 \in [W_u, W_r]$, and a firm's value function $F : \mathcal{W} \to \mathbb{R}$ such that the following hold:*

(i) The employed worker's promised utility W follows (5).

(ii) The operating firm's value function $F(W)$ solves (6), (7) and (8).

(iii) Given the market tightness function $\theta(W_0)$, the vacant firm's choice of entering market W_0 satisfies the free entry condition (9).

(iv) Given the market tightness function $\theta(W_0)$, the unemployed worker's utility function W_u and the choice of market to direct search W_0 satisfy condition (11).

Despite the presence of the dynamic contract, the model is still tractable. Thanks to the continuous-time environment, solving the general equilibrium boils down to solving a finite number of differential equations. The computational details are explained in Appendix D.

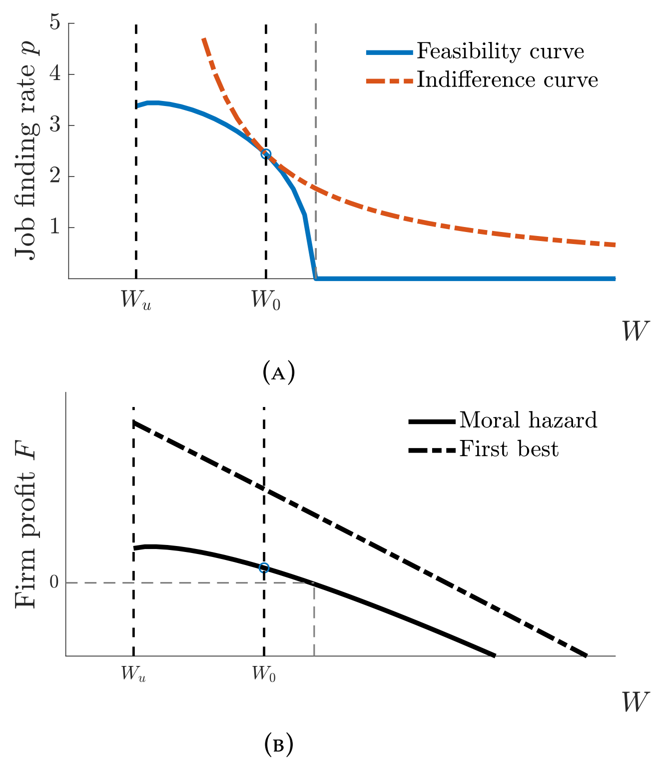


FIGURE 2: Endogenizing outside option W_u and initially promised utility W_0

Notes: Panel B shows that the firm profit function is hump-shaped with moral hazard problem. Panel A shows that the feasibility constraint for workers is given by the blue curve (the vacancies that firms would like to create). The job finding rate drops to zero when the firm profit drops to zero or below. The red curve represents workers' indifference curve in job searching. The tangent point W_0 determines the submarket that unemployed workers will collectively visit. The choice of W_0 in turn settles the promised value W_u for the unemployed.

As mentioned earlier, the values of W_0 and W_u lie at the interaction between optimal contract and the labor market equilibrium. How are these objects determined in equilibrium? Given a particular value of W_u, a firm can evaluate their profits by posting jobs with different promised utilities. Figure 2b illustrates that the profit function is hump-shaped. It is increasing in a very short interval at left end, where the worker and the firm's incentives are aligned. The profit quickly declines as promised utility increases as the wage bill accumulates. This profit function in turn implies a locus of market tightness and promised utility through the free entry condition (9), which is depicted by the blue solid line in Figure 2a. Finally, workers find their ideal submarket W_0 to enter, which correspond to the tangent point between their indifference curve and the feasibility curve.

In equilibrium, the value of W_u has to be consistent with workers' expectation, satisfying condition (11). This amounts to a fixed-point problem. All the other equilibrium objects can be derived according to the optimal contract once W_u and W_0 are determined.

3. Parameterization

In this section, we describe the parameterization strategy in order to quantify the model. Most of the parameters can be calibrated with standard moments on labor market flows, and we explain in details how the parameters related to moral hazard problem are determined.

3.1 Calibration

Table 1 summarizes the calibrated parameter values. The model period is a quarter, and the interest rate r for workers and firms is set to be 0.012 so that the annual rate is 4%. The arrival intensity of the separation shock λ_x is set to match the average job duration at 2.5 years. We choose a Cobb-Douglas matching function, $m(v, u) = \xi u^{\alpha} v^{1-\alpha}$. We set the vacancy cost k so that the steady-state labor market tightness is normalized to be 1. The matching efficiency ξ is set to match the monthly job finding rate at 0.45, and the matching elasticity α is estimated using detrended data on the job finding rate and the $v - u$ ratio following Shimer (2005). The unemployment benefit is set to match the replacement ratio at 20%.

Turn to workers' preferences. The utility function for consumption takes a standard CRRA form, $u(C) = \frac{C^{1-\eta}-1}{1-\eta}$. The risk aversion η is set to be 0.5, as in Sannikov (2008). Note that this value is lower than those commonly used in standard DSGE literature. This is because without endogenous capital accumulation, labor earnings are the only source of income and a value for η larger than one implies too strong income effects. In canonical labor search models, with $\eta > 1$, the employment would be counter-cyclical and it is common to specify linear utility function. We set $\eta = 0.5$ to ensure an empirically plausible income effect on labor supply.

In terms of the disutility of effort, we adopt a quadratic functional form, $\phi(A) = \chi_1(A^2 + \chi_2 A)$. We set χ_1 so that the aggregate labor supply is normalized to one, $\int_W A(W)g(W)dW = 1$. The coefficient of the linear term χ_2 is related to the amount of time workers spend on the temporary suspension stage when the promised utility approaching its reflective lower bound. We set $\chi_2 = 0.5$ so that workers spend negligible amount of time around this corner.

Parameter	Description	First Best	Baseline	Target
r	discount rate	0.012	0.012	4% annual return
λ_x	separation rate	0.1	0.1	job duration 2.5 yrs
ξ	matching efficiency	1.35	1.35	monthly job finding rate 0.45
α	matching elasticity	0.72	0.72	Shimer (2005)
η	risk aversion	0.5	0.5	Sannikov (2008)
z	aggregate productivity	1	1	normalization
b	unemployment benefit	0.2	0.2	replacement ratio $\frac{b}{c} = 20\%$
k	vacancy posting cost	0.203	0.0053	market tightness $\theta = 1$
χ_1	effort disutility	0.183	0.063	aggregate output equals 1
χ_2	effort disutility	0.5	0.5	—

Table 1: Parameters used for benchmark calibration

3.2 Estimating the variance of idiosyncratic factors

The key parameter in our calibration is the variance of idiosyncratic factors. It governs the underlying information friction faced by firms. Our baseline calibration strategy is to match the residual of performance pay in the PSID. The variance of the idiosyncratic factor is recovered by indirect inference: we choose the parameter such that when running the same regression using model generated data, the residual is the same as that using PSID performance pay data. As a benchmark, we use the residual of performance pay as the target, and as a robustness check, we use regular wage residual to perform the estimation.

Sample selection. We use the PSID data to conduct the estimation. The sample is from 1994 to 2019.[9] The PSID data are at a yearly frequency from 1994 to 1996 and at a biannual frequency from 1997 to 2019. This gives us 15 years of unbalanced panel data. The sample contains 15,571 individuals who contribute to a total of 62,950 person-year observations. Furthermore, the sample contains 34,401 worker-job matches. If we define a performance pay job as job that pays performance pay at least once in the whole employment span, then the sample contains 5,614 performance pay jobs that contribute to a total of 14,692 performance pay job-year observations.

Estimation strategy. To infer the variance of idiosyncratic factor σ^2, we use the volatility of performance pay conditional on observed workers' characteristics. The identification assumption is that firms do not possess more information than econometricians, and the only unobservable factor that causes the variance in the residual of performance pay is the variance of the idiosyncratic factor. We set σ^2 so that the volatility of model implied wage residual matches the volatility of performance pay residual.

We focus on the performance pay instead of the regular wage as the former is directly linked to firms' need to incentivise workers. As argued in Lemieux, MacLeod, and Parent (2009) and Doligalski, Ndiaye, and Werquin (2020), performance pay is a "cleaner" identifier for estimating the variance of idiosyncratic factor because performance pay is used directly to motivate the worker to exert effort and can also immediately reflect the worker's performance. If we instead use regular wage residual to estimate the variance of idiosyncratic factors, then one reasonable critique is that firms have a larger information set than econometricians. Using the performance pay residual can partially resolve such concern. Nevertheless, we experiment the specifications using regular wage residual in Appendix E.2, and the results are similar qualitatively.

We define a performance pay job as a job that pays performance pay at least once during the course of employment. Hourly performance pay is computed by dividing annual income from bonuses, tips, and commission fees (adjusted by CPI-U) by annual working hours. The performance pay residual is the residual part of hourly performance pay after controlling for worker-level observable characteristics, as well as worker-firm, match-level, time-invariant unobservable factors, year-specific aggregate unobservable factors, industry-cross-occupation-level unobservable factors and fixed effects. Built on the convention in Heathcote, Perri, and Violante (2010) and the

[9]The PSID began in 1968, but it only started to report the bonuses, tips, and commission income of the family head in 1994.

"industry standard" Mincer regression, we set the following regression specification.

$$\begin{aligned}\texttt{performance}_{i,j,t} = &\beta_{edu}\ \texttt{edu}_i + \beta_{expr}\ L(\texttt{expr}_{i,t}) + \beta_{tenu}\ L(\texttt{tenu}_{i,j,t}) + \beta_t\ \texttt{D}_t + \beta_{ind}\ \texttt{D}_{ind} + \beta_{occ}\ \texttt{D}_{occ} \\ &+ \beta_{i,j}\ \texttt{D}_{i,j} + \varepsilon_{i,j,t}.\end{aligned}$$

In this regression, $\texttt{performance}_{i,j,t}$ is the asinh hourly real performance pay of individual i working job j at time t. We run the regression on year dummies $\texttt{D}_t$, year of education $\texttt{edu}_i$, a cubic polynomial of potential experience $\texttt{expr}_{i,t}$, a quadratic polynomial of tenure $\texttt{tenu}_{i,j,t}$, industry dummies $\texttt{D}_{ind}$, occupation dummies $\texttt{D}_{occ}$, and finally, worker-job dummies $\texttt{D}_{i,j}$ to tease out the worker-job-level time-invariant component from the residual.

Then $\varepsilon_{i,j,t}$ is the residual in performance pay. $\texttt{std}\left[\varepsilon_{i,j}\right]$ is the data counterpart of the model-implied variance of the hourly compensation, $\texttt{std}\left[\texttt{asinh}\left(\frac{C}{A}\right)\right]$.[10] We also explore various specifications and the regression results which are summarized in Table 2. The variance of the idiosyncratic factor is chosen to match the estimation results. We use the regression results in Column (6) to calibrate the baseline model, and it indirectly infers σ to be 7.6.

4. Steady-state analysis

In this section, we explore the effects of moral hazard on labor market dynamics at the steady state. At the aggregate level, we show that when the moral hazard problem is present, the elasticity of the unemployment rate with respect to aggregate productivity is amplified and the wage dispersion becomes countery-cyclical. At the individual level, we show the individual wage dynamics display a wage scarring effect. Most of the insights developed in this section can be carried to the model with aggregate uncertainty.

4.1 Unemployment rate and the productivity shock

Unemployment Elasticity. How does unemployment rate respond to a change of the aggregate productivity in the long run? We compare the steady-state unemployment rate with different levels of aggregate productivity in the economies with and without the moral hazard problem.

[10] We use `asinh` function instead of log function to take away unit for dependent variable. `asinh` has a nice shape similar to log and $asinh(0) = 0$.

Dependent	asinh(performance)					
Method	OLS	OLS	FE	FE	FE	FE
Regressors	(1)	(2)	(3)	(4)	(5)	(6)
i.worker		✓	✓			
i.match				✓	✓	
i.year	✓		✓		✓	
i.industry	✓		✓		✓	
i.occupation	✓		✓		✓	
i.y# i.ind# i.occ		✓		✓		✓
Observations	14496	14332	13103	12885	12139	11897
Std of residual	.675	.648	.533	.510	.513	**.487**

[1] To match with the model, we use asinh hourly real performance pay: PSID provides yearly nominal performance pay, and we deflate it using CPI-U. Since we use the real performance pay, there's no need to control the time variable to detrend the performance pay.

[2] Controls that we use in the regressions but omit in this report table include: worker's education years `edu`, a cubic polynomial of worker's potential experience in the labor market `expr`, $\texttt{expr}^2$, $\texttt{expr}^3$, a quadratic polynomial of worker's tenure year in current job `tenu`, $\texttt{tenu}^2$.

Table 2: Regression table of performance pay residual

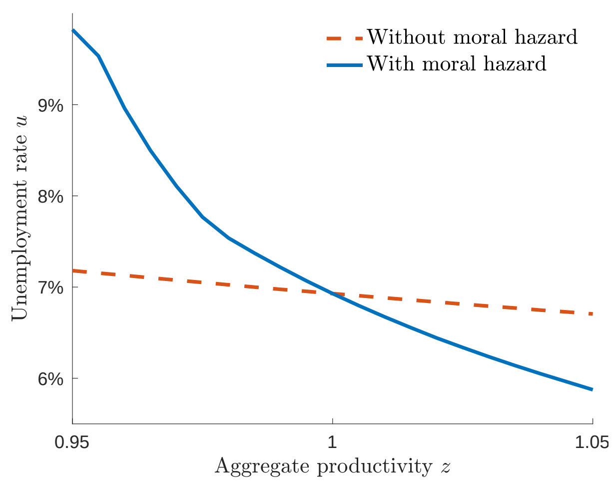


Figure 3: The elasticity of the unemployment rate with respect to z

Notes: The inverse relationship of unemployment rate and the productivity is steeper with moral hazard problem than without. It suggests that the unemployment rate is more elastic to the change in aggregate productivity when the moral hazard problem is taken into account.

The red dashed line in figure 3 plots how the unemployment rate u changes with aggregate productivity z in the first-best economy without moral hazard. The elasticity of unemployment with respect to productivity is quite low, which resembles the findings in the labor search literature Shimer (2005).

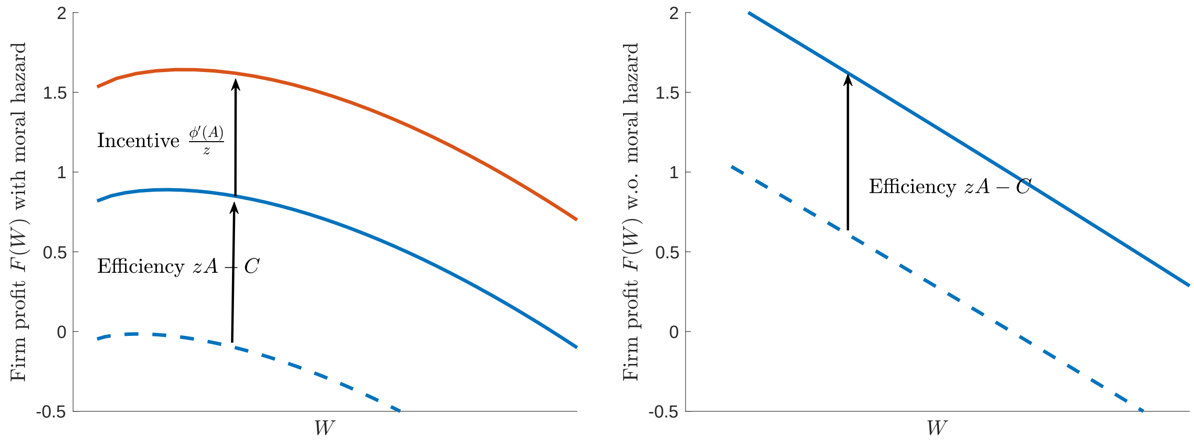
In contrast, the blue solid line plots the changes in unemployment in the baseline economy. Compared with the first-best economy, the unemployment rate is much more responsive to changes in the productivity shock. The reason for this additional responsiveness is that firms are more responsive in its posting behavior. Recall that $dY_t = zA_t dt + \sigma dB_t$. Firms can only observe the output process Y_t and but cannot observe the idiosyncratic factor. For a target A_t, a higher aggregate productivity shock z will help firms to form more accurate forecast about A_t as the idiosyncratic factor become relatively less important in shaping the output. This reduces worker's power in charging the information rent, leaving the moral hazard problem less severe. This encourages firms to post more jobs, which we explain in more details below.

Decomposition: Efficiency v.s. Incentive. To see how the firm's job posting incentive is modified, we revisit the HJB equations that determine the firm's profit in first-best economy and in the moral hazard economy:

$$\text{First best:} \qquad (r + \lambda_x) F(W_t) = \max_{A} \quad \underbrace{zA_t - C_t}_{\text{standard efficiency gain}}$$

$$\begin{aligned}\text{Moral hazard:} \quad (r + \lambda_x) F(W_t) = \max_{A,C} \quad & \underbrace{zA_t - C_t}_{\text{standard efficiency gain}} \\ & \underbrace{+ F'(W_t) \left[r \left(W_t - u(C_t) + \phi(A_t) \right) - (W_u - W_t) \lambda_x \right] + \frac{1}{2} F''(W_t) \left(\frac{r\phi'(A_t)\sigma}{z} \right)^2}_{\text{additional incentive gain}}\end{aligned}$$

Note that z appears in both the flow term and the volatility term of the HJB equation for the moral hazard economy, while z only enters the output flow term of HJB equation for the first-best economy. Ceteris paribus, when there is a permanent increase in z, the output flow will increase due to the substitution effect, and this is the standard efficiency gain. Furthermore, spending on motivating the worker will decrease due to a looser incentive constraint, which we label as the



(A) Moral hazard

(B) Without moral hazard

FIGURE 4: The response of a firm's profit to a change in z

Notes: We use these diagrams to show why vacancy posting is more responsive to productivity shocks when there exists moral hazard problem. Consider there is a positive productivity shock. The dashed blue curves are firm's profits before the rise of productivity. The efficiency gain from higher productivity moves profit functions to solid blue curves. However, with the moral hazard problem, higher z also relaxes the incentive constraints and reduces firms' costs in incentivising workers. This additional incentive gain brings firm's profit to the red curve and encourages firms to post more vacancies.

additional incentive gain for firms.[11]

In Figure 4a, the blue dashed line is the original profit function for a firm. When z permanently increases by 10%, the firm's profit will shift up to the blue solid line if we only consider the standard efficiency gain. When we also consider the incentive gain for the firm, the profit curve moves further up to the red line. In contrast, in figure 4b without moral hazard, when z increases by 10%, the profit curve only shifts from the dotted blue line to the solid blue line, and the size of the shift is due entirely to the standard efficiency gain.

This partial equilibrium result for the firm's profit change carries over to the firm's willingness to post jobs. As a result, the labor market is less tight and job finding rate for unemployed workers is higher. In steady state, the flow of workers entering the unemployment pool $\lambda_x dt\,(1-u)$ is equal to the flow of unemployed workers exiting the pool $(pdt)\,u$, and the steady state measure of unemployed workers is given by

$$u = \frac{\lambda_x}{\lambda_x + p\left(\theta(W_0^*)\right)}.$$

A higher job finding rate $p\left(\theta\left(W_0^*\right)\right)$ implies a lower unemployment rate.

To illustrate how this incentive provision motive appears in general equilibrium, Figure 5 de-

[11] $F''(\cdot) < 0$ is formally proved in Sannikov (2008), section 7.4, lemma 1.

composes the unemployment rate changes in the baseline economy into two parts: efficiency gain and incentive gain. The red curve represents the efficiency gain when the incentive gain is muted, and the blue curve captures the composite effect of both the efficiency gain and the incentive gain.

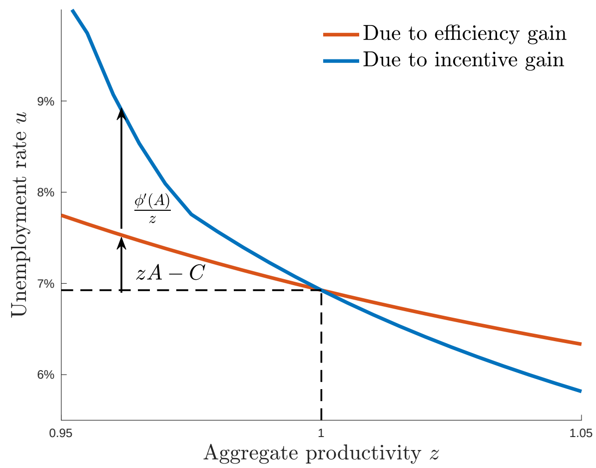


FIGURE 5: Decomposing the unemployment rate elasticity with respect to z

Notes: With the presence of moral hazard problem, the inverse relationship between unemployment and productivity demonstrated in Figure 3 is due to two effects: efficiency gain and incentive gain. The efficiency gain is reflected by the slope of red curve and the incentive gain is reflected by the difference in slopes of blue curve and red curve.

Role of ϱ. As discussed in section 2.1, it is possible that the aggregate productivity can also scale the effect of noise on output as specified by equation (1). As a robustness check, figure 6 shows that unemployment is more elastic with respect to aggregate productivity as ϱ decreases. When ϱ approaches to 0, the elasticity of unemployment corresponds to our baseline model. We can expect that as ϱ grows, the moral hazard has less effect on generating elastic labor market response and less in favor of our claims. Whereas when ϱ decreases, the moral hazard problem has a larger bite in creating more elastic employment.

4.2 Wage dispersion and the productivity shock

Next, we examine how wage dispersion responds to the change in aggregate productivity in the presence of the moral hazard effect. We claim that even without ex ante worker heterogeneity, the information friction alone could create counter-cyclical wage dispersion. Recall that the incentive constraint for the worker is

$$z\Psi_{b,t} \geq \phi'(A_t)$$

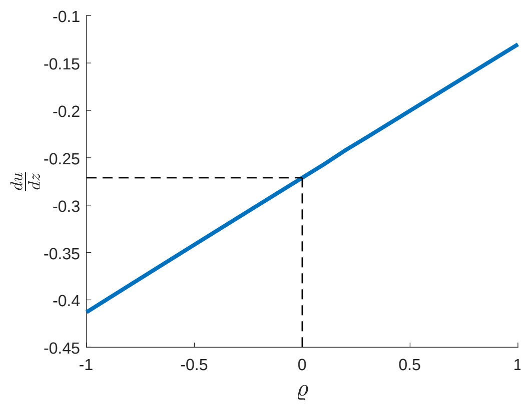


FIGURE 6: Elasticity of the unemployment rate with respect to z

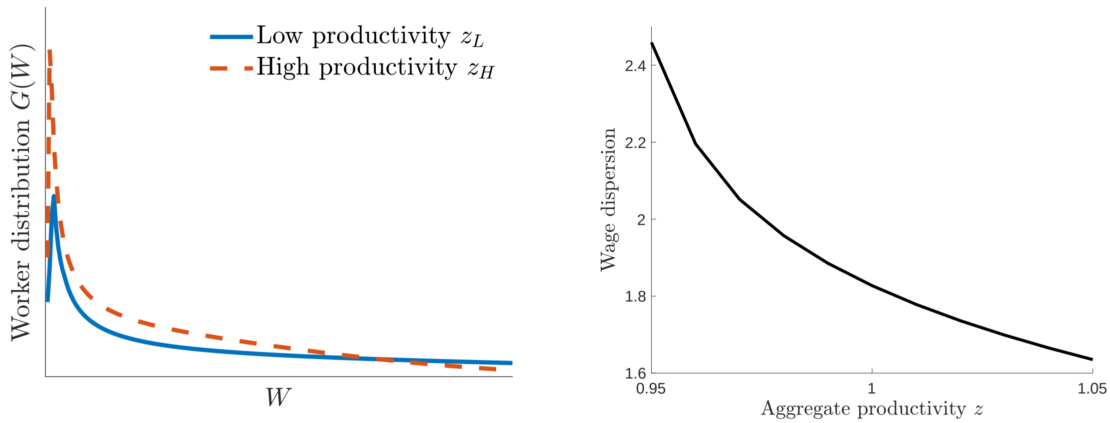
When the aggregate productivity z increases, for a fixed target effort plan A_t, the incentive constraint becomes more relaxed. To reduce the cost of exposing workers to the idiosyncratic factor, a firm will choose a lower level of exposure to motivate the worker. Put it differently, with higher productivity, the more relaxed incentive constraint implies that the firm has less need to expose the worker to risk. Consequently, compensation is less volatile over time and there is less cross-sectional dispersion.

This logic appears in the evolving process of the worker's promised utility.

$$dW_t = \left[r\left(W_t - u\left(C_t\right) + \phi\left(A_t\right)\right) - \left(W_u - W_t\right)\lambda_x\right]dt + r\frac{\phi'\left(A_t\right)}{z}\sigma dB_t$$

Aggregate productivity z enters the volatility term of the stochastic differential equation. The higher z is, the less volatile the promised utility is.

Wage dispersion is measured by the standard deviation of cross-sectional wages. Figure 7a shows that the distribution of promised utility is less fat-tailed when aggregate productivity is higher. The blue solid curve is the worker's distribution when setting $z = 1$, and the red dotted curve is that when setting z to be 10% higher. Since $C(W)$ and $A(W)$ are nearly monotonic transformations of promised utility, the wage dispersion defined by $\texttt{std}\left[\log \frac{C}{A}\right]$ inherits the pattern of the distribution of promised utility. Figure 7b shows that wage dispersion will decrease as aggregate productivity rises. The elasticity of wage dispersion with respect to aggregate productivity is -5.34 near the steady state ($z = 1$).



(A) Promised utility distribution: translated to same left end

(B) Countercyclical wage dispersion

FIGURE 7: Wage dispersion

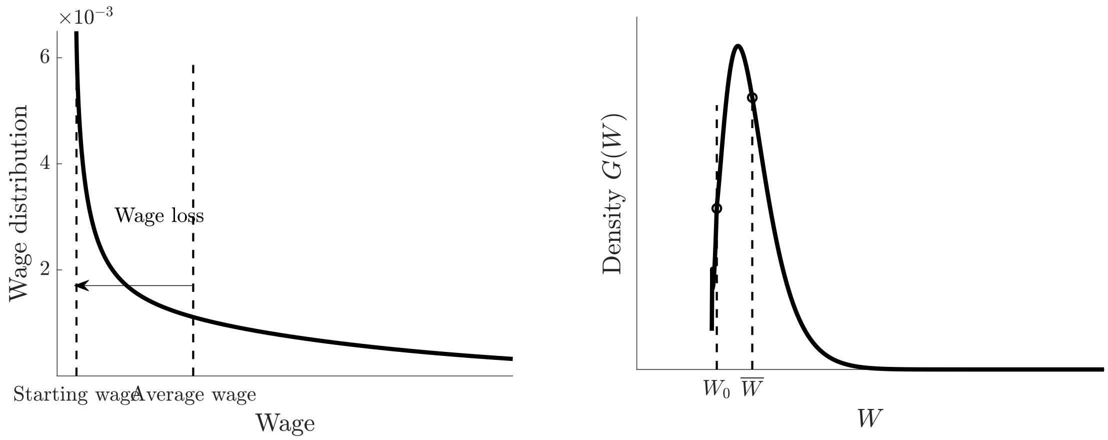
Notes: Panel A shows that the promised utility for workers is more concentrated if the productivity is higher, as the red curve is less fat-tailed than the blue curve. Panel B shows that the wage dispersion is negatively correlated with the productivity.

4.3 Individual wage dynamics

Now we turn to the individual wage dynamics. We document two features of the models that are consistent with the empirical regularities: workers will suffer a wage loss after replacement (a wage scaring effect) and a right-skewed wage distribution.

The baseline model with moral hazard displays a wage scarring effect, which is in contrast with the constant compensation pattern in the frictionless economy. When workers lose their current job and are then reemployed, the wage plummets. The first-year wage compared with the most recent payment from the last job will decrease by 54% according to our simulated data. The reason is that the worker's wage is back-loaded and it takes time to accumulate reputation.[12] Figure 8a visualizes the wage scarring effect, which is represented by the gap between the starting wage for new hires and the average wage for all employed workers. Jacobson et al. (1993), Couch and Placzek (2010), and Barnette and Michaud (2011) document that the wage decreases by 15% – 40% after job displacement. Our model-generated result lies a bit beyond the upper end of the range of estimates that these empirical papers report. A potential reason is that we assume that workers lose all reputation from past employment once removed from their current jobs and that workers have to build their reputation from scratch. However, in reality, workers can partially reveal their past

[12]Here "reputation" is a loosely defined term, and its model correspondence is a worker's promised utility.



(A) Wage scarring effect

(B) Distribution of promised utility

Figure 8: Individual wage dynamics

Notes: Panel A shows that when the separation shock hits, the worker will have to find another job and the starting wage is 54% less than the original job. Panel B illustrates the stationary distribution of worker's promised utility. The promised utility also drops from $\overline{W}$ to W_0 when the worker changes job.

performance with previous employers using resumes or recommendation letters, which smooths the job transition for workers.

The individual wage dynamics also imply that most workers are working with low promised utility. Figure 8b illustrates the stationary distribution of worker's promised utility and it is right-skewed. Once workers reach the lower bound, they will linger in the region with relatively low payment. Our model-generated data show that it takes workers 8.7 quarters to return to average wage after reaching the lower bound. The wage scarring effect is also reflected by the gap between average promised utility $\overline{W}$ and the initial promised utility W_0 in Figure 8b. The wage scarring effect helps to account for a Gini coefficient of 0.22 for the model-generated data, though there is no ex-ante heterogeneity among workers.

5. Dynamic model analysis

In this section, we extend the analysis to allow uncertainty on the aggregate productivity. We quantify the role of moral hazard in amplifying unemployment volatility and generating counter-cyclical wage dispersion.

5.1 Model with aggregate uncertainty

The dynamic model is quite similar to the model in steady state except that we now allow aggregate productivity z to vary over time. We assume that z_t is the aggregate productivity shock publicly known to all and it is a regime-switching shock that takes two possible values, $z_t \in \{z_L, z_H\}$. The regime switches when the productivity shock $N_{z,t}$ arrives with a Markov transition rate $\lambda_z(z_t)$. The transition matrix is thus given by

$$\begin{bmatrix} 1-\lambda_z(z_L) & \lambda_z(z_L) \\ \lambda_z(z_H) & 1-\lambda_z(z_H) \end{bmatrix}.$$

We use z^c to denote the complement state to the current state z. The aggregate productivity shock creates an additional state variable when we solve the model. We briefly describe the modifications of the model elements while leaving the full detail of the model to Appendix C.

Employed worker's problem. Now, the worker's promised utility is also exposed to the aggregate productivity shock $dN_{z,t}$, and the exposure process is $\Psi_{z,t}$. The worker's promised utility follows a stochastic process:

$$dW_t = \left[r\left(W_t - u\left(C(W_t)\right) + \phi\left(A(W_t)\right)\right)\right]dt + \Psi_{b,t}\sigma dB_t + \Psi_{x,t}\left(dN_{x,t} - \lambda_x dt\right) + \Psi_{z,t}\left(dN_{z,t} - \lambda_z\left(z_t\right)dt\right)$$

Operating firm's problem. We use $F(W,z)$ and $F(W,z^c)$ to characterize firm's profit under two aggregate states. An operating firm's recursive problem now can be summarized by a system of two HJB equations, where compared with the steady state HJB equation, two elements related to the change in the aggregate productivity shock z are added: (a) the immediate effect on the firm's profit $\lambda_z(z)\left[F(W+\Psi_z, z^c) - F(W,z)\right]$, (b) the indirect effect that shifts a worker's promised utility and through the income effect alters the firm's profit.

$$\begin{aligned}(r + \lambda_z(z) + \lambda_x)F(W,z) = \max_{A,C,\Psi_z}\; & zA - C + \lambda_z(z)F\left(W + \Psi_z, z^c\right) \\ & + F_W(W,z)\left[r\left(W - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z) - W\right) - \lambda_z(z)\Psi_z\right] \\ & + \frac{1}{2}F_{WW}(W,z)\left(\frac{r\phi'(A)}{z}\right)^2\sigma^2\end{aligned}$$

Vacant firm's problem. The expected payoff from hiring a new worker is now related to the aggregate state, making a firm's posting decision also vary with the aggregate state. The free entry condition for the firm is given by

$$k = q(\theta(z)) F(W_0(z), z) = \xi \theta(z)^{-\alpha} F(W_0(z), z).$$

Unemployed worker's problem. For unemployed workers, the search decision will also be altered by the aggregate productivity shock. Their value function comes from the utility gain from unemployment benefits, the change in utility from the change in aggregate state, and the expected search value.

$$(r + \lambda_z) W_u(z) = \max_{W_0(z) \in \mathcal{W}} ru(b) + \lambda_z W_u(z^c) + p(\theta(W_0(z), z))[W_0(z) - W_u(z)]$$

Block recursive equilibrium. In block recursive equilibrium, the worker's value and policy functions depend on the aggregate state of the economy only through the aggregate productivity z but not through the distribution of workers across different employment states. Without this refinement, firms and workers should at least keep track of the unemployment rate.[13] For this reason, our model can restrict the state variables to a total number of two (W, z) and is still tractable in the dynamic environment.

Dynamic model calibration. We have four additional parameters to calibrate, i.e., $z_L, z_H, \lambda_z(z_L)$, $\lambda_z(z_H)$. The calibration strategy is to match the moments of quarterly data on labor productivity (output per hour) and the frequency of recession indicator. We assume that aggregate productivity's deviation from the steady state is symmetric and the percentage deviation is Δz. Following the estimation in Shimer (2005), $\text{std}[\Delta z] = 0.02$, $\mathbb{E}[\Delta z_t \Delta z_{t+1}] = 0.878$.[14] Furthermore, the recession indicator suggests that the probability of low aggregate productivity is $\pi_{\Delta z L} = 0.141$. These moments indicate the values of our four parameters: $\Delta z_L = -2.87\%$, $\Delta z_H = 2.87\%$, $\lambda_z(z_L) = 0.105$, $\lambda_z(z_H) = 0.017$. A detailed discussion on the calibration of the aggregate productivity process is in Appendix E.1.

[13] Since we do not consider on-the-job search in our model, the recursive equilibrium should only additionally consist of the measure of employed workers u_t but not the whole distribution of employed workers $G(W)$. Nevertheless, Menzio and Shi (2011) formally show in Theorem 2 that block recursive equilibrium exists and is unique and achieves the social optimum. They also show there is no loss in generality in focusing on the block recursive equilibrium because all equilibria are block recursive.

[14] The data are from the U.S. Bureau of Labor Statistics. The labor productivity data are measured as the percentage change from the previous quarter at an annual rate, and the quarterly percentage change is approximately the annual percentage change divided by four. To eliminate the unit difference, we take the log of the labor productivity level data and obtain the percentage change for each quarter compared with the base year (1947 Q1). We also detrend the data by applying the HP filter 1600 on the time series data.

5.2 Unemployment volatility

Table 3 compares the unemployment volatility in the dynamic model with and without the moral hazard problem. The standard deviation of the unemployment rate in the information frictionless economy is quite low (0.016), while our baseline model brings the model much closer to the data (0.137) which amounts to roughly 70% of the unemployment rate volatility observed in the data.

Moments	std(z)	autocorr(z)	std(u)
Data	0.02	0.88	0.190
No moral hazard	0.02	0.88	0.016
Benchmark	0.02	0.88	0.137

Table 3: Unemployment volatility

Our result complements the existing literature on the Shimer's puzzle (Shimer, 2005). Previous studies have resorted to a high replacement ratio (Hagedorn and Manovskii, 2008) or rigid wages (Hall, 2005; Gertler and Trigari, 2009) in generating higher unemployment volatility. Our model takes a different approach and endogenously creates higher unemployment volatility in the presence of asymmetric information frictions. What drives this result is the counter-cyclical information rent, as we elaborated in the steady-state analysis in Section 4.1. Compared to the first-best economy, firms in the moral hazard economy are more willing to post jobs during a business upturn because they enjoy an additional profit gain from spending less on incentivizing workers.

5.3 Wage dispersion across business cycles

In terms of the cyclical property of the wage dispersion, the model's simulation results in Table 4 accord well with the empirical findings, displaying a counter-cyclical pattern. The correlation between output and wage dispersion is −0.34 in our model. The wage dispersion conditional on low aggregate productivity realization (0.69) is larger than that conditional on high aggregate productivity realization (0.66). These results are consistent with the findings in Storesletten, Telmer, and Yaron (2004) who empirically document that wage dispersion is higher in contraction periods than that in expansion periods.

In our economy, the cross-sectional dispersion of underlying idiosyncratic shocks is independent of the aggregate shock. The counter-cyclical wage dispersion is instead driven by the counter-cyclical

Moments	Baseline
`corr(GDP, wage disp.)`	-0.34
`wage disp.`$\lvert z_L$	0.69
`wage disp.`$\lvert z_H$	0.66
`autocorr(wage disp.)`	0.76

Table 4: Wage dispersion

information rent as discussed in Section 4.2. In business downturns, the idiosyncratic shocks loom larger than that of the aggregate productivity shock, which makes firm more exposed to the moral hazard problem. In response, firms optimally provide less insurance for workers, leaving wage compensation more responsive to workers' individual shocks. At the aggregate level, the wage dispersion displays a counter-cyclical pattern.

Quantitatively, such channel is not designed to account for the entire observed cyclicality in wage dispersion. The counter-cyclical wage dispersion is likely to be influenced by changes in labor force composition and heterogeneous exposures to aggregate output (Patterson, 2023; Guvenen, Ozkan, and Song, 2014). Nevertheless, our theory complements the existing theory in rationalizing the cross-sectional pattern without attributing to exogenous variations in underlying shocks.

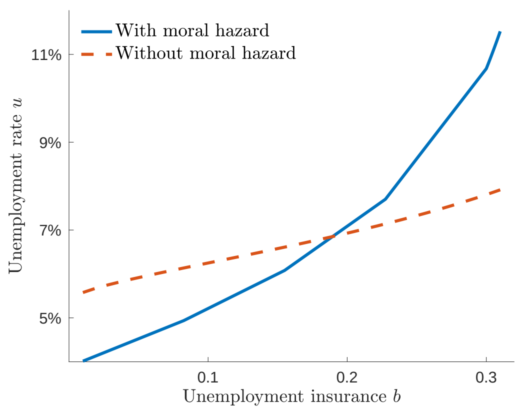
6. Policy implications of the model

In this section we will extend our discussion to the policy implications. We are especially interested in the unemployment insurance policy and the minimum wage policy, as these two policies are frequently employed by the government to improve the welfare for the bottom workers. But at the same time, these two policy tools are also likely to alter workers' incentives. We will discuss the interplay between the moral hazard problem and labor market policies.

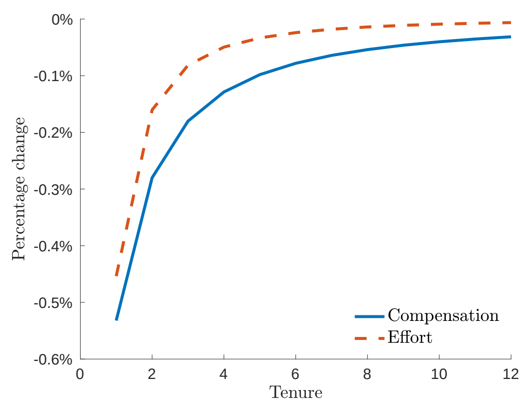
6.1 Unemployment insurance

In this part we will study the interaction between moral hazard problem and unemployment insurance policy. Even without the moral hazard problem, when workers are guaranteed a generous unemployment insurance plan, the general equilibrium effect indicates that this unemployment benefit encourages workers to wait longer for better jobs and thus that the unemployment rate will rise. However, when we consider the moral hazard problem, as in our baseline model, more generous unemployment insurance also undermines worker incentives: it is more difficult for firms to motivate workers since they become less afraid of being suspended. Figure 9a illustrates that the unemployment rate can quickly rise to an unfavorable level if the government decides to equip unemployed workers with a more generous unemployment insurance benefit. This seeming benefit for unemployed workers may backfire and in turn create more unemployed workers.

The dynamic setting of our model also allows us to look into the effect of higher unemployment benefit on workers in different stage of their career path. Figure 9b shows the impact of a 10% increase in unemployment insurance on worker's compensation plan and effort level. The policy affects junior workers the most, causing them to lose around 0.5% of their labor income and inducing them to put in 0.4% less effort in first tenure quarter. But the impact is minor for workers staying in their job for more than 3 years.



(A) Impact on aggregate employment



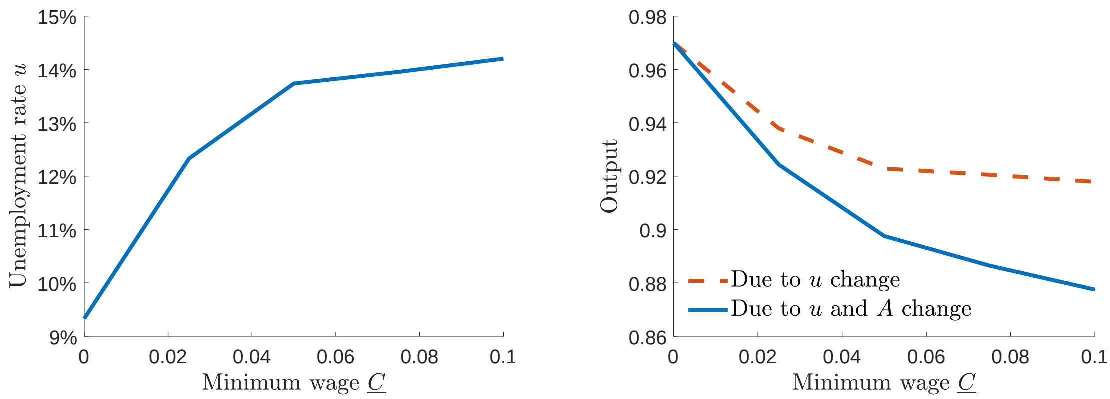
(B) Impact for different tenure workers

FIGURE 9: Changing unemployment insurance plan

Notes: Panel A shows that a more generous unemployment insurance b can lead to a higher unemployment rate with the moral hazard problem taken into account. As shown in panel B, the impact of a higher unemployment insurance is heterogeneous to workers in different tenure years: the junior workers are losing more compensation and offering less effort in the equilibrium.

6.2 Minimum wage

Textbook labor supply and demand curves suggest that setting a minimum wage may distort the labor market equilibrium and drive up unemployment. In the presence of a long-term incentive contract, the entire history of worker's output performance matters for the compensation, and all employed workers could hit the lower bound and receive the minimum wage at any time. Since workers are guaranteed a minimum labor income, pushing shirking becomes more difficult, thus exacerbating the moral hazard problem. Figure 10 shows that the unemployment rate grows rapidly when the government raises the minimum wage. The right panel shows that total output will decrease rapidly when minimum wage increases. Specifically, the red dotted curve represents the output decrease due to a higher unemployment rate when we shut down the incentive channel. The gap between the red dotted line and the blue solid line in the right panel represents the incentive distortion that causes less effort input and thus further reduced output.



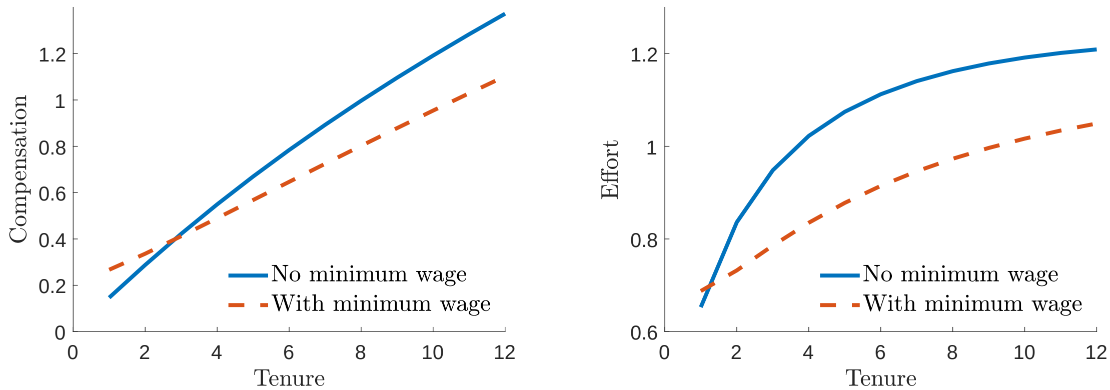
(A) Impact on employment

(B) Impact on output

FIGURE 10: Elasticity of unemployment and output to the change in the minimum wage

Notes: Panel A shows that a higher minimum wage requirement can lead to a higher unemployment rate with the moral hazard problem. Panel B shows that a higher minimum wage lowers the output level through two channels: employed workers are fewer and effort provided is less.

Figure 11 shows the impact of imposing minimum wage policy on workers in different stages of their tenure clock. Intuitively, when the government lifts the minimum wage floor, the more junior workers will benefit from the policy but senior worker will be hurt because the incentive distortion will dominate. Consider a minimum wage requirement that is 10% higher than unemployment benefit is now imposed. Then as shown in Figure 11, for first year worker, the compensation is in general higher than before, but for seasoned workers, both compensation and effort level are reduced.



(A) Impact on compensation plan

(B) Impact on effort level

FIGURE 11: Changing minimum wage

Notes: Consider a 10% rise in minimum wage. Panel A shows that junior workers will earn more than before but senior workers will earn less than before. Panel B shows that junior workers work harder but senior workers shirk more than before.

7. CONCLUSION

This paper develops a general equilibrium framework that incorporates the repeated moral hazard problem into a competitive labor market. The usage of block recursive equilibrium and the continuous-time method delivers a neat and tractable result even in a dynamic setting. We bring the model to micro-level data and quantify the model using labor market facts and indirect inference that estimates the variance of idiosyncratic factors. The quantified model has some implications for key macro labor market variables. First, when we account for the moral hazard problem, the counter-cyclical information rent on the worker side will create a more volatile unemployment rate and counter-cyclical wage dispersion. We quantify the size of unemployment rate volatility, which is over 8 times larger in the moral hazard model than that in information frictionless model. It can account for over 70% of the observed unemployment rate volatility in the data. Second, we quantify the correlation between wage dispersion and output, which is −0.34, showing strong countercyclicality. Third, this paper also discusses individual worker wage dynamics. The model-simulated sample generates a 54% initial-year wage scarring effect, and workers linger in the low wage zone for 8.7 quarters.

Future research could explore and advance in the following directions. To maintain model simplicity and deliver a clear message about how the moral hazard problem interplays with labor search frictions, we deliberately abstract from on-the-job search. A fully developed model could also take on-the-job search into consideration. An educated guess is that the moral hazard problem would be aggravated since this environment provides workers with better outside options. The estimation part in this paper could also be done in more cautious way. We used general accessible PSID data to estimate the idiosyncratic factors, but arguably, firms may have firm-level observable shocks that could affect a worker's labor income, and such shocks are observable to firms but unobservable to econometricians. An employer-employee matched dataset could mitigate this concern, and the firm-level average wage for workers could be used as an identifier for firm-level observable shocks.

References

Barnette, J., and A. Michaud (2011): "Wage scars from job loss," Discussion paper.

Bils, M., Y. Chang, and S.-B. Kim (2022): "How sticky wages in existing jobs can affect hiring," *American Economic Journal: Macroeconomics*, 14(1), 1–37.

Bolton, P., H. Chen, and N. Wang (2011): "A unified theory of Tobin's q, corporate investment, financing, and risk management," *The journal of Finance*, 66(5), 1545–1578.

Couch, K. A., and D. W. Placzek (2010): "Earnings losses of displaced workers revisited," *American Economic Review*, 100(1), 572–89.

DeMarzo, P. M., M. J. Fishman, Z. He, and N. Wang (2012): "Dynamic agency and the q theory of investment," *The Journal of Finance*, 67(6), 2295–2340.

DeMarzo, P. M., and Y. Sannikov (2006): "Optimal security design and dynamic capital structure in a continuous-time agency model," *The Journal of Finance*, 61(6), 2681–2724.

Diamond, P. A. (1982): "Wage determination and efficiency in search equilibrium," *The Review of Economic Studies*, 49(2), 217–227.

Dixit, A. (2013): *The art of smooth pasting*. Routledge.

Doligalski, P., A. Ndiaye, and N. Werquin (2020): "Redistribution with performance pay," *Available at SSRN 3581882*.

Gabaix, X. (2009): "Power laws in economics and finance," *Annu. Rev. Econ.*, 1(1), 255–294.

Gertler, M., and A. Trigari (2009): "Unemployment fluctuations with staggered Nash wage bargaining," *Journal of political Economy*, 117(1), 38–86.

Grochulski, B., and Y. Zhang (2019): "Termination as an incentive device," .

Guvenen, F., S. Ozkan, and J. Song (2014): "The nature of countercyclical income risk," *Journal of Political Economy*, 122(3), 621–660.

Hagedorn, M., and I. Manovskii (2008): "The cyclical behavior of equilibrium unemployment and vacancies revisited," *American Economic Review*, 98(4), 1692–1706.

Hall, R. E. (2005): "Employment fluctuations with equilibrium wage stickiness," *American economic review*, 95(1), 50–65.

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic dynamics*, 13(1), 15–51.

Holmstrom, B., and P. Milgrom (1987): "Aggregation and linearity in the provision of intertemporal incentives," *Econometrica: Journal of the Econometric Society*, pp. 303–328.

Jacobson, et al. (1993): "Earnings losses of displaced workers," *The American economic review*, pp. 685–709.

Kennan, J. (2010): "Private information, wage bargaining and employment fluctuations," *The Review of Economic Studies*, 77(2), 633–664.

Lamadon, T. (2016): "Productivity shocks, long-term contracts and earnings dynamics," *manuscript, University of Chicago*.

Lemieux, T., W. B. MacLeod, and D. Parent (2009): "Performance pay and wage inequality," *The Quarterly Journal of Economics*, 124(1), 1–49.

Li, Y., and C. Wang (2018): "Endogenous Labor Market Cycles," *Available at SSRN 3251664*.

Løkka, A. (2004): "Martingale representation of functionals of Lévy processes," *Stochastic Analysis and Applications*, 22(4), 867–892.

Menzio, G., and S. Shi (2010): "Block recursive equilibria for stochastic models of search on the job," *Journal of Economic Theory*, 145(4), 1453–1494.

——— (2011): "Efficient search on the job and the business cycle," *Journal of Political Economy*, 119(3), 468–510.

Moen, E. R. (1997): "Competitive search equilibrium," *Journal of political Economy*, 105(2), 385–411.

Moen, E. R., and Å. Rosén (2011): "Incentives in competitive search equilibrium," *The Review of Economic Studies*, 78(2), 733–761.

Mortensen, D. T., and C. A. Pissarides (1994): "Job creation and job destruction in the theory of unemployment," *The review of economic studies*, 61(3), 397–415.

Patterson, C. (2023): "The matching multiplier and the amplification of recessions," *American Economic Review*, 113(4), 982–1012.

Payne, J. (2018): "The disruption of long term bank credit," Discussion paper.

Phelan, C., and R. M. Townsend (1991): "Computing multi-period, information-constrained optima," *The Review of Economic Studies*, 58(5), 853–881.

Phelan, T. (2017): "On the optimal degree of inequality in business income," .

Sannikov, Y. (2008): "A continuous-time version of the principal-agent problem," *The Review of Economic Studies*, 75(3), 957–984.

Shimer, R. (2005): "The cyclical behavior of equilibrium unemployment and vacancies," *American economic review*, 95(1), 25–49.

Souchier, M. (2022): "The Pass-through of Productivity Shocks to Wages and the Cyclical Competition for Workers," Discussion paper, Working Paper.

Spear, S. E., and S. Srivastava (1987): "On repeated moral hazard with discounting," *The Review of Economic Studies*, 54(4), 599–617.

Storesletten, K., C. I. Telmer, and A. Yaron (2004): "Cyclical dynamics in idiosyncratic labor market risk," *Journal of political Economy*, 112(3), 695–717.

Tsuyuhara, K. (2016): "Dynamic contracts with worker mobility via directed on-the-job search," *International Economic Review*, 57(4), 1405–1424.

Appendices

A. Information structure

In the main paper, we described how firms receive the information about the output process Y_t but cannot distinguish the idiosyncratic factor dB_t and the worker's unobservable effort A_t. This section we will elaborate on the information structure of firm-worker pair and be specific about firm's and worker's information set.

Measurement. Let's first take a stand on measurement: assume the physical world idiosyncratic factor is dB_t, and B_t is a Brownian motion which generates sigma-algebra $\mathcal{F}_t$.

$$\mathcal{F}_t = \sigma\{B_s, s \leq t\}$$

Accordingly we denote the trajectory of idiosyncratic factor as $B_t = \omega^B(t)$ and the set of all possible trajectories to be $\Omega^B(t)$. We use this as our coordinate functions.

It is also known that any trajectory of idiosyncratic factor realization can be described by some elementary operation of cylinder sets C_t^B defined as

$$C_t^B(t_i, \mathcal{B}_i, i = 1, ...N) = \{\omega^B(t) \in \Omega^B(t) \mid B_{t_i} \in \mathcal{B}_i, i = 1, ..., N, t_i \in [0, t]\}$$

$\mathcal{B}_i$ are Borel measurable sets. N is the number of possible events. It is sufficient to define the probability measure $Pr\{\cdot\}$ on cylinder sets C_t^B and then to extend it to all events and all trajectories in $\mathcal{F}$ by the elementary properties of a probability measure. Now our task is to specify the probability $Pr\left(C_t^B(t_i, \mathcal{B}_i, i = 1, ...N)\right)$ for any points $t_1, ..., t_N \in [0, t]$ and Borel measurable sets $\mathcal{B}_1, ..., \mathcal{B}_N \subseteq \mathbb{R}$.

Let us for now denote the "objective" measure by $\mathbb{P}_t = \{P_t^B\}$, and by the way we define $\mathbb{P}_t$, B_t is obviously the Brownian motion under measure $\mathbb{P}_t$.

$$P_t^B = Pr\left(C_t^B(t_i, \mathcal{B}_i, i = 1, ...N)\right) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi t_i}} \int_{\mathcal{B}_i} \exp\left(-\frac{B^2}{2t_i}\right) dB$$

Similarly, the trajectory of output Y_t can be described by some elementary operations of cylinder sets C_t^Y

$$C_t^Y(t_i, \mathcal{B}_i, i = 1, ...N) = \{\omega^Y(t) \in \Omega^Y(t) \mid Y_{t_i} \in \mathcal{B}_i, i = 1, ..., N, t_i \in [0, t]\}$$

The worker can perfectly observe B_t, and he knows his own action A_t, so he can construct a "ture" probability measure $\mathbb{Q}_t^W = \{P_t^Y\}$ over Y_t from probability measure $\mathbb{P}_t$.

Since the increment of Brownian motion is independent, given $dY_t = zA_t dt + \sigma dB_t$, we can conclude that the conditional distributions of output with respect to action $G(dY_t \mid A_t)$ are independent across time.

$$G(dY_t \mid A_t) = G(zA_t dt + \sigma dB_t \mid A_t)$$

$$G(Y_t \mid A_{0\le s\le t}) = G(Y_0 + \int_0^t zA_s ds + \sigma B_t \mid A_{0\le s\le t})$$

$$P_t^Y = Pr\left(C_t^Y(t_i, \mathcal{B}_i, i = 1, ...N)\right) z = \prod_{i=1}^N \frac{1}{\sqrt{2\pi t_i}\sigma} \int_{\mathcal{B}_i} \exp\left(-\frac{(Y - \int_0^t zA_{s_i} ds - Y_0)^2}{2\sigma^2 t_i}\right) dY$$

Later when we calculate the expected discount utility for worker, we need to do it under $\mathbb{Q}^W$ measurement, because all we care about is the *probability distribution of* Y_t *process*, which not only depends on exogenous B_t process but also on endogenous worker's choice of A_t process.

On the other hand, firms cannot observe B_t but know the probability distribution of B_t trajectory. He assumes strategy profile taken by worker as $\{A_t^F\}$ (denoted in the sense of firm recommend effort) thus constructs his "subjective" probability measure over output process Y_t. Denote the probability measure of output used by firm to be $\mathbb{Q}^F = \{\widetilde{P}_t^Y\}$

$$G(dY_t \mid A_t^F) = G(zA_t dt + \sigma dB_t \mid A_t^F) = G(zA_t^F dt + \sigma dB_t^F)$$

$$G(Y_t \mid A_{0\le s\le t}^F) = G(Y_0 + \int_0^t zA_s^F ds + \int_0^t \sigma B_s^F \mid A_{0\le s\le t}^F)$$

$$\widetilde{P}_t^Y = Pr\left(C_t^Y(t_i, \mathcal{B}_i, i = 1, ...N)\right) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi t_i}\sigma} \int_{B_i} \exp\left(-\frac{(Y - \int_0^t zA_{s_i}^F ds - Y_0)^2}{2\sigma^2 t_i}\right) dY$$

Shifted process. Ex post, if the firm observes output process Y_t, for each effort strategy $(A_t^F)_{t\ge 0}$ assumed by the firm (equivalently, the choice of deterministic function $(A_t^F)_{t\ge 0}$), the process $B^F = (B_t^F)_{t\ge 0}$ can be referred pointwise by the firm reciprocally,

$$B_t^F = \frac{1}{\sigma}\left(Y_t - \int_0^t zA_s^F ds\right)$$

for each possible output trajectory $\omega^Y \in \Omega^Y$. We call B_t^F firm's subjective shifted diffusion term.

Effort process and induced measure. Noted that a deviation of worker's strategy is essentially a change in the measure that worker can use to evaluate the probability of the occurrence of certain paths of output. Because worker can take any arbitrary action A_t which is not necessarily A^F, there may exist a gap between objective diffusion term dB_t and firm's subjective shifted diffusion term dB_t^F even ex ante,

$$dB_t^F = \frac{1}{\sigma}(dY_t - zA_t^F dt) = \frac{1}{\sigma}(zA_t dt + \sigma dB_t - zA_t^F dt) = \frac{1}{\sigma}(zA_t dt - zA_t^F dt) + dB_t$$

B. Contracting problem

B.1 Perfect information contract

Proposition B.1. *With perfect information, workers exert constant effort $A_{FB}(W)$ and are compensated with a constant stream of consumption $C_{FB}(W)$:*

$$\begin{aligned} A_{FB}(W) &= \underset{A\in\mathcal{A}}{\operatorname{argmax}} \frac{zA - u^{-1}\left[rW + \phi(A) - (W_u - W)\lambda_x\right]}{r + \lambda_x}, \\ C_{FB}(W) &= u^{-1}\left[rW + \phi(A_{FB}) - (W_u - W)\lambda_x\right]. \end{aligned}$$

Proof. Firm will offer a first best contract that maximizes his profit

$$\max_C \mathbb{E}\left[-\int_0^\tau e^{-rt}(zA - C)dt\right]$$

subject to promise keeping condition towards the retired worker

$$\mathbb{E}\left[r\int_0^\tau e^{-rt}\left(u(C) - \phi(A)\right)dt + e^{-r\tau}W_u\right] \geq W$$

The dynamics of continuation value should not load on the idiosyncratic factor

$$dW = \left[r\left(W - u(C) + \phi(A)\right) - (W_u - W)\lambda_x\right]dt$$

The HJB equation for the firm should be

$$(r + \lambda_x)F(W) = \max_{A,C} \mathbb{E}\left[zA - C + F'(W)\left[r\left(W - u(C) + \phi(A)\right) - (W_u - W)\lambda_x\right]\right] \tag{12}$$

The optimality conditions are

$$z + rF'(W)\phi'(A) = 0, \quad -1 - rF'(W)u'(C) = 0$$

Now we guess W is not changing overtime in the perfect information case, thus by HJB equation, we will have the value function F being

$$F(W) = \frac{zA - C}{r + \lambda_x}$$

and the drift of W equals to zero

$$r\left(W - u(C) + \phi(A)\right) - (W_u - W)\lambda_x = 0 \tag{13}$$

$$W = \frac{r\left[u(C) - \phi(A)\right] + \lambda_x W_u}{r + \lambda_x}$$

further A and C take values that maximize $F(W)$ while deliver the promised utility

$$F_{FB}(W) = \max_{A \in \mathcal{A}} \frac{zA - u^{-1}\left[W + \phi(A) - (W_u - W)\frac{\lambda_x}{r}\right]}{r + \lambda_x}$$

The first order condition with respect to A is

$$z - \frac{\partial u^{-1}\left[W + \phi(A) - (W_u - W)\frac{\lambda_x}{r}\right]}{\partial\left[W + \phi(A) - (W_u - W)\frac{\lambda_x}{r}\right]}\phi'(A) = 0$$

Notice this is equivalent to

$$z - \frac{1}{u'(C)}\phi'(A) = 0$$

Thus the conjecture value function satisfies following optimality condition

$$\phi'(A) = zu'(C) \tag{14}$$

Now we verify the two optimality conditions in the original HJB equation (12). Notice that if the conjecture is right, then by (14), we only need to verify one of the optimality conditions. Take the derivative of (13) with respect to W is

$$r\left(1 - u'(C)\frac{\partial C}{\partial W} + \phi'(A)\frac{\partial A}{\partial W}\right) + \lambda_x = 0$$

which is equivalent to

$$-z\frac{\partial A}{\partial W} + \frac{\partial C}{\partial W} = \frac{\lambda_x + r}{ru'(C)}$$

So when we take first derivative of the conjectured value function, it gives us following equation

$$F'(W) = -\frac{1}{r + \lambda_x}\frac{r + \lambda_x}{ru'(C)} = -\frac{1}{ru'(C)}$$

Apparently it satisfies the first order conditions for original HJB equation. $\square$

B.2 Retirement plan

Lemma B.1. *If the firm retires the worker, the worker will receive constant consumption C_R overtime and the firm payoff will be $F_R(W)$.*

$$C_R(W) = u^{-1}\left[W - \frac{\lambda_x}{r}(W_u - W)\right], \quad F_R(W) = -\frac{u^{-1}\left[W - \frac{\lambda_x}{r}(W_u - W)\right]}{r + \lambda_x}$$

Proof. Firm will offer a retirement contract that maximizes his profit

$$\max_C \mathbb{E}\left[-\int_0^\tau e^{-rt} C dt\right]$$

subject to promise keeping to the retired worker

$$\mathbb{E}\left[r\int_0^\tau e^{-rt} u(C)\, dt + e^{-r\tau} W_u\right] \geq W$$

The promised utility satisfies a recursive equation and the continuous time version of it defines the law of motion of worker's promised utility

$$W = ru(C_R)\, dt + e^{-rdt}\left[(1 - \lambda_x dt)(W + dW) + \lambda_x dt W_u\right]$$

$$dW = \left[rW - \lambda_x (W_u - W) - ru(C_R)\right] dt$$

Drift of W is zero because promised utility for retired worker is constant overtime.

$$F_R(W) = -C_R dt + e^{-rdt}\left[(1 - \lambda_x dt)\left(F_R(W) + F_R'(W)\, dW\right)\right]$$

$$F_R(W) = \frac{-C_R}{r + \lambda_x}$$

□

B.3 Evolution of worker's promised utility

Under a given contract C_t, the worker can arbitrarily choose effort a_t and the worker's promised utility under such plan is defined as W_t^a

$$W_t^a = \mathbb{E}\left[r\int_t^\tau e^{-r(s-t)}\left[u(C_s) - \phi(a_s)\right] ds + e^{-r(\tau - t)} W_u\right]$$

Consider a worker's expected utility when taking an arbitrary action $a_{[0,\infty)}$,

$$\begin{aligned} V_t^a &= \left[r\int_0^t e^{-rs}\left[u\left(C_s\right)-\phi\left(a_s\right)\right]ds\right]+e^{-rt}\mathbb{E}\left[r\int_t^\tau e^{-r(s-t)}\left[u\left(C_s\right)-\phi\left(a_s\right)\right]ds+e^{-r(\tau-t)}W_u\right] \\ &= \left[r\int_0^t e^{-rs}\left[u\left(C_s\right)-\phi\left(a_s\right)\right]ds\right]+e^{-rt}W_t^a. \end{aligned} \tag{15}$$

Before the separation shock hits the employment match, the change of total expected value of the worker can be replicated (hedged) by loading on all realized martingale processes in this economy up to time t: idiosyncratic factor, aggregate shock, and compensated Poisson separation shock. The procedure is so called "Martingale representation theorem" formally proved in Løkka (2004).

$$V_t^a = V_0^a + r\int_0^t e^{-rs}\Psi_{b,s}\sigma_B dB_s + \int_0^t e^{-rs}\Psi_{x,s}\left(dN_s^x - \lambda_x ds\right) \tag{16}$$

Where Ψ_b, Ψ_x are respectively the sensitivity of the worker's continuation value to idiosyncratic factors, and to compensated separation shock. e^{-rt} is a convenient multiplier on these sensitivities. Taking derivative of (15) and (16) leads to

$$dW_t^a = r\left[W_t^a - u\left(C_t\right)+\phi\left(a_t\right)\right]dt + r\Psi_{b,t}\underbrace{\left(dY_t - z_t a_t dt\right)}_{\sigma dB_t} + \Psi_{x,t}\left(dN_{x,t}-\lambda_x dt\right)$$

B.4 Incentive compatibility constraint

Now we show that the incentive-compatible recommended action has to satisfy one shot deviation property. Consider worker's time t total expected discounted utility G_t that derived from strategy profile: $\left\{a_{[0,t)}, A_{[t,\infty)}\right\}$, where A_t is the firm recommended actions and W_t is the continuation value if worker follows firm suggested action.

$$V_t = \left[r\int_0^t e^{-rs}\left[u\left(C_s\right)-\phi\left(a_s\right)ds\right]ds\right]+e^{-rt}W_t$$

The incentive compatible condition at t is the same as Sannikov (2008)

$$A_t \in \arg\min_{a\in\mathcal{A}} r\phi\left(a\right) - r\Psi_{b,t}a_t z_t$$

The first order approach gives us the sufficient sensitivity of worker continuation value to idiosyncratic factors in order to provide worker enough incentive,

$$\Psi_{b,t} = \frac{\phi'\left(A_t\right)}{z_t}$$

We will discuss more on incentive compatibility condition in lemma B.2.

Lemma B.2. *An allocation $(C_t, A_t)_{t\geq 0}$ is incentive compatible if and only if the sensitivity process $\Psi_{b,t}$ satisfies*

$$\phi(A_t) - \Psi_{b,t} A_t z_t \leq \phi(a_t) - \Psi_{b,t} a_t z_t, \qquad \forall a \in \mathcal{A},\ 0 \leq t < \infty \tag{17}$$

almost everywhere.

Proof. In order to discuss worker's incentive, we should take worker's perspective to evaluate the contract. For any arbitrary strategy a taken by the worker and time $t \geq 0$, define a process $\widehat{V} := (\widehat{V}_t)_{t\geq 0}$ to be time-t expectation of the worker's total pay-off if he experienced the cost of effort from the strategy a before time t, and plans to follow the strategy A after time t.

$$\widehat{V}_t = r \int_0^t e^{-rs} \left[u(C_s) - \phi(a_s)\right] ds + e^{-rt} W_t$$

$\widehat{V}_t$ is defined in the spirit of one shot deviation principle in order to find the optimal plan in subgame perfect equilibrium (**SPE**). W_t represents the continuation utility if the worker sticks to A after t. Let us identify the drift of the process $\widehat{V}_t$ under the probability measure $\mathbb{Q}^A$.

It is obvious, by construction, the continuation value of the following two strategies: (a) taking action a_s in time interval $(0, t]$, and later switching to action A_s for time period (t, ∞), and (b) taking A_s at all time $(0, \infty)$, are the same. So the law of motion for the promised utility in both strategies are the same.

Notice here when we apply martingale representation theorem, the differential of $\widehat{V}_t$ satisfies

$$\begin{aligned} d\widehat{V}_t &= re^{-rt} \left[u(C_t) - \phi(a_t)\right] dt - re^{-rt} W_t dt + e^{-rt} r[W_t - (u(C_t) - \phi(A_t))]dt + e^{-rt} r\Psi_{b,t}\sigma dB_t \\ &= re^{-rt} \left[\phi(A_t) - \phi(a_t)\right] dt + re^{-rt}\Psi_{b,t} z(a_t - A_t)dt + re^{-rt}\Psi_{b,t}\sigma dB_t^A \end{aligned}$$

Then $re^{-rt} \left[\left(\phi(A_t) - \Psi_{b,t} z A_t\right) - \left(\phi(a_t) + \Psi_{b,t} z a_t\right)\right]$ is the drift of process $\widehat{V}_t^A$ under the probability measure $\mathbb{Q}^A$.

To prove "if"

Now suppose the inequality (17) failed to hold on a set of positive measure and define $\overline{A} = (\overline{A}_t)_{t\geq 0}$ to be any process satisfying

$$-\Psi_{b,t} z \overline{A}_t + \phi(\overline{A}_t) = \max_{A' \in \mathcal{A}} -\Psi_{b,t} z A' + \phi(A')$$

for all $t \geq 0$ almost surely. That is to say, the opposite of (17) holds for a set of positive measure if taking $A = \overline{A}$. Explicitly,

$$\left[\phi(A_t) - \Psi_{b,t} z A_t\right] - \left[\phi(\overline{A}_t) - \Psi_{b,t} z \overline{A}_t\right] > 0 \tag{18}$$

holds on a set of positive measure. Then the drift of $\widehat{V}_t^{\overline{A}}$ (under $\mathbb{Q}_t^{\overline{A}}$) is positive on a set of positive measure. Thus there exist a time $T > 0$ such that

$$\mathbb{E}^{\overline{A}}\left[\widehat{V}_T^{\overline{A}}\right] > \widehat{V}_0^{\overline{A}} = \widehat{V}_0^{A} = W_0(C, A)$$

That is to say for the time sets that (17) does not hold, if worker deviates from A to $\overline{A}$, his momentary drift of utility $\widehat{V}^A$ will be positive by (18). The worker gets utility $\mathbb{E}^{\overline{A}}\left[\widehat{V}_T^{\overline{A}}\right]$ if he follows $\overline{A}$ up to time t and then switches back to A. The utility is higher than sticking with A all the time. Then the strategy A is sub-optimal. We get a contradiction.

To prove "only if"

Suppose (17) holds for strategy A, then $\widehat{V}_t$ is a $\mathbb{Q}^A$- supermartingale for any alternative strategy a. Moreover, since the process $W(C, a)$ is bounded from below, we can add

$$\widehat{V}_\infty = r\int_0^\infty e^{-rs}\left[u(C_s) - \phi(a_s)\right] ds$$

as the last element of the super-MTG $\widehat{V}^a$. Therefore,

$$W_0(C, A) = \widehat{V}_0^F = \widehat{V}_0^a \geq \mathbb{E}^a[\widehat{V}_\infty] = W_0(C, a)$$

so the strategy A is at least as good as any alternative strategy a. □

B.5 Lower bound condition

In this section we will show that the lower bound condition is characterized by following equation,

$$F'(W_u) = \frac{F(W_u) - F_R(\frac{\lambda_x W_u}{r+\lambda_x})}{W_u - \frac{\lambda_x W_u}{r+\lambda_x}}$$

The recursive problem of the operating firm that can temporarily suspend the worker is

$$F(W_t) = \max\begin{Bmatrix}\max_{A,C,\Psi_b}\left\{(z_t A_t - C_t)\, dt + e^{-rdt} F(W_{t+dt})(1-\lambda_x dt)\right\}, \\ \max_C\left\{-C_t + e^{-rdt} F(W_{t+dt})(1-\lambda_x dt)\right\}\end{Bmatrix}$$

We will study the two options as two separate optimizing problem and combine their solutions to derive an optimal contract. Denote the firm's payoff when the worker takes low effort ($A = 0$) to be $L(W)$, and the firm's payoff when the worker takes non-zero effort to be $H(W)$. Denote the corresponding low-effort region of promised utility to be $\mathcal{W}^L = [W_u, W_s)$ and the high-effort region of promised utility to be $\mathcal{W}^H = [W_s, W_r]$. The firm's problem under high- and low- effort region can be summarized by two separate ordinary differential equations (ODEs). We refer the first ODE as high-effort ODE, and the second ODE as low-effort ODE

$$
\begin{aligned}
(r+\lambda_x) H(W_t) = \max_{A,C,\Psi_b} \quad & (z_t A_t - C_t) + H'(W_t)\left[r\left(W_t - u(C_t) + \phi(A_t)\right) - (W_u - W_t)\lambda_x\right] \\
& + \frac{H''(W_t)}{2}(r\Psi_{b,t}\sigma)^2 \\
(r+\lambda_x) L(W_t) = \max_{C} \quad & -C_t + L'(W_t)\left[r(W_t - u(C_t)) - \lambda_x(W_u - W_t)\right]
\end{aligned} \tag{19}
$$

We demonstrates the firm's profit under the optimal contract in Figure 12. In low effort region $\mathcal{W}^L$, the evolution of promised utility is deterministic and the firm's profit is a linear function in promised utility (see details in Lemma B.3). In high effort region $\mathcal{W}^H$, the evolution of promised utility is stochastic, and the firm's profit is hump shaped.

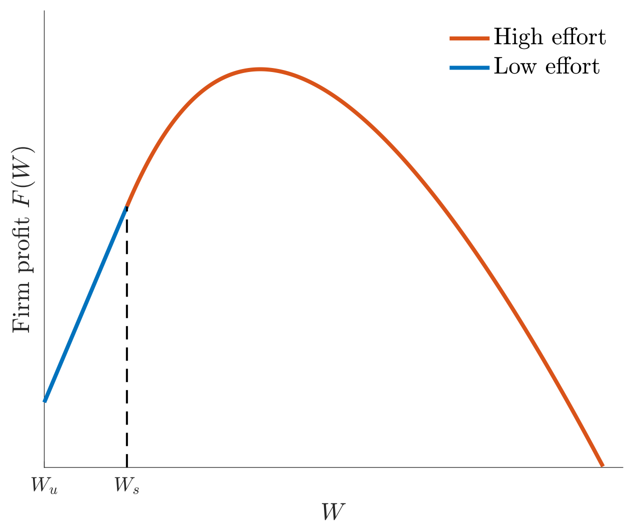


FIGURE 12: Low effort and high effort region

Lemma B.3. *The set of solutions $L(W)$ to Equation (19) consists of all straight lines with nonnegative slope passing through the point $\left(\frac{\lambda_x W_u}{r+\lambda_x}, F_R(\frac{\lambda_x W_u}{r+\lambda_x})\right)$.*

Proof. Intuitively, the solution to low-action ODE is to set $C_t = 0$ (taking first order derivative on the right hand side of low-action ODE). Now let's formally prove the solution and get more characteristics of value function $L(W)$. Define $\widetilde{W}$ as $u(C)$, the low-action ODE becomes

$$
(r+\lambda_x) L(W_t) = \max_{\widetilde{W}\geq 0} \ (r+\lambda_x) F_R\left(\frac{r\widetilde{W} + \lambda_x W_u}{r+\lambda_x}\right) + L'(W_t)\left[r\left(W_t - \widetilde{W}_t\right) - \lambda_x(W_u - W_t)\right] \tag{20}
$$

We will look for solutions to this ODE in the region $\mathcal{R} := \{(W, L) : L > F_R(W)\}$. If we denote the maximizer in (20) by W^*, the low-action ODE can be transformed into

$$(r+\lambda_x)\,L(W_t) = (r+\lambda_x)F_R(\frac{r\widetilde{W}+\lambda_x W_u}{r+\lambda_x}) + L'(W_t)\,[(r+\lambda_x)\,W_t - rW^* - \lambda_x W_u]$$

If $W^* = \frac{(r+\lambda_x)W-\lambda_x W_u}{r}$, then $L(W) = F_R(\frac{rW^*+\lambda_x W_u}{r+\lambda_x}) = F_R(W)$, which is not in $\mathcal{R}$. So $W^* \neq \frac{(r+\lambda_x)W-\lambda_x W_u}{r}$, then $L(W)$ is a straight line with slope $L'(W) \geq 0$. [15] The optimal W^* in (20) is $W^* = 0$ and $F_R(\frac{rW^*+\lambda_x W_u}{r+\lambda_x}) = F_R(\frac{\lambda_x W_u}{r+\lambda_x})$, so (20) reduces to

$$L(W_t) - F_R(\frac{\lambda_x W_u}{r+\lambda_x}) = L'(W_t)\left[W_t - \frac{\lambda_x}{r+\lambda_x}W_u\right]$$

Thus the set of solutions with nonnegative slope consists of all straight lines going through the point $(\frac{\lambda_x}{r+\lambda_x}W_u, F_R(\frac{\lambda_x W_u}{r+\lambda_x}))$. □

The optimal contract is obtained when W_s is set as low as possible because a larger support $[W_s, W_r]$ could sustain positive effort for a longer period. With a larger high-effort support, the volatility in the continuation value will not drive W_t to collide with W_s as quickly. Therefore, the low-effort region will degenerate into one point at W_u. To guarantee the consistency of the contract over the entire contracting space, the smooth pasting condition must be satisfied at point W_u.

B.6 Worker's distribution

In the steady state, the stationary distribution of employed workers $G(W_t)$ satisfies

$$\begin{aligned} 0 = & \underbrace{-\frac{\partial}{\partial W}\left[\left(r(W_t - u(C_t) + \phi(A_t)) - (W_u - W_t)\lambda_x\right) G(W_t)\right]}_{\substack{\text{Change in density from} \\ \text{the drift of promised utility}}} \\ & \underbrace{+\frac{1}{2}\frac{\partial^2}{\partial W^2}\left[\left(r\frac{\phi'(A_t)}{z_t}\sigma\right)^2 G(W_t)\right]}_{\substack{\text{Change in density from} \\ \text{the volatility of promised utility}}} - \underbrace{\lambda_x G(W_t)}_{\substack{\text{Exit from} \\ \text{exo. termination}}} \end{aligned} \tag{21}$$

By imposing proper boundary conditions, (21) can be solved. Here the lower bound is reflective and the upper bound is absorbing, then according to Dixit (2013) the corresponding boundary conditions are

$$\frac{\partial G(W_u)}{\partial W} = 0, \quad G(W_r) = 0.$$

[15]Cases where $L'(W) < 0$ are of less interests because they will not be part of the optimal contract.

B.7 Determination of the lower bound

The expected utility for an individual unemployed worker is $W_{u,t}$ and the promised utility for newly hired worker is $W_{0,t}$

$$W_{u,t} = ru\left(b\right)dt + e^{-rdt}\left[(1-p(\theta_t)dt)W_{u,t+dt} + p(\theta_t)dtW_{0,t+dt}\right]$$

At the steady state, $W_{u,t}$ and $W_{0,t}$ are constant. In the limit, $e^{-rdt} = 1 - rdt$

$$W_u = ru\left(b\right) + pdtW_0 + \left(1 - rdt - p(\theta)dt\right)W_u$$

The expected utility for the unemployed workers consists of two parts: utility from unemployment benefit, and utility from job searching.

$$rW_u = ru(b) + p(\theta)\left(W_0 - W_u\right)$$

C. Dynamic model analysis

Now we include regime switching productivity shock into the analysis. Productivity shock is publicly observable to all: z_t has two possible states, $z \in \{z_G, z_B\}$. The state switching follows a continuous time two-state Markov transition process with switching rate $\lambda_z\left(z_t\right)$. $dN_{z,t} = 1$ describes the times at which z state changes. The cumulative output is Y_t

$$dY_t = \left(z_tA_tdt + \sigma dB_t\right)\mathbb{1}_{[\tau,\infty)}$$

where A is unobservable labor hour input by worker. τ is the separation time.

C.1 Retirement plan

The firm can retire an worker at any time. Once worker is retired, he will exert zero effort and receive a constant flow of consumption. Denote the continuation value under aggregate state z to be W_z, and that under aggregate state z^c to be W_{z^c}. Denote the unemployment expected payoff under z and z^c to be respectively $W_u\left(z\right)$, $W_u\left(z_c\right)$.

If the aggregate state at the retirement point is z, then starting promised utility is W_z

$$u(C) = \frac{(\lambda_x + r)W_z}{r} - \frac{\lambda_x\left(\lambda_xW_u(z) + \lambda_z(z)W_u(z^c) + \lambda_z(z^c)W_u(z) + rW_u(z)\right)}{r\left(\lambda_x + \lambda_z(z) + \lambda_z(z^c) + r\right)}$$

$$F_R\left(W_z, z\right) = \frac{-C}{r + \lambda_x}$$

If the aggregate state at the retirement point is z^c, then starting promised utility is W_{z^c}

$$u(C) = \frac{(\lambda_x + r)W_{z^c}}{r} - \frac{\lambda_x\left(\lambda_x W_u(z^c) + \lambda_z(z)W_u(z^c) + \lambda_z(z^c)W_u(z) + rW_u(z^c)\right)}{r\left(\lambda_x + \lambda_z(z) + \lambda_z(z^c) + r\right)}$$

$$F_R\left(W_{z^c}, z^c\right) = \frac{-C}{r + \lambda_x}$$

Proof. The law of motion for continuation under both aggregate states are respectively

$$dW_z = \left[r\left(W_z - u(C)\right) - \lambda_x\left(W_u\left(z\right) - W_z\right) - \lambda_z(z)\psi_z\right] dt$$

$$dW_{z^c} = \left[r\left(W_{z^c} - u(C)\right) - \lambda_x\left(W_u(z^c) - W_{z^c}\right) - \lambda_z(z^c)\psi_{z^c}\right] dt$$

where $\psi_z = W_{z^c} - W_z$, $\psi_{z^c} = W_z - W_{z^c}$. And $\psi_{z^c} = -\psi_z$.

The HJB equations for a firm in retirement contract under both z and z^c states are

$$\begin{aligned}(r + \lambda_x + \lambda_z(z))\, F_R\left(W_z, z\right) = \max_{C,\psi_z} \quad & -C + \lambda_z(z)F_R\left(W_z + \psi_z, z^c\right) \\ & + F_R'\left(W, z\right)\left[r\left(W_z - u(C)\right) - \lambda_x\left(W_u(z) - W_z\right) - \lambda_z(z)\psi_z\right]\end{aligned} \tag{22}$$

$$\begin{aligned}(r + \lambda_x + \lambda_z(z^c))\, F_R\left(W_{z^c}, z^c\right) = \max_{C,\psi_{z^c}} \quad & -C + \lambda_z(z^c)F_R\left(W_{z^c} + \psi_{z^c}, z\right) \\ & + F_R'\left(W_{z^c}, z^c\right)\left[r\left(W_{z^c} - u(C)\right) - \lambda_x\left(W_u(z^c) - W_{z^c}\right) - \lambda_z(z^c)\psi_{z^c}\right]\end{aligned} \tag{23}$$

The first order conditions for (22) are

$$-1 - rF_R'(W, z)u'(C) = 0$$

$$-F_R'(W, z)\lambda_z(z) + \lambda_z(z)F_R'(W_z + \psi_z, z^c) = 0$$

The first order conditions for (23) are

$$-1 - rF_R'(W, z_c)u'(C) = 0$$

$$-F_R'(W, z_c)\lambda_z(z_c) + \lambda_z(z_c)F_R'(W_{z^c} + \psi_{z^c}, z) = 0$$

Use guess and verify method. We guess that the policy function C_t and ψ_t are chosen to make the drift of W_z and W_{z^c} equal to zero.

$$dW_z = \left[r\left(W_z - u(C)\right) - \lambda_x\left(W_u(z) - W_z\right) - \lambda_z(z)\psi_z\right] dt$$

$$dW_{z^c} = \left[r\left(W_{z^c} - u(C)\right) - \lambda_x\left(W_u(z^c) - W_{z^c}\right) - \lambda_z(z^c)\psi_{z^c}\right] dt$$

The linear equation system is

$$0 = r\left(W_z - u(C)\right) - \lambda_x\left(W_u(z) - W_z\right) - \lambda_z(z)\psi_z$$

$$0 = r\left(W_{z^c} - u(C)\right) - \lambda_x \left(W_u(z^c) - W_{z^c}\right) - \lambda_z(z^c)\psi_{z^c}$$

Jointly we can solve for $u(C)$ and ψ_z. Also we know $\psi_z = W_{z^c} - W_z = -\psi_{z^c}$.

$$u(C) = \frac{r\left(\lambda_z(z^c)W_z + \lambda_z(z)W_{z^c}\right) - \lambda_x\left[\lambda_z(z^c)(W_u(z) - W_z) + \lambda_z(z)(W_u(z^c) - W_{z^c})\right]}{r\left(\lambda_z(z) + \lambda_z(z^c)\right)} \tag{24}$$

$$\psi_z = \frac{-\lambda_x\left(W_u(z) - W_u(z^c)\right) + (r + \lambda_x)(W_z - W_{z^c})}{\lambda_z(z) + \lambda_z(z^c)}$$

$$\psi_z = -\frac{\lambda_x(W_u(z) - W_u(z^c))}{\lambda_x + \lambda_z(z) + \lambda_z(z^c) + r}$$

Given the drift term of W_z and W_{z^c} being zero, the HJB equations degenerate into

$$(r + \lambda_x + \lambda_z(z))\, F_R\left(W_z, z\right) = \max_{C,\psi_z} \quad -C + \lambda_z(z)F_R\left(W_z + \psi_z, z^c\right)$$

$$(r + \lambda_x + \lambda_z(z^c))\, F_R\left(W_{z^c}, z^c\right) = \max_{C,\psi_z} \quad -C + \lambda_z(z^c)F_R\left(W_{z^c} + \psi_{z^c}, z\right)$$

suppose C and ψ are already optimal choices, the HJB equation system delivers

$$F_R\left(W_z, z\right) = \frac{-C}{r + \lambda_x}, \quad F_R\left(W_{z^c}, z^c\right) = \frac{-C}{r + \lambda_x}$$

If the aggregate state at the retirement point is z, then starting promised utility is W_z. Replace W_{z^c} with $W_z + \psi_z$ in (24) we get the following,

$$u(C) = \frac{(\lambda_x + r)W_z}{r} - \frac{\lambda_x\left((\lambda_x + \lambda_z(z^c) + r)\,W_u(z) + \lambda_z(z)W_u(z^c)\right)}{r\left(\lambda_x + \lambda_z(z) + \lambda_z(z^c) + r\right)}, \tag{25}$$

$$F_R\left(W_z, z\right) = \frac{-C}{r + \lambda_x}.$$

If the aggregate state at the retirement point is z^c, then the starting promised utility is W_{z^c}

$$u(C) = \frac{(\lambda_x + r)W_{z^c}}{r} - \frac{\lambda_x\left((\lambda_x + \lambda_z(z) + r)\,W_u(z^c) + \lambda_z(z^c)W_u(z)\right)}{r\left(\lambda_x + \lambda_z(z) + \lambda_z(z^c) + r\right)},$$

$$F_R\left(W_{z^c}, z^c\right) = \frac{-C}{r + \lambda_x}.$$

To verify the solution is optimal, we check whether the optimality conditions are satisfied. Denote $u(C) = f(W)$ in (25).

$$F_R'\left(W_z, z\right) = \frac{-\left(u^{-1}\right)'\left(f(W)\right) f'(W)}{r + \lambda_x} = -\frac{\lambda_x + r}{r\left(r + \lambda_x\right) u'(C)} = -\frac{1}{ru'(C)}$$

where $f(W)$ is the right hand side of (25). Obviously, the allocation satisfies the first order conditions regarding C. At the same time $F_R'\left(W_z, z\right) = F_R'\left(W_{z^c}, z^c\right) = -\frac{1}{ru'(C)}$, so the optimality condition regarding ψ_z also satisfies. So we have verified the constant consumption plan is indeed the optimal retirement plan. $\square$

Figure 13 illustrates the firm's profit of a retirement contract. Intuitively, the firm's profit is larger under good aggregate state than that under bad aggregate state.

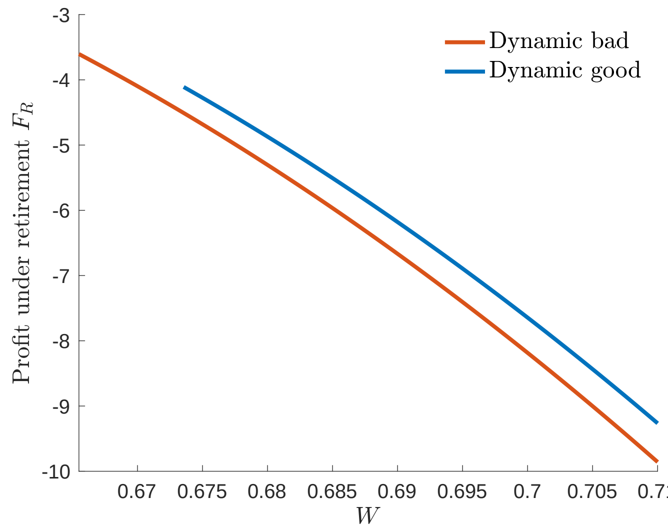


FIGURE 13: The elasticity of the unemployment rate with respect to z

C.2 Perfect information contract

The most effective way for the employer to deliver the promised utility is through compensating the worker a constant stream of consumption goods C and let the worker put in constant effort A despite of the aggregate state. Denote the continuation value under aggregate state z to be W_z, and that under aggregate state z^c to be W_{z^c}. And the unemployment expected payoff under z and z^c to be respectively $W_u(z)$, $W_u(z_c)$.

If the aggregate state at the retirement point is z, then starting promised utility is W_z

$$u(C) - \phi(A) = \frac{(\lambda_x + r)W_z}{r} - \frac{\lambda_x\left(\lambda_x W_u(z) + \lambda_z(z)W_u(z^c) + \lambda_z(z^c)W_u(z) + rW_u(z)\right)}{r\left(\lambda_x + \lambda_z(z) + \lambda_z(z^c) + r\right)}$$

$$F_{FB}\left(W_z, z\right) = \frac{zA - C}{r + \lambda_x}$$

If the aggregate state at the retirement point is z^c, then starting promised utility is W_{z^c}

$$u(C) - \phi(A) = \frac{(\lambda_x + r)W_{z^c}}{r} - \frac{\lambda_x\left(\lambda_x W_u(z^c) + \lambda_z(z)W_u(z^c) + \lambda_z(z^c)W_u(z) + rW_u(z^c)\right)}{r\left(\lambda_x + \lambda_z(z) + \lambda_z(z^c) + r\right)}$$

$$F_{FB}\left(W_{z^c}, z^c\right) = \frac{z^c A - C}{r + \lambda_x}$$

Proof. The law of motion for continuation under both aggregate states are respectively

$$dW_z = \left[r\left(W_z - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z) - W_z\right) - \lambda_z(z)\psi_z\right]dt$$

$$dW_{z^c} = \left[r\left(W_{z^c} - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z^c) - W_{z^c}\right) - \lambda_z(z^c)\psi_{z^c}\right]dt$$

where $\psi_z = W_{z^c} - W_z$, $\psi_{z^c} = W_z - W_{z^c}$. And $\psi_{z^c} = -\psi_z$. The HJB equation for a firm under both z and z^c states are

$$\begin{aligned}(r + \lambda_x + \lambda_z(z))F_{FB}(W_z, z) = \max_{C,\psi_z} \quad & -C + \lambda_z(z)F_{FB}\left(W_z + \psi_z, z^c\right) \\ & + F'_{FB}(W, z)\left[r\left(W_z - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z) - W_z\right) - \lambda_z(z)\psi_z\right]\end{aligned} \tag{26}$$

$$\begin{aligned}(r + \lambda_x + \lambda_z(z^c))F_{FB}(W_{z^c}, z^c) = \max_{C,\psi_{z^c}} \quad & -C + \lambda_z(z^c)F_{FB}(W_{z^c}, z) \\ & + F'_{FB}(W_{z^c}, z^c)\left[r\left(W_{z^c} - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z^c) - W_{z^c}\right) - \lambda_z(z^c)\psi_{z^c}\right]\end{aligned} \tag{27}$$

The first order conditions for (26) are

$$-1 - rF'_{FB}(W, z)u'(C) = 0$$

$$z + rF'_{FB}(W_z, z)\phi'(A) = 0$$

$$-F'_{FB}(W, z)\lambda_z(z) + \lambda_z(z)F'_{FB}(W_z + \psi_z, z^c) = 0$$

The first order conditions for (27) are

$$-1 - rF'_{FB}(W, z_c)u'(C) = 0$$

$$z^c + rF'_{FB}(W_{z^c}, z^c)\phi'(A) = 0$$

$$-F'_{FB}(W, z_c)\lambda_z(z_c) + \lambda_z(z_c)F'_{FB}(W_{z^c} + \psi_{z^c}, z) = 0$$

Use guess and verify method. Guess the policy function C_t and ψ_t are chosen to make the drift of W_z and W_{z^c} equal to zero.

$$dW_z = \left[r\left(W_z - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z) - W_z\right) - \lambda_z(z)\psi_z\right]dt$$

$$dW_{z^c} = \left[r\left(W_{z^c} - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z^c) - W_{z^c}\right) - \lambda_z(z^c)\psi_{z^c}\right]dt$$

The linear equation system is

$$0 = r\left(W_z - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z) - W_z\right) - \lambda_z(z)\psi_z$$

$$0 = r\left(W_{z^c} - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z^c) - W_{z^c}\right) - \lambda_z(z^c)\psi_{z^c}$$

Jointly we can solve for $u(C)-\phi(A)$ and ψ_z. Also we know $\psi_z = W_{z^c} - W_z = -\psi_{z^c}$.

$$u(C)-\phi(A) = \frac{r\left(\lambda_z(z^c)W_z + \lambda_z(z)W_{z^c}\right) - \lambda_x\left[\lambda_z(z^c)(W_u(z)-W_z) + \lambda_z(z)(W_u(z^c)-W_{z^c})\right]}{r\left(\lambda_z(z)+\lambda_z(z^c)\right)} \tag{28}$$

$$\psi_z = \frac{-\lambda_x\left(W_u(z)-W_u(z^c)\right) + (r+\lambda_x)(W_z - W_{z^c})}{\lambda_z(z)+\lambda_z(z^c)}$$

$$\psi_z = -\frac{\lambda_x(W_u(z)-W_u(z^c))}{\lambda_x+\lambda_z(z)+\lambda_z(z^c)+r}$$

Given the drift term of W_z and W_{z^c} being zero, the HJB equations degenerate into

$$(r+\lambda_x+\lambda_z(z))\,F_{FB}\left(W_z, z\right) = \max_{C,\psi_z} \quad zA - C + \lambda_z(z)F_{FB}\left(W_z+\psi_z, z^c\right)$$

$$(r+\lambda_x+\lambda_z(z^c))\,F_{FB}\left(W_{z^c}, z^c\right) = \max_{C,\psi_{z^c}} \quad z^cA - C + \lambda_z(z^c)F_{FB}\left(W_{z^c}+\psi_{z^c}, z\right)$$

Suppose C and ψ are already optimal choices, the HJB equation system delivers

$$F_{FB}\left(W_z, z\right) = \frac{zA-C}{r+\lambda_x}$$

$$F_{FB}\left(W_{z^c}, z^c\right) = \frac{z^cA-C}{r+\lambda_x}$$

If the aggregate state at the retirement point is z, then starting promised utility is W_z. Replace W_{z^c} with $W_z+\psi_z$ in (28) we got,

$$u(C)-\phi(A) = \frac{(\lambda_x+r)W_z}{r} - \frac{\lambda_x\left((\lambda_x+\lambda_z(z^c)+r)\,W_u(z) + \lambda_z(z)W_u(z^c)\right)}{r\left(\lambda_x+\lambda_z(z)+\lambda_z(z^c)+r\right)}, \tag{29}$$

$$F_{FB}\left(W_z, z\right) = \frac{zA-C}{r+\lambda_x}.$$

If the aggregate state at the retirement point is z^c, then starting promised utility is W_{z^c}

$$u(C)-\phi(A) = \frac{(\lambda_x+r)W_{z^c}}{r} - \frac{\lambda_x\left((\lambda_x+\lambda_z(z)+r)\,W_u(z^c) + \lambda_z(z^c)W_u(z)\right)}{r\left(\lambda_x+\lambda_z(z)+\lambda_z(z^c)+r\right)},$$

$$F_{FB}\left(W_{z^c}, z^c\right) = \frac{z^cA-C}{r+\lambda_x}.$$

To verify the solution is optimal, we check whether the optimality conditions are satisfied. Taking derivative w.r.t W on both sides of (29)

$$u'(C)\frac{\partial C}{\partial W} - \phi'(A)\frac{\partial A}{\partial W} = \frac{\lambda_x+r}{r}.$$

Then the derivative of First Best firm profit is

$$F'_{FB}\left(W_z, z\right) = \frac{1}{r + \lambda_x}\left(z\frac{\partial A}{\partial W} - \frac{\partial C}{\partial W}\right) = -\frac{1}{ru'(C)}.$$

Obviously, the allocation satisfies the first order conditions regarding C and A. At the same time $F'_R\left(W_z, z\right) = F'_R\left(W_{z^c}, z^c\right) = -\frac{1}{ru'(C)}$, so the optimality condition regarding ψ_z also satisfies. So we have verified the constant consumption plan is indeed the optimal contract. □

C.3 Constrained optimal contract

Operating firm's problem. The operating firm propose a long term contract that maximize expected profit

$$\max_C \mathbb{E}\left[\int_0^\tau e^{-rt}\left(z_t A dt - C_t dt\right)|z_0\right]$$

subject to deliver intial promised utility W_0 to the worker

$$\mathbb{E}\left[\int_0^\tau e^{-rt}\left(u\left(C_t\right) - \phi\left(A\right)\right)dt + e^{-r\tau}W_u\left(z_\tau\right)\right] = W_0$$

and the contract must be incentive compatible.

Employed worker's problem. There exist exogenous separation shock N_x and aggregate regime switching shock N_z. Then worker's promised utility can be described by a stochastic differential equation

$$dW_t^a = \left[r\left(W_t^a - u\left(C_t\right) + \phi\left(a_t\right)\right)\right]dt + \Psi_{b,t}\sigma_B dB_t + \Psi_{x,t}\left(dN_{x,t} - \lambda_x dt\right) + \Psi_{z,t}\left(dN_{z,t} - \lambda_z\left(z_t\right)dt\right)$$

where $\Psi_{x,t}$ satisfies

$$\Psi_{x,t} = W_u\left(z_t\right) - W_t$$

When neither N_x nor N_z has arrived yet, the continuation value is described by the following stochastic differential equation

$$dW_t = \left[r\left(W_t - u\left(C_t\right) + \phi\left(A_t\right)\right) - \lambda_x\left(W_u\left(z\right) - W_t\right) - \lambda_z\left(z\right)\Psi_{z,t}\right]dt + \Psi_{b,t}\sigma dB_t$$

The incentive-compatible constraint is

$$z_t\Psi_{b,t} = r\phi'\left(A_t\right)$$

Operating firm's recursive problem. The firm's problem can be written into recursive problem with value function depending on two state variables W, z

$$\begin{aligned}
F(W,z) = &\max_{A,C,\Psi_z} z_t A dt - C_t dt \\
&+ e^{-rdt}\left[(1-\lambda_x dt - \lambda_z(z)dt)\left(F(W,z) + F_W(W,z)\,dW + \frac{1}{2}F_{WW}(W,z)\,d\langle W\rangle\right)\right. \\
&\left.+\lambda_z(z)dtF(W+\Psi_z, z^c) + \lambda_x dt V(z)\right]
\end{aligned}$$

The HJB equation is

$$\begin{aligned}
rF(W,z) = &\max_{A,C,\Psi_z} z_t A - C_t + \lambda_z(z)\left(F(W+\Psi_z, z^c) - F(W,z)\right) - \lambda_x F(W,z) \\
&+ F_W(W,z)\left[r\left(W - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z) - W\right) - \lambda_z(z)\Psi_z\right] \\
&+ \frac{1}{2}F_{WW}(W,z)\left(\frac{r\phi'(A)}{z}\right)^2\sigma^2
\end{aligned}$$

and the coupled HJB equation for z^c is

$$\begin{aligned}
rF(W,z^c) = &\max_{A,C,\Psi_{z^c}} z_t^c A - C_t + \lambda_z(z^c)\left(F(W+\Psi_{z^c}, z) - F(W,z^c)\right) - \lambda_x F(W,z^c) \\
&+ F_W(W,z^c)\left[r\left(W - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z^c) - W\right) - \lambda_z(z^c)\Psi_{z^c}\right] \\
&+ \frac{1}{2}F_{WW}(W,z^c)\left(\frac{r\phi'(A)}{z^c}\right)^2\sigma^2
\end{aligned}$$

By collecting items, we can write the HJB equation system into

$$\begin{aligned}
\left(r + \lambda_z(z) + \lambda_x\right)F(W,z) = &\max_{A,C,\Psi_z} z_t A - C_t + \lambda_z(z)F(W+\Psi_z, z^c) \\
&+ F_W(W,z)\left[r\left(W - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z) - W\right) - \lambda_z(z)\Psi_z\right] \\
&+ \frac{1}{2}F_{WW}(W,z)\left(\frac{r\phi'(A)}{z}\right)^2\sigma^2
\end{aligned}$$

and the coupled HJB equation for z^c is

$$\begin{aligned}
\left(r + \lambda_z(z^c) + \lambda_x\right)F(W,z^c) = &\max_{A,C,\Psi_z} z_t^c A - C_t + \lambda_z(z^c)F(W+\Psi_{z^c}, z) \\
&+ F_W(W,z^c)\left[r\left(W - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z^c) - W\right) - \lambda_z(z^c)\Psi_{z^c}\right] \\
&+ \frac{1}{2}F_{WW}(W,z^c)\left(\frac{r\phi'(A)}{z^c}\right)^2\sigma^2
\end{aligned}$$

Optimality conditions. For $z \in \mathcal{Z}$, the following first order conditions hold

$$[A] \quad z + rF_W(W,z)\,\phi'(A) + \frac{\sigma^2 r^2}{z^2} F_{WW}(W,z)\,\phi'(A)\,\phi''(A) = 0$$

$$[C] \quad -1 - rF_W(W,z)\,u'(C) = 0$$

$$[\Psi_z] \quad F_W(W,z) = F_W(W + \Psi_z, z^c)$$

Boundary conditions. For $F(W,z)$, the boundary conditions are

$$F'(W_u(z),z) = \frac{F(W_u(z)) - F_R(\frac{\lambda_x W_u(z)}{r+\lambda_x})}{W_u(z) - \frac{\lambda_x}{r+\lambda_x}W_u(z)}$$

$$F(W_r(z),z) = F_R(W_r(z),z)$$

$$F'(W_r(z),z) = F_R'(W_r(z),z)$$

For $F(W,z^c)$, the boundary conditions are

$$F'(W_u(z^c),z^c) = \frac{F(W_u(z)) - F_R(\frac{\lambda_x W_u(z^c)}{r+\lambda_x})}{W_u(z^c) - \frac{\lambda_x}{r+\lambda_x}W_u(z^c)}$$

$$F(W_r(z^c),z^c) = F_R(W_r(z^c),z^c)$$

$$F'(W_r(z^c),z^c) = F_R'(W_r(z^c),z^c)$$

Notice that $W_r(z)$ is the retirement point and $W_u(z)$ is the outside option.

C.4 Labor market equilibrium

This subsection sets up the problems for firms and workers in the directed search market.

Vacant firm's problem. When the aggregate state is z, observing a menu of submarket $(W_0(z), \theta(z))$, firms simultaneously choose whether or not to create vacancies and where to locate them. The firm's expected benefit of creating a vacancy in submarket $W_0(z)$ is the product of job filling probability $q(\theta(z))$, and the expected value of hiring a worker is $F(W_0(z),z)$. k is the cost of posting vacancies per unit of time. We assume that it's a constant across all aggregate states. In the equilibrium, only a subset of markets $(W_0(z), \theta(z))$ open, for those what do open, the tightness $\theta(z)$ is positive and consistent with the firm free entry equilibrium outcome if and only if

$$k = q(\theta(z))\,F(W_0(z),z) = \xi\,(\theta(z))^{-\alpha}\,F(W_0(z),z) \tag{30}$$

From the firm's free entry condition (30), for the submarket $W_0(z)$ that opens, the corresponding market tightness is

$$\theta\left(W_0(z), z\right) = \left(\frac{\xi F\left(W_0(z), z\right)}{k}\right)^{\frac{1}{\alpha}}$$

It is obvious that the market tightness $\theta\left(W_0(z)\right)$ moves in the same direction as the firm's payoff $F\left(W_0\right)$ moves with the promised utility W_0.

Unemployed worker's problem. For an unemployed worker, b is the unemployment benefit flow which we also assume to be constant across all aggregate states. At aggregate state z, any unemployed worker will take $(W_0(z), \theta\left(z\right))$ menu as given. His search decision is to choose which submarket $W_0(z)$ to visit. If he visits submarket $W_0(z)$, his job finding probability is $p\left(\theta\right)$. If he succeeds, he enters in a new employment relationship which gives him the lifetime utility $W_0(z)$. Otherwise he stays unemployed. The expected utility for an individual unemployed worker is $W_{u,t}(z)$.

$$\begin{aligned} W_{u,t}(z) = & \max_{X(z), X(z^c)} ru(b)dt \\ & + e^{-rdt}[(1 - \lambda_z(z)dt)\left[\left(1 - p\left(\theta\left(W_0(z), z\right)\right)dt\right) W_{u,t+dt}(z) + p\left(\theta\left(W_0(z), z\right)\right)dtW_0(z)\right] \\ & + \lambda_z(z)dt\left[\left(1 - p\left(\theta\left(W_0(z^c), z^c\right)\right)dt\right) W_{u,t+dt}(z^c) + p\left(\theta\left(W_0(z^c), z^c\right)\right)dtW_0(z^c)\right]] \end{aligned} \tag{31}$$

where $p(\theta)$ is endogenized by

$$p(\theta(W_0(z), z)) = \xi\left(\frac{\xi F\left(W_0(z), z\right)}{k}\right)^{\frac{1-\alpha}{\alpha}}$$

Denote the policy function that solves optimization problem (31) to be $W_0^*(z)$. Note here we refine the equilibrium to be block recursive equilibrium, the worker does not need to care about the unemployment rate, or any other distributional variables in this economy.

$$(r + \lambda_z)W_u(z) = ru(b) + \lambda_z W_u(z^c) + p\left(\Theta\left(W_0(z)\right)\right)\left(W_0(z) - W_u(z)\right)$$

$$(r + \lambda_{z^c})W_u(z^c) = ru(b) + \lambda_{z^c} W_u(z) + p\left(\Theta\left(W_0(z^c)\right)\right)\left(W_0(z^c) - W_u(z^c)\right)$$

Distribution of workers. The evolution of the distribution of workers' promised utilities can be summarized by the Kolmogorov forward equation (KFE). Let $G(W, z)$ denote the density function of an employed worker's promised utility W and aggregate state z. Let $u(z)$ denote the measure of unemployed workers. For simplicity, we suppress the argument t in the density function. The KFE equation follows

$$\begin{aligned}\frac{\partial G(W,z)}{\partial t} &= -\frac{\partial}{\partial W}\left[\left[r\left(W_t - u\left(C_t\right) + \phi\left(A_t\right)\right) - \lambda_x\left(W_u(z) - W_t\right) - \lambda_z(z)\Psi_{z,t}\right]G(W,z)\right] \\ &+ \frac{1}{2}\frac{\partial^2}{\partial W^2}\left[\left(\frac{r\phi'(A)}{z}\right)^2\sigma^2 G(W,z)\right] \\ &- \lambda_x G(W,z) + \left[\left(1-\varphi'(W,z)\right)G\left(W-\varphi(W,z),z\right) - G(W,z)\right]\lambda_z(z)\end{aligned}$$

where $\varphi(W,z)$ is defined by the equivalence,

$$w + \psi(w) = W \implies w = W - \varphi(W)$$

The law of motion for the measure of unemployed workers is

$$du(z) = \lambda_x dt\,(1-u(z)) - \left(p(\theta(z))dt\right)u(z).$$

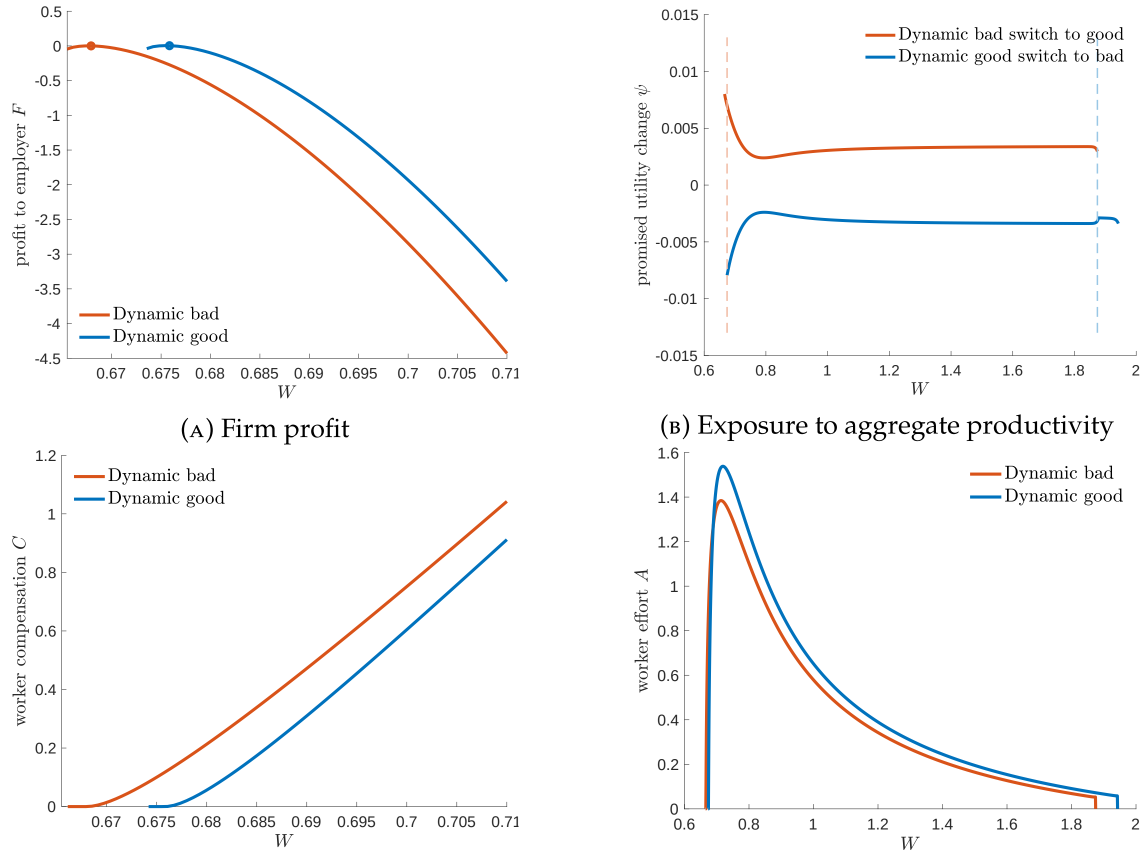
C.5 Dynamic result

The equilibrium result of contracting problem is summarized by Figure 14. The red curves correspond to optimal contract result in lower aggregate productivity level and the blue curves correspond to that in higher aggregate productivity level. Figure 14a shows that firms earn more profit when aggregate productivity improves. The scatters denote the submarkets that unemployed workers choose to visit and workers will collectively choose a market that promises higher utility at the higher aggregate state. Figure 14b plots the exposure of worker's promised utility to the aggregate productivity shock. When the aggregate state improves, the worker's promised utility will increase. Figure 14c shows that worker's consumption plan is also improved when the good productivity shock arrives. Figure 14d confirms that worker's effort is procyclical– the substitution effect dominates the income effect.

D. Computation algorithm

D.1 Optimal contract in steady state

In order to compute the optimal contract in the steady state, we use iterate-and-update method. We first guess the outside option W_u, the slope of profit function at the lower bound F_p, the promised utility at the retirement point W_r simultaneously. The individual worker takes W_u, F_p, W_r as given and the optimal contract $F(W)$ can be obtained by solving the ODE using implicit method as shown below. Given the contracting state



(A) Firm profit

(B) Exposure to aggregate productivity

(C) Compensation

(D) Recommend effort

FIGURE 14: Dynamic model optimal contract

space $W \in [W_u, W_r]$ and boundary value conditions $F'(W_u) = F_p$, $F(W_r) = F_R(W_r)$, we can numerically solve the HJB equation

$$(r + \lambda_x) F(W) = \max_{A,C} r(zA - C) + F'(W) \left[r \left(W - u(C) + \phi(A) \right) - \lambda_x (W_u - W) \right] + \frac{1}{2} F''(W) \left(\frac{r\phi'(A)\sigma_B}{z} \right)^2$$

We use the retirement profit function and its policy functions $A = 0$, $C = u^{-1}(W - \frac{\lambda_x}{r}(W_u - W))$ as the initial guess. Denote

$$\rho = r + \lambda_x, \quad G = r(zA - C), \quad MU = r(W - u(C) + \phi(A)) - (W_u - W)\lambda_x, \quad S = \frac{r\phi'(A)\sigma_B}{z}$$

HJB equation can be approximated by the finite difference equation, noted that we adopt the upwind scheme to work with the forward/backward approximation.

$$\rho F_i = G_i + \frac{F_{i+1} - F_i}{\Delta_i} MU_i^+ - \frac{F_i - F_{i-1}}{\Delta_i} MU_i^- + \frac{F_{i+1} - F_i - F_i + F_{i-1}}{\Delta_i^2} \frac{S_i^2}{2}$$

Where "+" on subscript is to take positive values and "-" on subscript is to take negative absolute values. Collect terms we get,

$$\rho F_i = G_i + \left(\frac{MU_i^-}{\Delta_i} + \frac{S_i^2}{2\Delta_i^2} \right) F_{i-1} + \left(-\frac{MU_i^+}{\Delta_i} - \frac{MU_i^-}{\Delta_i} - \frac{S_i^2}{\Delta_i^2} \right) F_i + \left(\frac{MU_i^+}{\Delta_i} + \frac{S_i^2}{2\Delta_i^2} \right) F_{i+1}$$

Denote

$$x_i = \frac{MU_i^-}{\Delta_i} + \frac{S_i^2}{2\Delta_i^2}, \quad y_i = -\frac{MU_i^+}{\Delta_i} - \frac{MU_i^-}{\Delta_i} - \frac{S_i^2}{\Delta_i^2}, \quad z_i = \frac{MU_i^+}{\Delta_i} + \frac{S_i^2}{2\Delta_i^2}$$

Denote matrix $\boldsymbol{A}$ as the intensity matrix or transition matrix,

$$\boldsymbol{A} \equiv \begin{bmatrix} y_1 & z_1 & 0 & \dots & 0 & 0 & 0 \\ x_2 & y_2 & z_2 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & x_{N-1} & y_{N-1} & z_{N-1} \\ 0 & 0 & 0 & \dots & 0 & x_N & y_N \end{bmatrix}$$

Then the HJB equation can be written into

$$\rho \boldsymbol{F} = \boldsymbol{G} + \boldsymbol{A}\boldsymbol{F}$$

where

$$\boldsymbol{F} = \begin{bmatrix} F_1 \\ \vdots \\ F_N \end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix} G_1 \\ \vdots \\ G_N \end{bmatrix}.$$

After rearrange the terms,

$$(\rho \boldsymbol{I} - \boldsymbol{A})\boldsymbol{F} = \boldsymbol{G}.$$

Denote $\boldsymbol{B} = \rho \boldsymbol{I} - \boldsymbol{A}$, except with the "dirty fix" on 1-st and N-th row to incorporate the boundary conditions. Accordingly we need to fix vector $\boldsymbol{G}$ on both ends to incorporate the boundary conditions. where $G_1 = F_p, G_N = F_R(W_r)$. The boundary conditions are embedded in the first and last row of equation.

$$
\boldsymbol{B} \equiv \begin{bmatrix} -\frac{1}{\Delta_1} & \frac{1}{\Delta_1} & 0 & \dots & 0 & 0 & 0 \\ -x_2 & \rho - y_2 & -z_2 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -x_{N-1} & \rho - y_{N-1} & -z_{N-1} \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}.
$$

The HJB equation becomes

$$
\boldsymbol{B} \cdot \boldsymbol{F} = \boldsymbol{G}.
$$

Then we can solve for $\boldsymbol{F}$,

$$
\boldsymbol{F} = \boldsymbol{B}^{-1} \boldsymbol{G}.
$$

The updated $F(W)$ can be used to solve the updated policy functions A and C from the optimality conditions

$$
z_t + rF'(W_t)\,\phi'(A_t) + r^2 F''(W_t)\left(\frac{\sigma}{z}\right)^2 \phi'(A_t)\,\phi''(A_t) = 0
$$

$$
-1 - rF'(W_t)\,u'(C_t) = 0
$$

From this optimality condition, the updated C, A and ψ as functions of W are known. Then we can use these policy functions as the new guess and perform the iteration until $F(W)$ converges.

Next, we update the value of W_u, F_p, W_r simultaneously. We can also update W_u through HJB equation (11),

$$
W_u = \max_{W_0 \in \mathcal{W}} \frac{1}{r} \left[ru(b) + p\left(\Theta\left(W_0\right)\right)\left(W_0 - W_u\right) \right]
$$

To maximize the expected search value, individual worker determines which is the optimal submarket to visit, denote as W_0^*. For both computation efficiency and accuracy concern, we adopt the Golden Search method in finding the maximized searching value of the worker.

We can update the value of F_p through the smooth pasting condition (8),

$$
F_p = \frac{F(W_u) - F_R(\frac{\lambda_x W_u}{r+\lambda_x})}{W_u - \frac{\lambda_x}{r+\lambda_x} W_u}
$$

We can update the value of W_r by the smooth pasting condition at the upper bound (7),

$$
F'(W_r) = F_R'(W_r) = \left[-\frac{ru^{-1}\left[W - \frac{\lambda_x}{r}\left(W_u - W\right)\right]}{r + \lambda_x} \right]' = -\frac{1}{u'\left[W_r - \frac{\lambda_x}{r}\left(W_u - W_r\right)\right]}
$$

We can implicitly solve and update W_r.

Then we can perform the iterate-and-update strategy until W_u, F_p, W_r all respectively converge.

D.2 Worker's distribution in steady state

In steady state, unemployment worker measure is given by

$$u = \frac{\lambda_x}{\lambda_x + p\left(\Theta(X)\right)}$$

We use $g(W, t)$ to denote the measure of workers over their promised utilities.

$$\int_{W_u}^{W_r} g(W, t) dW = 1 - u$$

The Fokker-Planck equation or KFE describes the distribution transition for all workers,

$$\begin{aligned}\frac{\partial g(W,t)}{\partial t} &= -\frac{\partial}{\partial W}\left[\left(r(W_t - u(C_t) + \phi(A_t)) - (W_u - W_t)\lambda_x\right) g(W,t)\right] \\ &+ \frac{1}{2}\frac{\partial^2}{\partial W^2}\left[\left(r\frac{\phi'(A_t)}{z_t}\sigma_B\right)^2 g(W,t)\right] - \lambda_x g(W,t)\end{aligned}$$

Except for workers with promised utility W_0, the distribution transition has an additional injection of new workers, denoted as $I(W_0, t)$,

$$\begin{aligned}\frac{\partial g(W,t)}{\partial t} &= -\frac{\partial}{\partial W}\left[\left(r(W_t - u(C_t) + \phi(A_t)) - (W_u - W_t)\lambda_x\right) g(W,t)\right] \\ &+ \frac{1}{2}\frac{\partial^2}{\partial W^2}\left[\left(r\frac{\phi'(A_t)}{z_t}\sigma_B\right)^2 g(W,t)\right] - \lambda_x g(W,t) + I(W_0,t)\end{aligned}$$

where $I(W_0, t) = pdt\ u$. And the boundary conditions are: (a) a reflecting lower boundary at $W = W_u$ and (b) an absorbing upper boundary at $W = W_r$:

$$-\left[\left(r(W_u - u(C_t) + \phi(A_t)))\right) g(W_u,t)\right] + \frac{1}{2}\frac{\partial}{\partial W}\left[\left(r\frac{\phi'(A_t)}{z_t}\sigma_B\right)^2 g(W_u,t)\right] = 0$$

$$g(W_r, t) = \int_0^t J(W_r, t) dt$$

The stopping/exit rate through upper boundary $W = W_r$ per unit of time is $J(W_r, t)$

$$J(W_r, t) = -\left[\left(r(W_r - u(C_t) + \phi(A_t)) - (W_u - W_r)\lambda_x\right) g(W_r, t)\right] + \frac{1}{2}\frac{\partial}{\partial W}\left[\left(r\frac{\phi'(A_t)}{z_t}\sigma_B\right)^2 g(W_r, t)\right]$$

Now we perform finite difference to compute $g(W, t)$. Notation is the same as in the last subsection.

$$MU = r(W - u(C) + \phi(A)) - (W_u - W)\lambda_x, \quad S = \frac{r\phi'(A)\sigma_B}{z}$$

KFE can then be written into

$$\frac{\partial g(W, t)}{\partial t} = -\frac{\partial(MU(W)g(W))}{\partial W} + \frac{1}{2}\frac{\partial^2(S^2(W)g(W))}{\partial W^2} - \lambda_x g(W)$$

The KFE for stationary probability density function can be approximated by finite difference equation. Again we use upwind scheme and implicit method,

$$\frac{g_i - g_i^0}{\Delta t} = -\left[\frac{g_i MU_i^+ - g_{i-1}MU_{i-1}^+}{\Delta_i} + \frac{g_i MU_i^- - g_{i+1}MU_{i+1}^-}{\Delta_i}\right] + \frac{g_{i+1}S_{i+1}^2 - g_i S_i^2 - g_i S_i^2 + g_{i-1}S_{i-1}^2}{2\Delta_i^2} - \lambda_x g_i$$

Collect terms we will get

$$\frac{g_i - g_i^0}{\Delta t} = \left(\frac{MU_{i-1}^+}{\Delta_i} + \frac{S_{i-1}^2}{2\Delta_i^2}\right) g_{i-1} + \left(-\frac{MU_i^+}{\Delta_i} - \frac{MU_i^-}{\Delta_i} - \frac{S_i^2}{\Delta_i^2}\right) g_i + \left(\frac{MU_{i+1}^-}{\Delta_i} + \frac{S_{i+1}^2}{2\Delta_i^2}\right) g_{i+1} - \lambda_x g_i$$

Adopt the same notations as in the last subsection,

$$x_i = \frac{MU_i^-}{\Delta_i} + \frac{S_i^2}{2\Delta_i^2}, \quad y_i = -\frac{MU_i^+}{\Delta_i} - \frac{MU_i^-}{\Delta_i} - \frac{S_i^2}{\Delta_i^2}, \quad z_i = \frac{MU_i^+}{\Delta_i} + \frac{S_i^2}{2\Delta_i^2}$$

Then the finite difference KFE becomes

$$\frac{g_i - g_i^0}{\Delta t} = z_{i-1}g_{i-1} + y_i g_i + x_{i+1}g_{i+1} - \lambda_x g_i \tag{32}$$

Notice that the adjoint matrix for transition matrix $\boldsymbol{A}$ from last subsection is

$$\boldsymbol{A}^T = \begin{bmatrix} y_1 & x_2 & 0 & \dots & 0 & 0 & 0 \\ z_1 & y_2 & x_3 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & z_{N-2} & y_{N-1} & x_N \\ 0 & 0 & 0 & \dots & 0 & z_{N-1} & y_N \end{bmatrix}$$

which coincides with the coefficient on the right hand side of (32). And the finite difference of KFE can be vectorized into the following equation

$$\boldsymbol{g} - \boldsymbol{g}^0 = \boldsymbol{A}^T \boldsymbol{g} \Delta t - \lambda_x \boldsymbol{g} \Delta t$$

We can rearrange terms,

$$\left[(1 + \lambda_x \Delta t) \, \boldsymbol{I} - \Delta t \, \boldsymbol{A}^T \right] \boldsymbol{g} = \boldsymbol{g}^0$$

With the exception that at grid W_0, the KFE is

$$\boldsymbol{g} - \boldsymbol{g}^0 = \boldsymbol{A}^T \Delta t - \lambda_x \boldsymbol{g} \Delta t + pdt \ u$$

After collecting terms,

$$\left[(1 + \lambda_x \Delta t) \, \boldsymbol{I} - \Delta t \, \boldsymbol{A}^T \right] \boldsymbol{g} = \boldsymbol{g}^0 + pdt \ u.$$

Denote $\boldsymbol{D} = (1 + \lambda_x \Delta t) \, \boldsymbol{I} - \Delta t \, \boldsymbol{A}^T$, except with the fix on 1-st and N-th row to incorporate the boundary conditions.

For the reflective lower bound, the finite difference gives us

$$\frac{\partial g_1}{\partial t} = \left(\frac{\mu_2^-}{\Delta} + \frac{S_2^2}{2\Delta^2} \right) g_2 - \left(\frac{\mu_1^+}{\Delta} + \frac{S_1^2}{2\Delta^2} \right) g_1 - \lambda_x g_1 = 0$$

Accordingly, we modify the first row of intensity matrix $\boldsymbol{D}$,

$$x_1 = 0, \quad y_1 = \frac{\mu_1^-}{\Delta} + \frac{S_1^2}{2\Delta^2} - \frac{\mu_1^+}{\Delta} - \frac{S_1^2}{\Delta^2} = -\frac{\mu_1^+}{\Delta} - \frac{S_1^2}{2\Delta^2}, \quad x_2 = \frac{\mu_2^-}{\Delta} + \frac{S_2^2}{2\Delta^2}$$

For the absorbing upper bound, we follow the practice introduced by (Gabaix, 2009).

$$\frac{\left(g_{N-1} - g_{N-1}^0\right)}{\Delta t} = g_{N-2} \left(\frac{\mu_{N-2}^+}{\Delta} + \frac{1}{2} \frac{S_{N-2}^2}{\Delta^2} \right) + g_{N-1} \left(-\frac{\mu_{N-1}^+}{\Delta} - \frac{\mu_{N-1}^-}{\Delta} - \frac{S_{N-1}^2}{\Delta^2} \right) - \lambda_x g_{N-1}$$

$$\frac{\left(g_N - g_N^0\right)}{\Delta t} = g_{N-1}\left(\frac{\mu_{N-1}^+}{\Delta} + \frac{1}{2}\frac{S_{N-1}^2}{\Delta^2}\right) - \lambda_x g_N$$

We modify the last and second to last row of intensity matrix $\boldsymbol{D}$,

$$z_{N-1} = \frac{\mu_{N-1}^+}{\Delta} + \frac{1}{2}\frac{S_{N-1}^2}{\Delta^2}, \quad z_{N-2} = \frac{\mu_{N-2}^+}{\Delta} + \frac{1}{2}\frac{S_{N-2}^2}{\Delta^2}, \quad y_{N-1} = -\frac{\mu_{N-1}^+}{\Delta} - \frac{\mu_{N-1}^-}{\Delta} - \frac{S_{N-1}^2}{\Delta^2}, \quad x_N = 0$$

The finite difference scheme can be vectorized into the following equation,

$$\boldsymbol{D} \cdot \boldsymbol{g} = \boldsymbol{g}^0$$

where $\boldsymbol{g}^0$ is the old density distribution, except that at the lower bound we modify to incorporate the boundary condition,

$$g_1^0 = 0$$

Besides, at $W = W_0$, $g_{W_0}^0 = g_{W_0}^0 + pdt\ u$. Then we can solve for $\boldsymbol{g}$.

$$\boldsymbol{g} = \boldsymbol{D}^{-1}\boldsymbol{g}^0 \tag{33}$$

Obviously, the stationary distribution function can be numerically solved running forward by (33), till $\boldsymbol{g}$ converges.

D.3 Optimal contract in dynamic setting

We first guess the unemployment value $W_{u,G}, W_{u,B}$, retirement continuation value $W_{r,G}, W_{r,B}$ and $F_{p,G}, F_{p,B}$. Where $F_{p,G}, F_{p,B}$ are the first order derivatives of firm profit function at the lower bound W_u. We also guess the value functions under two aggregate states to be $F(W_z, z)$ and $F(W_{z^c}, z^c)$. Given the contracting state space $W_z \in [W_{u,B}, W_{r,B}]$ and boundary value conditions, we can numerically solve the HJB equation for aggregate state $z = z_B$,

$$\begin{aligned}
(r + \lambda_x + \lambda_z(z))\, F(W, z) = &\max_{A, C, \Psi_z} z_t A - C_t + \lambda_z(z) F(W + \Psi_z, z^c) \\
&+ F_W(W, z)\left[r\left(W - u(C) + \phi(A)\right) - \lambda_x\left(W_u(z) - W\right) - \lambda_z(z)\Psi_z\right] \\
&+ \frac{1}{2} F_{WW}(W, z)\left(\frac{\phi'(A)}{z}\right)^2 \sigma^2
\end{aligned}$$

Given the contracting state space $W \in [W_{u,G}, W_{r,G}]$ and boundary value conditions, we can numerically solve the coupled HJB equation for $z^c = z_G$,

$$\begin{aligned}(r + \lambda_x + \lambda_z(z^c))\, F\,(W, z^c) = \max_{A,C,\Psi_{z^c}} &\; z_t^c A - C_t + \lambda_z(z^c) F\,(W + \Psi_{z^c}, z) \\ &+ F_W\,(W, z^c)\left[r\left(W - u\,(C) + \phi\,(A)\right) - \lambda_x\,(W_u\,(z^c) - W) - \lambda_z\,(z^c)\,\Psi_{z^c}\right] \\ &+ \frac{1}{2} F_{WW}\,(W, z^c)\left(\frac{\phi'\,(A)}{z^c}\right)^2 \sigma^2\end{aligned}$$

We will demonstrate the algorithm of solving the HJB equation by looking at HJB equation associated with state z, and the HJB equation associated with state z^c can be solved likewise. Denote

$$\rho = r + \lambda_x + \lambda_z(z), \quad G = zA - C + \lambda_z(z) F\,(W + \Psi_z, z^c)$$

$$MU = r\left(W - u\,(C) + \phi\,(A)\right) - (W_u(z) - W)\,\lambda_x - \lambda_z\,(z)\,\Psi_z, \quad S = \frac{\phi'\,(A)\,\sigma}{z}$$

HJB equation for $F(W, z)$ can be approximated by finite difference equation

$$\rho F_i = g_i + \frac{F_{i+1} - F_i}{\Delta_i} MU_i^+ - \frac{F_i - F_{i-1}}{\Delta_i} MU_i^- + \frac{F_{i+1} - F_i - F_i + F_{i-1}}{\Delta_i^2}\frac{S_i^2}{2}$$

We can solve $F(W_z, z)$ from this finite difference equation using exactly the same procedure we introduced in the steady state model. Similarly we can update $F(W_{z^c}, z^c)$. The updated $F(W_z, z)$ and $F(W_{z^c}, z^c)$ can be used to solve the updated policy functions A, C and ψ_z from the optimality conditions

$$z_t + rF'\,(W_z, z)\,\phi'\,(A_t) + r^2 F''\,(W, z)\left(\frac{\sigma}{z}\right)^2 \phi'\,(A_t)\,\phi''\,(A_t) = 0$$

$$-1 - rF'\,(W, z)\,u'\,(C_t) = 0$$

$$F'(W_z, z) = F'(W_z + \psi_z, z^c)$$

Now we update the value of $W_{u,z}, W_{u,z^c}, F_{p,z}, F_{p,z^c}, W_{r,z}, W_{r,z^c}$ simultaneously. The procedure is similar to that in the Steady State computation. We do this iteratively until everything converges.

D.4 Dynamic model worker distribution

Denote $g(W_z, z, t)$ as the measure of worker distribution over their promised utilities at aggregate state z, and $g(W_{z^c}, z^c, t)$ the measure of distribution at aggregate state z^c.

$$\int_{W_u(z)}^{W_r(z)} g(W_z, z, t) dW = 1 - u(z)$$

The KFE that describes the distribution transition for all W at aggregate state z is

$$\begin{aligned}\frac{\partial g(W_z,z,t)}{\partial t} &= -\frac{\partial}{\partial W_z}\left[\left(r(W_z-u(C_t)+\phi(A_t))-(W_{u,z}-W_z)\lambda_x-\psi_{z,t}\lambda_z(z)\right)\right)g(W_z,z,t)\Big]\\ &+\frac{1}{2}\frac{\partial^2}{\partial W_z^2}\left[\left(r\frac{\phi'(A_t)}{z_t}\sigma_B\right)^2 g(W_z,z,t)\right]-\lambda_x g(W_z,z,t)\end{aligned}$$

At $W_0(z)$, the distribution transition has an additional injection of new workers, denoted as $I(W_0(z),t)$,

$$\begin{aligned}\frac{\partial g(W_z,z,t)}{\partial t} &= -\frac{\partial}{\partial W_z}\left[\left(r(W_z-u(C_t)+\phi(A_t))-(W_{u,z}-W_z)\lambda_x-\psi_{z,t}\lambda_z(z)\right)g(W_z,z,t)\right]\\ &+\frac{1}{2}\frac{\partial^2}{\partial W_z^2}\left[\left(r\frac{\phi'(A_t)}{z_t}\sigma\right)^2 g(W_z,z,t)\right]-\lambda_x g(W_z,z,t)+I(W_0(z),z,t)\end{aligned}$$

The boundary conditions are: reflecting lower boundary at $W_z=W_{u,z}$,

$$-\left[\left(r(W_{u,z}-u(C_t)+\phi(A_t)))\right)g(W_{u,z},z,t)\right]+\frac{1}{2}\frac{\partial}{\partial W_z}\left[\left(r\frac{\phi'(A_t)}{z_t}\sigma_B\right)^2 g(W_{u,z},z,t)\right]=0$$

and absorbing upper boundary at $W_z=W_{r,z}$.

$$g(W_{r,z},z,t)=\int_0^t J(W_{r,z},z,t)dt$$

The stopping/exit rate through upper boundary $W_z=W_{r,z}$ per unit of time is $J(W_{r,z},z,t)$.

$$\begin{aligned}J(W_{r,z},z,t) &= -\left[\left(r(W_{r,z}-u(C_t)+\phi(A_t))-(W_{u,z}-W_{r,z})\lambda_x\right)g(W_{r,z},z,t)\right]\\ &+\frac{1}{2}\frac{\partial}{\partial W_z}\left[\left(r\frac{\phi'(A_t)}{z_t}\sigma\right)^2 g(W_{r,z},z,t)\right]\end{aligned}$$

If the aggregate productivity shock arrives, the aggregate state will switch from z to z^c. The measure of distribution under the new aggregate state instantly change to $g(W_{z^c},z^c,t)$, which is transformed from the former distribution measure $g(W_z,z,t)$.

$$g(W_{z^c},z^c,t)=g(W_z+\psi_z,z^c,t)=g(W_z,z,t)$$

The rest finite difference procedure resembles that in the steady state model computation, we will omit the details for the sake of simplicity.

E. Calibration and estimation

E.1 Stochastic aggregate productivity

The labor productivity data is from the U.S. Bureau of Labor Statistics. The labor productivity data are measured as the percentage change from the previous quarter at an annual rate, and the quarterly percentage change is approximately the annual percentage change divided by four. The sample period is from 1994 Q1 to 2019 Q4. To make the model comparable with the data, the process we will be looking at is the percentage deviation from labor productivity in steady state, which we normalize to be 1. To simplify the notation, rewrite $\lambda_z(z_H) = \lambda_{zH}$ and $\lambda_z(z_L) = \lambda_{zL}$. Define $\Delta z_i = \frac{z_i - \mathbb{E}[z]}{\mathbb{E}[z]}, i \in \{L, H\}$. We assume the percentage deviation to be symmetric, $\Delta z_H = -\Delta z_L$. We then match the moments of process $\{\Delta z\}$ with the data. The moments include: the variance $\texttt{var}[\Delta z]$, the autocorrelation $\mathbb{E}[\Delta z_t \Delta z_{t+1}]$ and the stationary distribution $[\pi_{\Delta zL}, \pi_{\Delta zH}]$. Given the productivity shock process is a Markov process, the model implied $\{\Delta z\}$ moments are

$$\mathbb{E}[\Delta z] = \frac{\Delta z_L \lambda_{zH} + \Delta z_H \lambda_{zL}}{\lambda_{zL} + \lambda_{zH}}, \quad \texttt{std}[\Delta z] = \sqrt{\frac{\lambda_{zL}\lambda_{zH}(\Delta z_L - \Delta z_H)^2}{(\lambda_L + \lambda_H)^2}}$$

$$\texttt{corr}[\Delta z_t \Delta z_{t+1}] = \frac{\mathbb{E}[\Delta z_t \Delta z_{t+1}] - \mathbb{E}[\Delta z]^2}{var[\Delta z]} = 1 - \lambda_{zL} - \lambda_{zH}$$

$$\pi_{\Delta zH} = \frac{\lambda_{zL}}{\lambda_{zL} + \lambda_{zH}}, \quad \pi_{\Delta zL} = \frac{\lambda_{zH}}{\lambda_{zL} + \lambda_{zH}}$$

The stationary distribution is calibrated by looking at the monthly data of NBER based Recession Indicators for the United States. The probability $\pi_{\Delta zL}$ is calibrated by calculating the proportion of the recession periods in the whole sample period. Since we assume $\Delta z_H = -\Delta z_L = \Delta z$, collectively, $\Delta z, \lambda_{zL}, \lambda_{zH}$ can be calibrated by the following equation system

$$\lambda_{zL} = \frac{1 - \pi_L}{\pi_L}\lambda_{zH}, \quad \Delta z = \sqrt{\frac{[\texttt{std}(\Delta z)]^2}{4\pi_L \pi_H}}, \quad 1 - \lambda_{zL} - \lambda_{zH} = \texttt{corr}[\Delta z_t \Delta z_{t+1}]$$

The labor productivity time series suggests that $\pi_{\Delta zB} = 0.141$. Following Shimer (2005), we estimate $\texttt{std}[\Delta z] = 0.02$, $\mathbb{E}[\Delta z_t \Delta z_{t+1}] = 0.878$. Then we can get the value of four parameters $\Delta z_L = -2.87\%$, $\Delta z_H = 2.87\%$, $\lambda_{z_L} = 0.105$, $\lambda_{zH} = 0.017$.

E.2 Alternative calibration strategy

For the robustness check, we also calibrate the model to match the PSID regular wage residual. The identification assumption is that firms do not possess more information than econometricians, and the only unobservable factor that causes the variance in the wage residual is the variance of the idiosyncratic factor. Using this strictest calibration strategy, the standard deviation of real wage residual is 0.242.

To recover the regular wage residual, we regress the hourly wage on worker-job observables. The regression results are in Table 5.

Dependent	asinh(wage)					
Method	OLS	OLS	FE	FE	FE	FE
Regressors	(7)	(8)	(9)	(10)	(11)	(12)
i.worker			✓	✓		
i.match					✓	✓
i.year	✓		✓		✓	
i.industry	✓		✓		✓	
i.occupation	✓		✓		✓	
i.y# i.ind# i.occ		✓		✓		✓
Observations	60411	60300	56825	56710	40088	39910
Std of residual	.525	.511	.321	.315	.247	**.242**

TABLE 5: Regression table of regular wage residual

We use the regression results in Column (12) to gauge our alternative calibration, and the inferred σ is 1.1. The parameterization will be adjusted accordingly and are shown in Table 6.

Parameter	Description	First Best	Alternative	Target
k	vacancy posting cost	0.203	0.083	market tightness $\theta = 1$
χ_1	effort disutility	0.183	0.166	aggregate output equals 1
σ	idiosyn. volatility	0	1.1	micro evidence

TABLE 6: Parameters used for alternative calibration

The simulation results are summarized in the Table 7. We can see that even with this most conservative calibration strategy, the results still hold qualitatively. The unemployment volatility is higher in the economy with moral hazard. The wage dispersion is counter-cyclical when we consider the moral hazard problem.

Unemployment volatility $\texttt{std}(u)$		**Wage dispersion** $\texttt{std}\left[\texttt{asinh}\left(\frac{C}{A}\right)\right]$	
Data	0.190	$\texttt{corr(GDP, wage disp.)}$	-0.1550
No moral hazard	0.016	$\texttt{wage disp.}\vert z_L$	0.398
Alternative calibration	0.0181	$\texttt{wage disp.}\vert z_H$	0.379

Table 7: Unemployment volatility and wage dispersion under alternative calibration