Monopsony with Dynamic Wage Contracts

Ishaana Talesara Arthur Wickard

March 20, 2026

Click for latest version

Abstract

Monopsony power is often measured by interpreting firm wage and labor responses to shocks through static models. But when workers face frictions to changing jobs, employment adjusts gradually and workers respond to changes in the total value of a job—not just the current wage. We develop a general equilibrium dynamic monopsony model where firms contract with risk-averse workers over idiosyncratic shocks. We show that the shock-identified labor supply elasticity depends on the persistence of the shock, worker risk aversion, and the horizon over which it is estimated. These forces induce a wedge between the inverse shock-identified labor supply elasticity and the wage markdown. We estimate the model using U.S. Census employer-employee matched data. The small and persistent wage response to temporary shocks is consistent with firms insuring risk-averse workers. Search frictions explain why employment continues to rise even after wages have started to fall. We find the average worker's wage is marked down 8.3%. By contrast, the static model approach of inverting the shock-identified labor supply elasticity implies a markdown estimate as wide as 26%. Lastly, we show that firm employment dynamics are not efficient: insurance distorts the job ladder, preventing productivity-improving job transitions from occurring.

We are incredibly grateful to Daron Acemoglu, Iván Werning, and Martin Beraja for their ongoing support and guidance. We also thank Patrick Adams, David Autor, Marc de la Barrera, Ricardo Caballero, Alex Carrasco, Nathan Lazarus, Dean Li, Jack Liang, Pedro Martínez-Bruera, Giuseppe Moscarini, Nina Roussille, Bradley Setzler, Christian Wolf, Henry Zhang, Whitney Zhang, and many participants at the MIT Macroeconomics, Labor Economics, and Applied Microeconomics talks for their comments. We also thank Shital Sharma for his help with the Census data and disclosure process. Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 3060. (CBDRB-FY25-P3060-R12493). This research uses data from the Census Bureau's Longitudinal Employer-Household Dynamics Program, which was partially supported by the following National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1745302. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Ishaana Talesara (Department of Economics, Massachusetts Institute of Technology; talesara@mit.edu); Arthur Wickard (Department of Economics, Massachusetts Institute of Technology; wickard@mit.edu).

1 Introduction

What determines how firms adjust to idiosyncratic shocks and what does this response tell us about the extent of labor market power? Static monopsony models suggest that wage markdowns can be estimated by inverting the labor supply elasticity, which has led a large literature to do exactly this.[0] This approach leads to implausibly large wage markdowns—if firms exert such a large degree of wage-setting power then why are they not more profitable?[1]

In this paper, we show that, when considering realistic dynamics, inverting the shock-identified labor supply elasticity does not recover the markdown. First, when workers face frictions to switching jobs, employment adjusts gradually and workers respond to changes in the total value of a job—not just the current wage. Second, the same initial wage increase may imply very different total earnings gains depending on whether the firm's shock was temporary or persistent. Third, risk-averse workers will prefer smoother wage changes, which may influence how firms respond to shocks (Knight, 1921). In the presence of search frictions, risk-aversion, and temporary shocks, the shock-identified labor supply elasticity is a complex, dynamic object which does not neatly map onto the wage markdown.

To motivate the need to model dynamics of frictional labor adjustment, we use U.S. Census matched employer-employee data and estimate the response of wages and employment to firm-level shocks to value added per worker. Labor adjusts slowly, continuing to rise even as the wage response fades. This means that the shock-identified labor supply elasticity is more elastic when estimated over longer horizons. We also find a relatively small and persistent wage response to a temporary shock to firm performance—consistent with insurance motives smoothing the firm's wage-setting behavior. To provide more direct evidence of insurance, we exploit variation in firms' ability to smooth wages due to financial constraints (Ellul et al., 2018). We show that firms facing tighter financial constraints exhibit a larger wage response to idiosyncratic shocks, consistent with these firms providing less insurance to workers. This pattern holds across a broad range of measures of financial constraints.

To explain these patterns, we develop a monopsony model with dynamic wage contracts. Firms post vacancies with state-contingent wage contracts, which allow them to offer insurance over idiosyncratic shocks. Risk-averse workers face search frictions: they infrequently receive a random job offer. Workers choose whether or not to accept based on the expected utility of the posted contract. Firms commit to the contracts, while workers cannot commit

[0]For example, the following papers use experimental/quasi-experimental variation in wages to identify a labor supply elasticity which they invert to estimate wage markdowns: Kroft et al. (2025); Amodio and De Roux (2024); Caldwell and Oehlsen (2023); Azar et al. (2022); Kline et al. (2019)

[1]In a review piece, Manning (2021) notes "The amount of monopsony power estimated in many studies is so high as to raise questions about how it can be reconciled with observed levels of profits."

to staying at the firm. As firms experience idiosyncratic productivity shocks, their value to keeping (and recruiting) a worker changes which leads them to offer wage contracts which vary with their performance. But risk aversion limits how much firms pass through these shocks to wages: workers value a single-period wage increase less than the same total compensation spread over time. Firms thus face a tradeoff between adjusting labor in response to a shock (which requires adjusting wages) and insuring workers. Labor accumulates slowly and responds to changes in the total value of the contract.

The model allows us to develop the notion of the *cross-sectional* labor supply elasticity—how much larger is the employment of firms which offer permanently higher wages?—and contrast it with the *shock-identified* labor supply elasticity. Steady-state, cross-sectional analysis is key to estimating wage markdowns,[2] but it does not tell us how to interpret the labor supply elasticities that researchers estimate using idiosyncratic shocks. We show that, when a shock is not fully permanent, there is no horizon over which the inverse shock-identified elasticity equals the markdown. Our model shows how the level and dynamics of the shock-identified elasticity can identify the cross-sectional elasticity and, in turn, the wage markdown.

It is difficult to disentangle market power and risk aversion empirically, since both affect the wage and labor pass-through of productivity shocks. Risk-averse workers value temporary wage changes less than permanent ones of the same discounted present dollar value. Additionally, workers respond to changes in the value of the contract, but we empirically observe only a limited set of wage changes. To address this challenge, we leverage the fact that the firm provision of insurance has a unique dynamic pattern: risk aversion makes it costly for firms to lower the promised utility after the shock has dissipated. So when workers are risk neutral, the firm will raise wages sharply after a productivity shock and then lower them quickly as the shock dissipates. But when workers are risk averse, the wage response is relatively more persistent as risk aversion induces wage smoothing.

We use simulated method of moments to estimate our model to match the dynamic response of wages and labor to idiosyncratic shocks that we present in Section 2, as well as several aggregate moments. The average worker's wage is marked down 8.3%—a relatively modest markdown compared with the literature, which typically finds markdowns between 15% and 50% (Azar and Marinescu, 2024). If we adopted the static model approach of inverting the shock-identified labor supply elasticity, we would have estimated a markdown as wide as 26%. We also estimate a coefficient of relative risk aversion of 5.97, suggesting that workers are considerably risk-averse: the average worker would give up roughly 3% of

[2]This steady state cross sectional analysis is common in dynamic monopsony models (as popularized by Manning (2003)).

pay to avoid a 10% symmetric gamble in wages. Indeed, risk aversion is necessary to explain the small, relatively flat wage response to idiosyncratic shocks. When we re-estimate our model assuming workers are risk neutral, we dramatically overshoot the level of the wage response in the first years following the shock and find that it fades away too quickly in the later years.

The response of wages to idiosyncratic productivity shocks (the "wage pass-through") is also commonly used to estimate wage markdowns, independently of how labor responds (Kline, 2025; Lamadon et al., 2022; Card et al., 2018). Using our estimated model, we show that the extent of worker risk aversion strongly influences the level and persistence of the wage pass-through. If workers were risk neutral, we would expect an average 0.086% wage increase in the six year window following the 1% shock—almost three times larger than the 0.03% average increase we observe. This shows that the choice to model the firm provision of wage insurance highly influences our understanding of what drives the firm's wage response to shocks. The empirical fact that firms adjust wages in response to idiosyncratic productivity shocks is evidence of the *existence* of monopsony power—there is not one market wage—but using the wage pass-through to infer the *extent* of market power is very sensitive to modeling decisions.

Finally, we show that disentangling market power from insurance is important because insurance has direct consequences for the economy. Since firms face a tradeoff between insuring workers and adjusting labor following a shock, wage insurance prevents some productivity-improving job-to-job transitions from occurring. Relative to a social planner who can perform redistribution but is constrained by search frictions, labor responds too slowly to shocks—especially when workers are risk-averse. This suggests that the firm provision of insurance distorts the job ladder and may reduce labor market dynamism.

Literature. We contribute to a large and growing literature that models and quantifies labor market power.[3] Manning (2021) highlights that the extent of monopsony power typically estimated is so large as to be puzzling—if firms have such enormous market power, why are profits not higher? Recent work has proposed resolutions to this puzzle: Bloesch et al. (2024) show that adjusting the wage markdown to account for recruiting costs can reconcile measured labor supply elasticities with observed firm profits. This parallels Seegmiller (2021); Chan et al. (2020) which both show that labor adjustment costs can influence the wage markdown. Berger et al. (2022) and Agostinelli et al. (2025) both show that the labor supply elasticity which is typically estimated empirically is not a sufficient statistic for the wage markdown, highlighting the role of strategic interactions and human capital accumulation, respectively. Similarly, we show that in the presence of search frictions and risk

[3]See Kline (2025); Azar and Marinescu (2024); Manning (2021) for recent surveys

aversion, the shock-identified labor supply elasticity does not map onto the cross-sectional elasticity relevant for wage markdowns.[4] Neglecting these dynamics can lead to a substantial overestimate of market power.

To model the dynamics of frictional labor adjustment, we build on the "dynamic monopsony" framework of Manning (2003), where search frictions and wage posting in the style of Burdett and Mortensen (1998) give firms monopsony power. The other common microfoundation for monopsony power is job-differentiation, where the absence of labor market frictions mean that labor responds instantaneously to a wage change (Berger et al., 2022; Lamadon et al., 2022; Chan et al., 2020). Dynamic monopsony models all feature worker-level dynamics, as workers receive random job offers and are exogenously separated into unemployment, but often have no idiosyncratic firm shocks or other firm dynamics (Bontemps et al., 2000; Jarosch et al., 2024; Gottfries and Jarosch, 2023). Gouin-Bonenfant (2022) and Bloesch et al. (2024) feature idiosyncratic shocks, but we directly focus on how these firm dynamics shape the measurement and interpretation of monopsony power. Additionally, a distinctive feature of our model is that we incorporate worker risk aversion and allow firms to offer state-contingent wage contracts. This allows us to show how risk aversion shapes workers' responses to temporary wage changes and how the firm provision of insurance affects wage-setting behavior.

By modeling how monopsonistic firms provide wage insurance and using this to match the empirical wage pass-through that we estimate, we also contribute to the literature measuring and interpreting the wage pass-through. The wage pass-through refers to the elasticity of wages with respect to value added per worker (or other measures of firm performance). This empirical object is often interpreted through a market power framework and labeled as "rent sharing" (Van Reenen, 1996; Card et al., 2016, 2018; Kline et al., 2019; Lamadon et al., 2022; Garin and Silvério, 2023) or through an insurance framework and labeled as "risk sharing" (Guiso et al., 2005; Friedrich et al., 2019; Balke and Lamadon, 2022). We show that both worker risk aversion and search frictions can reduce the wage pass-through, which suggests that using the level of pass-through to infer wage markdowns without modeling insurance, as in Card et al. (2018) and Lamadon et al. (2022), will bias the estimate of market power.

This paper builds on the literature studying wage contracts by writing a novel model with firms that contract with workers over *idiosyncratic* productivity shocks and where *random* search frictions influence the worker's ability to commit. Early papers study the insurance-incentive tradeoff in wage contracts in a competitive setting where the worker's

[4]This point that a short-run labor supply elasticity might not be enough to determine wage markdowns dates back to Boal and Ransom (1997), but by formalizing the relationship between shock-identified and cross-sectional labor supply elasticities we show how the researcher can infer the extent of monopsony power through how firms respond to shocks.

ability to quit for an exogenous outside option limits the extent of insurance (e.g., Baily (1974); Azariadis (1975); Harris and Holmstrom (1982)).[5] Recent work has studied wage contracting in the context of directed (competitive) search which is tractable due to having a block recursive equilibrium (Menzio and Shi, 2010; Balke and Lamadon, 2022; Souchier, 2022). We overcome the tractability problems associated with an equilibrium that is not block recursive to allow us to study wage contracting over idiosyncratic shocks within the workhorse monopsony model of random search. Finally, relative to Moscarini and Postel-Vinay (2013, 2016); Fukui (2020) who adopt random search models with wage contracting over aggregate shocks to study business cycle dynamics, we model idiosyncratic shocks which are the most directly related to the empirical monopsony literature. We also show that idiosyncratic shocks lead to inefficient job-to-job transitions even with risk-neutral workers, which contrasts with Moscarini and Postel-Vinay (2013) where the decentralized equilibrium with aggregate shocks features efficient labor reallocation.

Road Map. Section 2 offers motivational empirical evidence on how wages and employment respond to shocks and how this varies by firm financial constraints. Section 3 develops our model. We characterize the optimal contract in 4. We quantify the model in Section 5. Finally, we turn to the planner's problem and show how idiosyncratic shocks distort the job ladder in Section 6.

2 Empirical Patterns

In this section, we present several empirical facts. First, we study the dynamics of how idiosyncratic shocks pass through to wages and employment. We show that the wage pass-through is small and relatively persistent over time, while the labor response accumulates in the years following the shock. This motivates a model with search frictions (to explain labor dynamics) and insurance (to explain the small, persistent wage pass-through). Finally, we show that wages and employment respond more to shocks at financially constrained firms which may be less able to offer insurance.

2.1 Data

We use several US Census Bureau data products: the Revenue-Enhanced Longitudinal Business Database (LBD), the Longitudinal Employer-Household Dynamics (LEHD), and the Quarterly Financial Report. We complement these with Compustat data for publicly traded

[5] Similarly, Rudanko (2009, 2011); Kudlyak (2014) study wage contracting in a search-and-matching model without on-the-job search

firms.

Longitudinal Employer-Household Dynamics (LEHD) The LEHD is a Census Bureau employer-employee matched dataset containing restricted-use microdata with quarterly earnings for jobs in the United States. The LEHD data is created by combining data shared by states on state Unemployment Insurance earnings records and the Quarterly Census of Employment and Wages with other censuses and surveys. We have access to data from 28 states and from 1997 to 2022. We annualize earnings by adjusting for the number of quarters a worker worked at a firm, following Sorkin (2018) and Abowd et al. (2003). Details of the data construction are in Appendix A1. Our primary variable that we construct using the LEHD is a residualized measure of average annual earnings at the firm, where individual earnings are annualized, and then adjusted by removing person–state fixed effects. This allows us to control for composition changes when studying the response of wages to shocks.

Longitudinal Business Database (LBD) The U.S. Census Bureau Longitudinal Business Database (LBD) is a restricted-use establishment-level panel that links the Census Business Register across years to track all U.S. employer businesses with paid employees. The LBD is built from the Business Register and integrates administrative tax records with Census surveys. It provides annual coverage at both the establishment and firm levels, spanning all industries and states. We have access to the LBD from 1997 to 2022. The revenue-enhanced LBD also contains firm-level revenue. We use the LBD primarily for employment, payroll, revenue, and NAICS. Note that our payroll and employment measures are for the first quarter of the year and revenue is an annual measure.

Quarterly Financial Report (QFR) The Quarterly Financial Report (QFR) is a survey of firms conducted by the US Census Bureau, providing income statements and balance sheets. The survey targets manufacturing companies with total assets of $250,000 or more, and companies in other industries (mining, wholesale trade, retail trade, information, or professional and technical services) with total assets of $50 million or more. Relative to Compustat, the QFR provides data for smaller and private firms, which are likely to be more financially constrained than Compustat firms. We use the QFR to provide new descriptive evidence of the correlation between the wage pass-through and measures of financial constraints. We have access to the QFR in Census years—every 5 years from 1997 to 2022. When constructing firm-level measures of financial constraints used in our analysis in Section 2.3 we take the average across all available years to generate a time-invariant measure.

Compustat We link our firm-level Census Bureau datasets to Compustat, which allows us to construct firm-level measures of value added per worker. Compustat provides standardized annual and quarterly financial statement information for publicly traded U.S. corporations. This allows us to define value-added as the sum of operating income before depreciation, changes in inventories, and labor expenses, as in Donangelo et al. (2019). We calculate labor expenses as the product of Compustat employment and LEHD earnings.[6] This also allows us to define the input cost share. Specifically, we calculate non-labor input costs as being the difference between revenue and total value-added and define the ratio of non-labor input costs to total revenue as the non-labor input cost share.

2.2 Dynamic responses of wages and employment to shocks (+ cut by financial constraints)

Relative to existing models of monopsony, a key contribution of ours is to match the dynamics of wages and employment in response to idiosyncratic shocks.

To understand how firms respond to shocks, we estimate the following relationship between changes to value added per worker and cumulative firm outcomes over time.

$$\log(Y_{j,t+\tau}) - \log(Y_{j,t}) = \alpha + \beta_\tau(\log(\text{VA pw})_{j,t+1} - \log(\text{VA pw}_{j,t})) + \mathbf{X_{jt}} + \varepsilon_{\mathbf{jt}} \tag{1}$$

Y is (residualized) firm level wages or firm employment. We estimate equation 1 for $\tau \in [1, 6]$ to capture dynamics in the response. Control variables $\mathbf{X_{jt}}$ include two digit NAICS by year fixed effects and controls for 2 lags of the growth of Y and 2 lags of the growth of value added per worker.[7] Standard errors are clustered by gvkey (the firm identifier) and year.

To isolate exogenous, idiosyncratic shocks, we instrument for the change in value added per worker $\log(\text{VA pw})_{j,t+1} - \log(\text{VA pw})_{j,t}$ using (non-labor) input cost share shocks. We use the change in the cost share—the ratio of non-labor input costs to revenue—between $t-1$ and t as our instrument for changes in value-added per worker. We use the input cost share, rather than input costs in levels, because this isolates changes in profitability without mechanically picking up changes in quantity.[8] This allows us to isolate changes in wages and

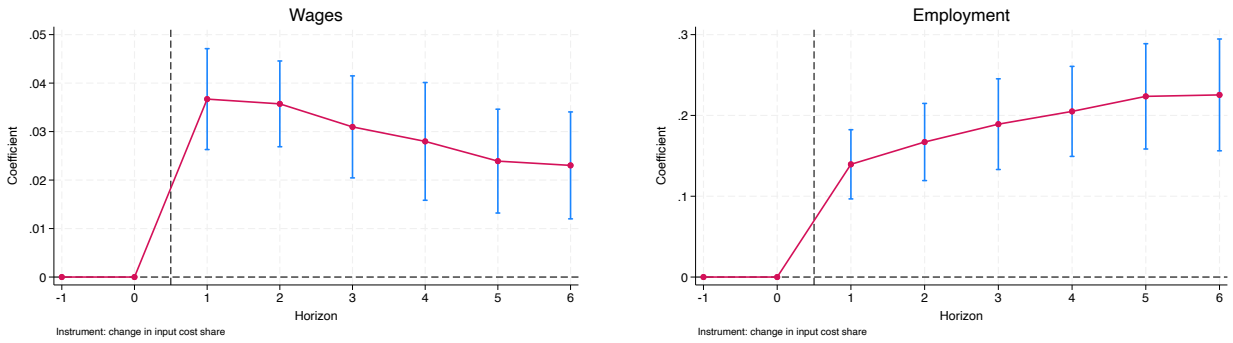
[6] This is because wage information is not readily available for most firms in Compustat. LEHD earnings are annualized as described in Appendix A1

[7] This specification is a Jordà (2005) local projection: separate linear regressions by horizon, so β_τ is the τ-step impulse response.

[8] For example, a labor supply shock could induce a change in quantity produced which would imply higher input expenditure without any change in profitability

employment driven by shocks to labor demand, rather than to labor supply. By including NAICS by year fixed effects we also isolate idiosyncratic cost shocks, which is desirable because sectoral labor supply elasticities are unlikely to be the same as firm-level elasticities (Souchier, 2022; Garin and Silvério, 2023). As an example, our instrument would pick up if a manufacturer's gas contract was renegotiated—but *not* oil price shocks (fixed effects will expunge industry shocks). Similarly, our instrument would pick up if a restaurant's food wholesaler goes bankrupt—but *not* a business cycle shock (due to our time fixed effects).

Figure 1: Impulse responses to value-added per worker shocks



(a) Wages (b) Employment

Notes: These figures show the impulse response of wages and employment to value-added per worker changes (where the input cost share is the instrument). Our wage measure is the firm's LEHD annualized earnings with person-state fixed effects residualized. Employment is Compustat employment. Horizons are years. All specifications include NAICS2×year fixed effects and two lags of wage changes and value-added per worker changes. Standard errors are two-way clustered by firm and year. This exhibit is from FSRDC Project Number 3060 (CBDRB-FY25-P3060-R12493).

Figure 1 plots the coefficients β_τ. There are several empirical patterns here that motivate our model. First, the response of wages to this value-added shock is small. The one year wage pass-through is 0.0367. Card et al. (2018) discuss that the elasticity of wages with respect to value added per worker typically falls between 0.05 and 0.15. They also note that studies which estimate the pass-through of firm shocks to individual wages find smaller values, sometimes below 0.05, suggesting that the fact we control for worker-state fixed effects in constructing our wage measure may explain our small wage pass-through. Our shock is also fairly transitory; when we estimate the annual persistence of the shock,[9] we find $\hat{\rho} = 0.561$ (0.042). Our model will explicitly treat this shock as temporary and match the measured persistence.

[9]Specifically, we estimate $\log(\text{VA pw}_{j,t+2}) - \log(\text{VA pw}_{j,t}) = \alpha + \rho\left(\log(\text{VA pw}_{j,t+1}) - \log(\text{VA pw}_{j,t})\right) + \mathbf{X_{jt}} + \varepsilon_{\mathbf{jt}}$. Our estimate $\hat{\rho}$ captures the extent that a shock to value added per worker in $t = 1$ persists in $t = 2$

Next, note that the wage response is fairly stable. Despite the value-added shock being non-permanent, by year six the wage pass-through is 0.023—not too dissimilar from the one year wage pass-through of 0.037. In our model, insurance will be key to matching both the small wage pass-through as well as the relatively flat wage response.

Finally, we can see that labor accumulates gradually in response to the shock. This means that the labor supply elasticities implied by the analysis also increase when estimated over a longer horizon. The one year labor supply elasticity is 3.8 which is standard for the literature—Azar and Marinescu (2024) note that values between 2 and 6 are standard. By year six, the wage response has fallen somewhat while labor has continued to increase—resulting in a labor supply elasticity of 9.78. The empirical fact that the labor supply elasticities may be larger when estimated over a wider horizon is known[10], and our paper interprets these dynamics through our model. Intuitively, when workers infrequently receive job offers it takes time for employment to adjust following a wage change. Additionally, even when the firm has begun to reduce the wage back to its baseline levels, it takes time for workers to receive better offers. These two forces both can lead to a larger labor supply elasticity when estimated over a wider horizon.

To get more directly at how insurance influences the wage and labor response to shocks, we estimate the following equation:

$$\begin{aligned}\log(Y_{j,t+\tau}) - \log(Y_{j,t}) = \alpha + \beta_\tau \Delta \log(\text{VA pw})_{j,t+1} + X_{jt} \\ + \gamma_\tau \left[\Delta \log(\text{VA pw})_{j,t+1} \times L_{j,t}\right] + \varepsilon_{jt}.\end{aligned} \tag{2}$$

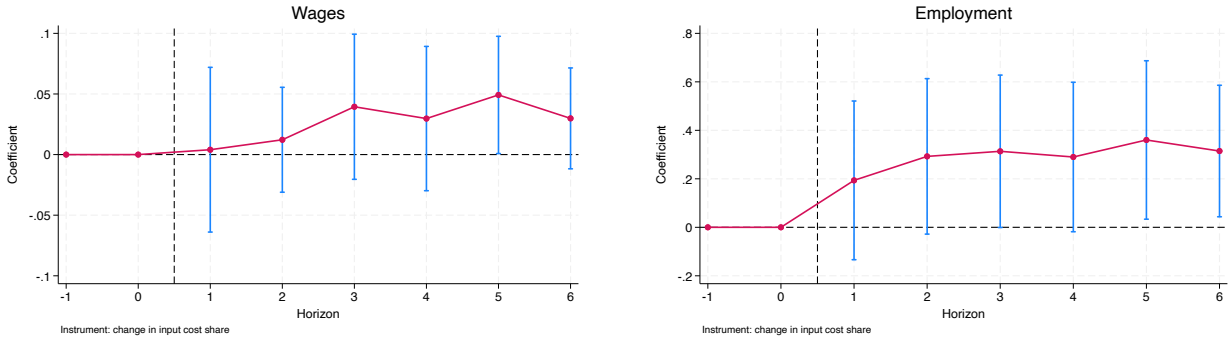
This is the same as the previous equation 1 but it now contains an interaction for firm leverage, $L_{j,t}$. Levered firms may face financial constraints which make offering insurance to workers against shocks more costly. For example, in response to a negative shock they may have to cut wages and employment despite this being costly—and they may have to ex-ante pay higher wages due to this risk. Indeed, the corporate finance literature has noted that leverage may result in a firm's workers being exposed to more risk.[11]

In Figure 2 we plot the continuous interaction γ_τ. Indeed, more levered firms have more volatile wages and labor. Going from a firm with a leverage ratio of 0.25 to 0.75 would predict an average 0.0137pp (46%) larger wage pass-through in the six year post period and an average 0.147pp (77%) larger employment response.

[10]Seegmiller (2021) establishes this fact using stock returns as an instrument and finds a similar pattern in the dynamics of the labor supply elasticity

[11]For example, Giroud and Mueller (2017) show that levered firms cut employment more during the Great Recession and Sharpe (1994) finds that levered firms have more cyclical employment

Figure 2: Heterogeneous response by firm leverage



(a) Wages

(b) Employment

Notes: These figures plot the estimated interaction coefficients γ_τ from Eq. 2, which capture how the response of wage and employment growth to value-added-per-worker changes varies with firm leverage $L_{j,t}$. Our wage measure is the firm's LEHD annualized earnings with person-state fixed effects residualized. Employment is Compustat employment. Horizons are years. All specifications include NAICS2×year fixed effects and two lags of wage changes and value-added-per-worker changes. Standard errors are two-way clustered by `gvkey` and year. This exhibit is from FSRDC Project Number 3060 (CBDRB-FY25-P3060-R12493).

2.3 The wage pass-through and financial constraints

In this section, we explore heterogeneity by financial constraints and find, broadly, that less financially constrained firms have a lower pass-through of revenue per worker into wages, suggesting insurance may also be relevant to explaining patterns in the relationship between wages and firm performance. This motivates our goal of understanding how insurance affects firms' wage-setting behavior.

How does the pass-through of revenue per worker into wages vary by measures of financial constraints? This exercise uses several measures of financial constraints that we measure in the Quarterly Financial Report microdata (QFR). The QFR allows us to observe the financial constraints for smaller, private firms—including those with as little as $250,000 in assets. This gives us much more heterogeneity in financial constraints than Compustat (Crouzet and Mehrotra, 2020).

We use several firm-level measures of financial constraints that are common in the corporate finance literature: (1) log total assets (Hadlock and Pierce, 2010) (2) the ratio of short-term liabilities to total liabilities (Kalemli-Özcan et al., 2022) (3) the ratio of total liabilities to total assets (Sharpe, 1994) (4) the ratio of short-term liabilities to total assets (Tian et al., 2015) and (5) cash per worker (Campello et al., 2010).

Table 1: Pass-through of $\Delta \log \text{rev pw}$ into $\Delta \log w$ by financial constraints ($\hat{\beta}_2$)

	(1) $\Delta \log w$	(2) $\Delta \log w$	(3) $\Delta \log w$	(4) $\Delta \log w$	(5) $\Delta \log w$
$\Delta \log \text{rev pw} \times F$	−0.00727	0.0160	−0.00985	0.0125	−0.00233
	(0.000344)	(0.00205)	(0.00240)	(0.00483)	(0.000258)
$F =$	log total assets	$\frac{\text{ST debt}}{\text{Total debt}}$	$\frac{\text{Total debt}}{\text{Total assets}}$	$\frac{\text{ST debt}}{\text{Total assets}}$	Cash per worker
1 SD effect in F (as % of baseline $\hat{\beta}_2$)	−147.4	33.9	−20.8	12.6	−41.6
N	3,080,000	3,080,000	3,080,000	3,080,000	3,080,000

Notes: Outcome is $\Delta \log w_{it}$ (LBD payroll per worker). Establishment–year data weighted by establishment share of firm employment. SEs clustered at the firm level. Models include NAICS 4-digit × year fixed effects and main effects for $\Delta \log \text{rev pw}$ and F. "1 SD effect in F, as % of baseline $\hat{\beta}_2$" $= 100 \times \frac{(1 \text{ SD effect of } F \text{ on } \hat{\beta}_2)}{0.009907}$. This exhibit is from FSRDC Project Number 3060 (CBDRB-FY25-P3060-R12493).

$$\Delta \log(\text{w})_{it} = \beta_0 + \beta_1 \Delta \log(\text{revenue pw})_{f(i)t} + \beta_2 \Delta \log(\text{revenue pw})_{f(i)t} \times F_{f(i)} + \beta_3 F_{f(i)} + \gamma_{kt} + \varepsilon_{it} \tag{3}$$

where i indexes establishments, $f(i)$ and denotes firms, $F_{f(i)}$ is the measure of financial constraints, $\log(w)_{it}$ is LBD payroll per worker. We cluster at the firm level, and γ_{kt} denotes fixed effects for NAICS 4 k by year t.

In Table 1 we report the values of $\hat{\beta}_2$ for our five different measures of financial constraints. To aid interpretation, we also report the effect of a one standard deviation change in each financial constraint measure as a percentage of the baseline pass-through coefficient. These results broadly reveal that less financially constrained firms have smaller wage pass-throughs. Firms with more total assets may be more able to spend down or borrow to smooth against shocks. Indeed, they have a lower wage pass-through. Similarly, firms with cash per worker may be more liquid and able to smooth their payroll—this is consistent with these firms also having a smaller wage pass-through. Firms with a higher share of short-term liabilities (either relative to total liabilities or relative to total assets) may be more vulnerable to rollover risk which could translate to more volatile pay. We do see that both having a higher share of short-term liabilities to total liabilities and short-term liabilities to total assets are correlated with having a larger wage pass-through. Lastly, we would expect that firms which are more levered—that is, have a higher share of total liabilities to total assets—would be less able to borrow in response to shocks and would pass more risk onto their workers. We find the opposite of this, but this might be explained by the fact that levered firms tend to

be larger and larger firms generally have lower productivity-wage pass-throughs (Rajan and Zingales, 1995). Notably, in our Compustat sample which is large, public firms, we have shown in Figure 2 that levered firms have a larger wage pass-through.

We have shown that financially constrained firms have larger wage pass-throughs, that the wage response to shocks is small and relatively flat, and that the labor response accumulates over time. Now, we write a model which will allow us to use the wage and labor dynamics to disentangle market power and insurance.

3 Model

We write a random search model with on the job search (Burdett and Mortensen, 1998), idiosyncratic shocks, risk-averse workers, and dynamic wage contracts. This model parsimoniously allows us to model how insurance and monopsony both shape the wage and labor response to shocks—and, in turn, understand which empirical moments are required to estimate the level of market power.

By adding idiosyncratic shocks, we are able to match the empirical within-firm labor supply elasticity and the within-firm wage pass-through which are estimated using shocks. We will refer to these objects as the *shock-identified* labor supply elasticity and *shock-identified* wage pass-through. We allow firms to contract over these shocks when hiring risk-averse workers—this means the firm's role as an insurer can shape the wage and labor dynamics. These dynamics are not just shaped by worker risk aversion but also by firm market power which influences the firm's ability to recruit and retain more workers as it adjusts the compensation.

3.1 Market Power in Search Models

We begin by offering an intuition for what generates monopsony power in this class of models. In random search monopsony models, firms have market power because workers face search frictions. The use of Burdett and Mortensen (1998) as a model of "dynamic monopsony" was popularized by Manning (2003).[12] Any active firm will recruit all the unemployed workers it contacts, but it will only successfully poach employed workers who are compensated less. Since these employed workers only receive a job offer with odds λ_e, this means that a firm will not immediately lose all of its workers to a competitor if it pays a lower wage—as would be true in the competitive case. The extent of these search frictions affects the competitive pressure firms face and, in the edge case where there are infinite search frictions and employed

[12] "Dynamic" refers to modeling labor flows, not modeling productivity dynamics

workers never receive competing offers, firms will mark down the wage fully to the worker's outside option.

Search frictions are not the only force that influences firm monopsony power in this setting. Firms vary in their permanent productivity which means that a firm's relative position on the job ladder can influence the amount of local competition it is exposed to. Gouin-Bonenfant (2022) formalizes and quantifies how increased dispersion in the productivity distribution can lead to a smaller labor share.

Search frictions and dispersion in the productivity distribution are two key parameters that influence monopsony power because they affect how firms are able to accumulate workers in the long run. If a firm permanently offers a more desirable contract, how many workers will it accumulate? Proposition 1 formalizes this concept as the *cross-sectional* labor supply elasticity.

3.2 Environment

Time is discrete and all agents discount future utility at a rate β. There is a unit mass of workers who can be either employed or unemployed. There is a unit mass of firms which can either be active or inactive (employing zero workers). Workers are hand-to-mouth and have a utility function $u(\cdot)$. This assumption can be justified by the fact that many workers, even in wealthy countries, have little or no liquid wealth (Kaplan et al., 2014).[13] Firms face idiosyncratic shocks. Firms post state-contingent wage contracts, and risk-averse workers accept a job they match with if the expected value of taking the job exceeds the value of not taking the job. Both employed and unemployed workers search. Search frictions govern the probability that they receive job offers.

3.2.1 Timing

Each period, the timing of actions is as follows:

1. Production and wage payments take place in time t; unemployed workers receive flow utility b
2. Exogenous separations occur with odds δ
3. Firms post vacancies with state-contingent wage contracts
4. Idiosyncratic productivity shocks are drawn s_{t+1}

[13]For an exploration of savings in a model of wage contracting see Souchier (2025)

5. Unemployed and employed workers match with vacancies with odds λ_u, λ_e and decide whether to accept or reject the job offer

6. Endogenous separations occur

3.2.2 Contracts

Firms post and commit to state-contingent wage contracts. The future history of productivity shocks is observable to both workers and firms and so can be contracted on. Firms can commit to the contract, but workers cannot commit to not leave for a better employer if they receive a competing offer. The firm is subject to a fairness constraint—it must pay all workers the same wage. This fairness constraint rules out tenure-based contracts, as in Burdett and Coles (2003), and counter-offers as in Cahuc et al. (2006). This also means that firms cannot offer insurance to workers over the risk of exogenous separations. Our strict fairness constraint is common within the wage contracting literature (the Burdett and Mortensen (1998) wage posting tradition features only a single posted wage) because it aids tractability and because internal equity is a real managerial concern (Bewley, 1999; Breza et al., 2018).

3.2.3 Unemployed Workers

Unemployed workers earn a flow utility b and receive an offer from a random employer with probability λ_u. They can choose to accept or reject the offer.

$$V_U = u(b) + \beta(1 - \lambda_u)V_U + \beta\lambda_u \int_{\underline{V}}^{\bar{V}} \max\{V_U, V_0\} dF(V_0) \tag{4}$$

The first term b, is the flow payment for unemployment. In the second term, the worker does not receive a job offer and stays unemployed. In the third term, the worker receives a job offer with probability λ_u and accepts if the value from the job exceeds the value of unemployment. $F(\cdot)$ denotes the distribution of posted utilities.

3.2.4 Employed Workers

Employed workers are risk-averse and receive utility $u(w_t)$ from the wage w_t paid out by their current job's wage contract at the state realized in t. Search is random. They receive job offers with odds λ_e and can choose to accept or reject the offer, but cannot use the offer to renegotiate pay at their current employer. Notably, they cannot commit to stay at their current employer if the offer they receive exceeds the value of staying. They exogenously

separate into unemployment with probability δ and may endogenously separate into unemployment if the value of their contract falls below the value of being unemployed. The value of employed workers is thus given by V_t.

$$V_t = u(w_t) + \beta(1-\lambda_e)(1-\delta)\mathbb{E}_{s_{t+1}}\left[\max(V_{t+1}, V_U)\right] + \beta\lambda_e(1-\delta)\mathbb{E}_{s_{t+1}}\left[\int \max(V_0, V_{t+1}, V_U)dF(V_0)\right] + \beta\delta V_U \quad (5)$$

In Equation 5, the first term is the utility from wages received today. The second term describes the worker who does not receive a job offer and does not separate exogenously. The worker will stay at their current firm unless the value of staying is worse than quitting into unemployment. The third term describes the worker receiving a job offer and accepting it if the value exceeds the value of staying at the current job and the value of unemployment. The fourth term describes the worker exogenously separating into unemployment.

3.2.5 Firms

Firms vary in their permanent productivity z_j and also face idiosyncratic, temporary productivity shocks s_{jt}, making their total productivity in t $(1+s_{jt})z_j$. The productivity shocks are drawn from the following mean-reverting process:

$$s_{jt} = \rho s_{j,t-1} + \varepsilon_{jt}$$

The firm's productivity level and its labor, L_t, determine its output which is produced using a linear production technology, $(1 + s_{jt})z_j L_{jt}$. The firm's labor evolves as its incumbent workers separate into unemployment or leave for another job and it recruits new workers through its posted wage contract. As discussed previously, the wage contracts are state-contingent and firms can commit. Instead of optimizing over the optimal wage to offer in every possible state (which is an infinite dimensional problem), we can follow Spear and Srivastava (1987) and write the firm's problem recursively. The firm takes its incumbent workers' promised utility as a state variable and chooses which utility to post in each state next period—subject to the promise-keeping constraint that its incumbents must receive the utility (current wage and future promised utilities) they were promised.

The firm solves:

$$\Pi(V_t, L_t, s_t, z) = \max_{w_t, V_{t+1}(s_{t+1})} ((1+s_t)z - w_t)L_t + \beta\mathbb{E}_{s_{t+1}}\left[\Pi(V_{t+1}, L_{t+1}, s_{t+1}, z)\right] \quad (6)$$

subject to a promise-keeping constraint:

$$\begin{aligned} V_t = u(w_t) + \beta(1-\lambda_e)(1-\delta)\,\mathbb{E}_{s_{t+1}}[V_{t+1}] \\ + \beta\lambda_e(1-\delta)\,\mathbb{E}_{s_{t+1}}\left[\int \max\left(V_0, V_{t+1}\right) dF(V_0)\right] + \beta\delta V_U \end{aligned} \tag{7}$$

Firms post promised utilities V for each state s for $t+1$, which affects the number of workers $L_{t+1}(V_{t+1}, L_t)$ in each state. The promise-keeping constraint implies a wage w_t that the firm must pay out to incumbent workers to satisfy the utility they were promised V_t. For example, to reduce the number of workers it has (say, in response to a bad shock), the firm must cut the posted V from one period to the next. When cutting V, the firm will have to pay a high enough wage to incumbent workers to satisfy this constraint. The impact of the promise-keeping constraint grows as the firm has more incumbent workers. If workers are risk neutral, the promise-keeping constraint does not influence the firm's problem since utility is transferrable; the firm will choose the future utilities to post to maximize profits and then will pay the incumbents the wage required to satisfy the promise. Paying the worker a large sum at once becomes costly when the worker is risk-averse.

Labor evolves as follows:

$$\begin{aligned} L_{t+1}(V_{t+1}, L_t) = L_t(1-\delta)\left(1-\lambda_e(1-F(V_{t+1}))\times\mathbb{I}[V_{t+1}>V_U]\right) + \ \lambda_u u \times \mathbb{I}[V_{t+1}>V_U] \\ + \lambda_e(1-u)(1-\delta)G(V_{t+1}) \end{aligned} \tag{8}$$

The first term reflects that the firm will keep incumbents who do not exogenously separate, leave for a better job, nor endogenously separate. The second term reflects hiring from unemployment. The third term reflects poaching of employed workers from other firms. $G(\cdot)$ denotes the distribution of employed workers' current promised utilities.

3.3 Equilibrium Definition

Let $\mu(V, L, s, z)$ refer to the joint distribution of firms across incumbent promised utility, labor, productivity shocks, and productivity levels. A *recursive equilibrium* is a set of value functions, policies, and distributions $\{V_U,\, V^{POL}(V, L, s, z),\, \Pi(V, L, s, z), \mu(V, L, s, z)\}$ such that (i) firms' policies maximize profits subject to the promise-keeping constraint and fairness constraint (ii) workers decide whether to accept job offers to maximize utility (iii) the distribution $\mu(V, L, s, z)$ is consistent with firms' posting strategies and workers' acceptance rules (iv) labor market clearing holds: $\int L(V, L, s, z)d\mu(V, L, s, z) + u = 1$.

The equilibrium is a fixed point in the distribution $\mu(V, L, s, z)$. Note that the cross-

sectional distributions of posted contracts and employment across promised utilities, $F(V)$ and $G(V)$, are implied from $\mu(V, L, s, z)$.

4 Characterizing the optimal contract

Here we characterize the equilibrium by considering the firm's first order conditions associated with the optimization problem (6). The firm chooses the incumbent workers' wage payments for this period w_t and the posted state-contingent promised utilites $V_{t+1}(s_{t+1})$.

$$\mu_t u'(w_t) = L_t \tag{9}$$

$$\mu_{t+1} = \mu_t \left[(1-\lambda_e)(1-\delta) + \lambda_e(1-\delta)F(V_{t+1})\right] + \Pi_L L'(V_{t+1}) \tag{10}$$

where μ is the Lagrange multiplier on the promise-keeping constraint. Recall that the promise-keeping constraint pins down the wage, given the incumbent's previous promised utility. To cut the expected promised utility in $t+1$, firms must pay workers a sufficiently high wage to keep previous promises. Equation 9 is the FOC with respect to w_t. Here, we see that the firm trades off the cost of a higher wage, L_t (since they must pay all incumbents the higher wage), against the benefit of a higher wage, that they relax the promise-keeping constraint by $u'(w_t)$, scaled by the benefit to the firm of a looser promise-keeping constraint μ_t. Note that cost of a higher wage increases with the number of incumbents. The curvature of the utility function also matters: since $u'(\cdot)$ falls with w, delivering a given increase in promised utility through the wage is more expensive as the wage increases. Equation 10 is the FOC with respect to V_{t+1}. The left hand size reflects tomorrow's promise-keeping constraint–the cost of increasing V_{t+1} is a tighter promise-keeping constraint tomorrow. The benefit of increasing V_{t+1} is made up of two terms. The first term is the benefit of loosening today's promise-keeping constraint μ_t (scaled by share of workers who stay–which will increase in response to an increase in V_{t+1}). The second term is the direct effect on profits Π of having more workers, scaled by the increase in workers associated with the increase in V_{t+1}.

4.1 Solution method

Burdett and Mortensen (1998) and a broad class of related models are tractable because the equilibrium is ranking preserving—the ranking of firms by posted wages is the same as the

ranking of firms by productivity [14]. By allowing firms to contract over idiosyncratic risk, our economy no longer has a ranking preserving equilibrium. A firm with a higher permanent productivity experiencing a negative temporary shock may post a promised utility above or below a firm with low permanent productivity experiencing a positive shock.

The equilibrium is characterized by a fixed point in the distribution of firms over promised utilities, employment, and productivity states. Solving for a fixed point in four continuous state variables is computationally challenging. To make progress, we develop a first-order approximation to the full risk equilibrium, which permits estimation and analytical characterization of firms' responses to shocks.

Specifically, we consider a first order perturbation of the zero-risk steady state equilibrium. The economy is in the zero-risk ($s = 0$) steady state equilibrium at time 0 and then learn that at time $t = 1$ there will be a small amount of idiosyncratic risk—$s_1 \in \{-\epsilon, +\epsilon\}$ with equal odds for all firms. This small shock has a persistence ρ so that $s_{j+1} = \rho^j s_1$. Note that, to first order, this introduction of a small amount of idiosyncratic risk does not change $F(V)$ or other equilibrium objects.

To see this, let $F(\bar{V}; \epsilon)$ denote the cumulative distribution function with ϵ-sized symmetric shocks, $V(z, s)$ denote the value posted by a firm with permanent productivity z who has experienced a shock $s \in \{-\epsilon, +\epsilon\}$, and $J(z)$ be the distribution of permanent productivity levels.

$$F(\bar{V}; \epsilon)|_{\epsilon=0} = \int Pr\left(\mathbb{I}\{V(z, s) \leq \bar{V}\}\right) dJ(z)$$

Linearizing around our small shock $s = \{-\epsilon, \epsilon\}$,

$$F(\bar{V}; \epsilon)|_{\epsilon=0} = \int Pr\left(\mathbb{I}\{V(z, 0) + V_s(z, 0)s \leq \bar{V}\}\right) dJ(z)$$

$$= \frac{1}{2}\int \mathbb{I}\{V(z, 0) + V_s(z, 0)\epsilon \leq \bar{V}\} dJ(z) + \frac{1}{2}\int \mathbb{I}\{V(z, 0) - V_s(z, 0)\epsilon \leq \bar{V}\} dJ(z)$$

Let z^* solve $V(z^*; 0) = \bar{V}$ so this is the productivity cutoff where $J(z^*) = F(\bar{V}; 0)$. So following the positive shock we have $V(z_+(\epsilon), 0) + V_s(z_+(\epsilon), 0)\epsilon = \bar{V}$. Linearizing around z^*, we have $V(z_+(\epsilon), 0) = V(z^*, 0) + V_z(z^*, 0)(z_+(\epsilon) - z^*)$. We can rearrange to show that, to first order:

[14] For example, this is also true in Gouin-Bonenfant (2022) where wages are exclusively a function of productivity or in Moscarini and Postel-Vinay (2013) where there are aggregate shocks but no idiosyncratic shocks

$$z_+(\epsilon) = z^* - \frac{V_s(z^*,0)}{V_z(z^*,0)}\epsilon$$

Similarly for the negative shock, we can show:

$$z_-(\epsilon) = z^* + \frac{V_s(z^*,0)}{V_z(z^*,0)}\epsilon$$

We know that $F(\bar{V};\epsilon)|_{\epsilon=0} = \frac{1}{2}J(z_+(\epsilon)) + \frac{1}{2}J(z_-(\epsilon))$. So we can see that this does not change to first order with ϵ:

$$\frac{dF(\bar{V};\epsilon)}{d\epsilon} = -\frac{1}{2}J'(z_+(\epsilon))\frac{V_s(z^*,0)}{V_z(z^*,0)} + \frac{1}{2}J'(z_-(\epsilon))\frac{V_s(z^*,0)}{V_z(z^*,0)} = 0$$

Intuitively, a small amount of idiosyncratic risk will not change the rank-preserving equilibrium properties of the zero-risk equilibrium. This is desirable because it means that we can solve for the productivity-wage pass-through while holding fixed the equilibrium objects from the zero-risk steady state—which can be solved for in closed form.

4.2 The wage markdown

What is the relevant labor supply elasticity for estimating wage markdowns? What can the flow labor response to shocks tell us about the markdown? And how does the markdown depend on worker risk aversion? In this section, we will leverage our first order approximation to show that the wage markdown depends on the cross-sectional labor supply elasticity. How employment adjusts to shocks can also be used to infer the markdown—but this relationship is not direct. Risk aversion reduces the worker's reservation wage which widens the average wage markdown, but this is the only channel through which risk aversion affects firm's wage-setting power.

Proposition 1 (Wage markdown)**.** *To first order in the extent of idiosyncratic risk s, the firm with permanent productivity z has a wage markdown, $m(z) \equiv \frac{z-w(z)}{w(z)}$, equal to:*

$$m(z) = \frac{1-q(z)}{u'(w(z))\frac{w(z)L_V}{L(z)}}(1-\beta q(z)) \tag{11}$$

$$= \frac{1}{\varepsilon_{Lw}^{CROSS}} \tag{12}$$

where L_V is the flow labor response to a change in promised utility, $q(z) \equiv (1-\delta)[1-\lambda_e(1-F(V(z)))]$ is the share of workers who quit each period, and ε_{Lw}^{CROSS} is the elasticity of steady-state employment with respect to an increase in the wage.

See Appendix A3 for proof

Proposition 1 highlights that the elasticity of steady-state employment with respect to a permanent increase in the wage can be inverted to estimate the wage markdown in a search model. We refer to this object as the *cross-sectional* labor supply elasticity because it is analogous to comparing the employment at firms with permanently different productivity levels (which offer permanently different wages).

What does the firm response to shocks tell us about the wage markdown? Consider a firm which increases its promised utility and, in the following period, its workforce expands. What does this flow labor response, L_V, tell us about the markdown? The expression 11 relates how labor adjusts to a shock to the markdown. Notably, this expression does not contain the frequently empirically estimated shock-identified labor supply elasticity—what percent does employment respond to a percent increase in wages induced by a productivity shock—which we will discuss in more detail in Section 4.3. Here, we focus on how firm wage and employment dynamics can be used to estimate wage markdowns[15]. We ask, what does L_V tell us about the long-run labor response to a permanent wage increase?

The flow labor response, L_V, differs in two important ways from the cross-sectional labor response. First, labor adjusts slowly over time. This means that if the firm maintained the same promised utility over future years, it would accumulate more workers. Some simple arithmetic allows us to relate the short-run labor response to a steady-state labor response. When we write $\frac{dL_t}{dV_1} = (1-q)\frac{dL_{t-1}}{dV_1} + L_V$ we can solve for the steady state and get $\frac{dL_{CROSS}}{dV} = \frac{L_V}{q}$. This means that we need to adjust the flow labor response by the fraction of workers which quit each period. Second, L_V is how labor responds to a change in value, not a change in the wage. The firm can increase the worker's promised utility with many different types of wage contracts that may vary in their cost to the firm, but in steady-state this is achieved through a constant, permanent wage increase. To determine the costs increasing the value of the contract through a permanent wage increase, we can differentiate the promise-keeping constraint with respect to a permanent promised utility increase to find that the corresponding wage change is $\frac{1-\beta(1-\delta)[1-\lambda_e(1-F(V(z)))]}{u'(w(z))}$ (details in Appendix A3).

Proposition 2 (Risk aversion and the wage markdown)**.** *To first order in the extent of idiosyncratic risk s, increasing worker risk aversion lowers the reservation wage and results in wider wage markdowns $m(z) = \frac{z-w(z)}{w(z)}$ when the search efficiency of unemployed workers is greater than that of employed workers, $\lambda_u > \lambda_e$.*

There is no effect of worker risk aversion on the wage markdown when the reservation

[15]Empirical estimates of the cross-sectional labor supply may be confounded by amenities and other permanent differences across firms. Therefore, it may be easier to find credible empirical estimates of how firms respond to shocks

wage is held constant.

See Appendix A3 for proof

Do equilibrium wage markdowns depend on worker risk aversion? There are two channels through which risk aversion can affect markdowns: (1) through the reservation wage and (2) through the extent of between-firm competition. While the arrival rate of job offers to employed workers and the productivity distribution itself both can affect the extent of local competition, worker risk aversion has no such direct effect. Intuitively, this can be thought of as capturing the notion that firms cannot insure workers against permanent shocks.

But risk aversion does affect the reservation wage. Unemployed workers are exposed to risk in the job search process—they trade off waiting longer in unemployment and potentially receiving a better job offer with having to spend more time receiving only unemployment benefits for income. Taking a job quickly means workers give up the higher search efficiency of being in unemployment, but risk averse workers value a potentially high income job relatively less.

This means that if the reservation wage is held fixed, worker risk aversion has no effect on wage markdowns. In contrast, even when the reservation wage is constant, worker risk aversion may influence the contract value change that we infer from the empirical wage changes. This means that worker risk aversion influences how we use the firm response to shocks to estimate market power.

4.3 The pass-through of idiosyncratic shocks

How do shocks influence the firm's wage decision? How does the shock-identified labor supply elasticity differ from the cross-sectional labor supply elasticity? What does the firm response to shocks tell us about the extent of market power? To get at these questions, we will study how the firm's choice of the promised utility to offer its workers responds to the shock. From these promised utility changes we will use the labor flow equation to back out the labor response and the promise-keeping constraint to solve for the wage response.

We can differentiate the policy function for the promised utility to offer at time $t+1$, $V^{POL}(s_{t+1}, L_t, V_t)$, with respect to the shock that hit at time 1, s_1. [16]

$$\frac{dV^{POL}(s_{t+1}, L_t, V_t)}{ds_1} = V_s^{POL}\frac{ds_{t+1}}{ds_1} + V_L^{POL}\frac{dL_t}{ds_1} + V_V^{POL}\frac{dV_t^{POL}}{ds_1} \tag{13}$$

There are three channels through which the shock affects the promised utility. The direct channel, V_s^{POL}, captures how the shock boosts profitability. This incentivizes the firm to

[16] note that because we are evaluating at the zero risk steady state, we know that the partial derivatives are constant over time to first order

have a larger workforce, so they increase the promised utility. The strength of this channel over time depends on the persistence of the shock, $\frac{ds_{t+1}}{ds_t}$.

Next, as the firm accumulates (or sheds) workers following the shock, its incentive to offer a higher promised utility changes. When the firm has more workers, offering a higher promised utility will scale up the retention of a larger number of workers. So immediately after the shock, $\frac{dL_0}{ds_1}$ will be zero, but in the following periods the promised utility will respond to changes in the workforce through the V_L^{POL} channel.

Lastly, the promised utility of the firm's incumbent workers impacts its choice of a promised utility the following period when the workers are risk-averse. When the workers are risk neutral, the promise-keeping constraint does not distort the firm's choice of the next period promised utility—utility is transferable so the firm will "pay out" its incumbents whatever is required to allow for the posting decision next period to recruit/retain the profit-maximizing number of workers. But when the workers are risk-averse, it becomes costly to shift the next period promised utility far from today's promised utility. Cutting the promised utility requires boosting the wage and, with risk-averse workers, a sufficiently large decrease might be *impossible* to achieve with a wage payment that satisfies the promise-keeping constraint. Similarly to the V_L^{POL} channel, the V_V^{POL} kicks in over time as the increase to the promised utility in period t, $\frac{dV_t}{ds_1}$, makes it cheaper to keep the promised utility high in V_{t+1} (and vice versa for the negative shock).

This final channel is critical for understanding how insurance impacts the dynamics of wages and labor following a shock. Risk aversion smooths the promised utility response to the shock—making it more costly to adjust the promised utility initially, but then causing the effect to be more persistent as it is costly to sharply return the promised utility to its steady state value even as the shock has dissipated.

The promised utility response determines the wage response and the labor response through the promise-keeping constraint 49 and the labor flow equation 8 respectively. First, we can differentiate the labor flow equation at period $t+1$ with respect to the shock which hit at period 1:

$$\frac{dL_{t+1}}{ds_1} = \left[(1-\lambda_e)(1-\delta) + \lambda_e(1-\delta)F(V_{t+1})\right]\frac{dL_t}{ds_1} + L_V\frac{dV_{t+1}^{POL}}{ds_1} \tag{14}$$

The increase in promised utility in period $t+1$ directly affects the labor that period as it causes the firm to retain more incumbents and recruit more new hires. This change is captured in the partial derivative of the flow labor equation 8 with respect to V_{t+1} and is scaled up by the magnitude of $\frac{dV_{t+1}}{ds_1}$. Additionally, labor accumulates over time as some fraction of the firm's incumbent workers will either not receive an offer ($(1-\lambda_e)$) or receive

an offer from a worse firm ($\lambda_e F(V_{t+1})$) and will not be exogenously separation ($1-\delta$). This means that the entire history of promised utility responses to the shock feeds into the labor response at period $t+1$ through this $\frac{dL_t}{ds_1}$ term which represents the change in incumbent workers.

Similarly, to understand what the path of promised utility response implies about how wages adjust to this productivity shock, we differentiate the promise-keeping constraint 49.

$$\frac{dV_t^{POL}}{ds_1} = u'(w)\frac{dw_t}{ds_1} + \beta\left[(1-\lambda_e)(1-\delta) + \lambda_e(1-\delta)F(V_{t+1})\right]\frac{dV_{t+1}^{POL}}{ds_1} \tag{15}$$

Rearranging this expression we have the wage response, $\frac{dw_t}{ds_1}$ as a function of the current and future promised utility changes following the shock:

$$\implies \frac{dw_t}{ds_1} = \frac{\frac{dV_t^{POL}}{ds_1} - \beta\left[(1-\lambda_e)(1-\delta) + \lambda_e(1-\delta)F(V_{t+1})\right]\frac{dV_{t+1}^{POL}}{ds_1}}{u'(w)} \tag{16}$$

Intuitively, if the current promised utility is higher then, all else equal, this increases the wage the firm will pay out—and the extent of this change depends on the worker's marginal utility $u'(w)$. But, if the firm also increases the promised utility *next* period then this reduces its need to increase wages to fulfill the promise-keeping constraint.

Proposition 3 (Shock-identified labor supply elasticity)**.** *In the first period following an idiosyncratic productivity shock, the shock-identified labor supply elasticity is as follows:*

$$\varepsilon_{shock,Lw} = \frac{L_V u'(w)\frac{w}{L}}{1-\beta q(z)\frac{\frac{dV_2}{ds_1}}{\frac{dV_1}{ds_1}}}$$

Inverting this shock-identified labor supply elasticity fails to recover the wage markdown.

See Appendix A3 for proof

Inverting this shock identified labor supply elasticity does not produce the wage markdown—how do these two values compare? Recall from Proposition 1 the flow labor formulation of the wage markdown is $m(z) = \frac{1-q(z)}{u'(w(z))\frac{w(z)L_V}{L(z)}}(1-\beta q(z))$. The inverse shock-identified labor supply elasticity does not account for labor accumulating over time (the fraction of workers who quit each period which is on the numerator of $m(z)$). This biases us to estimate a wider markdown by inverting the shock-identified—that is, we would overestimate the extent of market power.

Additionally, firms may choose wage payments over time implement the same V change differently when responding to a temporary shock. While the permanently more productive firm pays a constantly higher wage, a firm who is temporarily more productive may front-load

its wage payments during the periods that it is most productive. If the wage is decreasing over time, this would make $\frac{\frac{dV_2}{ds_1}}{\frac{dV_1}{ds_1}}$ less than 1. This would also lead us to overestimate the extent of market power. Intuitively, the same wage change would produce a larger labor response if it is permanent and so estimating a labor supply elasticity using a more temporary shock can make labor appear more inelastic.

It is not guaranteed that inverting this labor supply elasticity (which is estimated immediately following the shock) will imply too much market power. As we have discussed, as the firm increases its promised utility and accumulates more workers the dynamic channels—V_L^{POL} and V_V^{POL}—will kick in and this can cause $\frac{dV_2}{ds_1} > \frac{dV_1}{ds_1}$. Especially if the shock is very persistent (so the direct channel V_s^{POL} remains strong) and the workers are risk averse (so the V_V^{POL} is very relevant), it is possible for the wage to *increase* in the years following the shock. If the wage increases strongly in the years following the shock, then it is possible that using the first period wage change would result in a shock-identified labor supply elasticity which is more elastic than the cross-sectional elasticity.

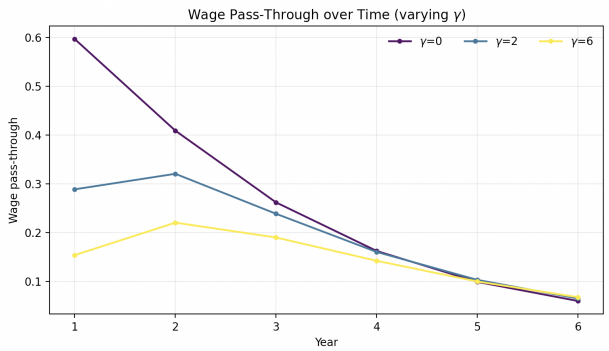
What if we calculate the labor supply elasticity several periods after the shock or take an average over a longer window? Empirically, we know that the labor supply elasticity t periods after the shock, $\frac{\frac{dL_{t+1}}{ds_1}w}{\frac{dw_{t+1}}{ds_1}L}$, is more elastic over time (i.e., as t is larger). As more time goes on, the labor response accumulates and, as the shock fades, the firm lowers the wage. These both can reduce the measured labor supply elasticity. At an extreme, if the firm maintained a high wage for three periods and then cut the wage back to baseline, if we estimated a labor supply elasticity at period four we would deem the firm's labor supply elasticity was infinitely elastic.

The shock-identified labor supply elasticity is a complex, dynamic object and it does not directly map onto the wage markdown. But we can leverage the dynamic wage and labor response to shocks to estimate wage markdowns. By modeling the firm's choice of the optimal contract we can understand how wages vary beyond the years we observe and by modeling the worker's behavior we can use their risk aversion, the arrival rate of jobs, and other relevant parameters to map these wages changes onto an implied promised utility change. This allows us to infer the wage markdown.

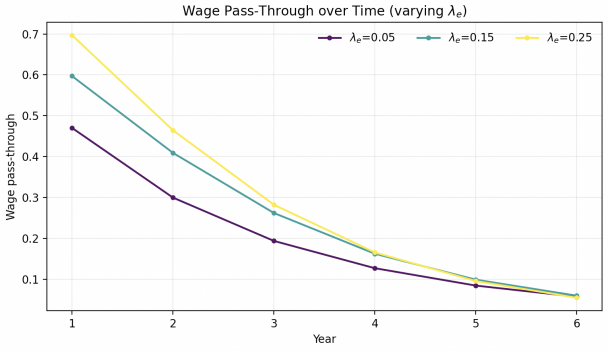
4.4 Identification intuition

We have characterized how firms contract over idiosyncratic shocks. Now, we will discuss how our model can leverage the firm's wage and labor response to shocks to estimate the extent of monopsony power and the degree of wage insurance. First, we need to take the empirical wage changes, infer future unobserved wage changes, and convert this to a change

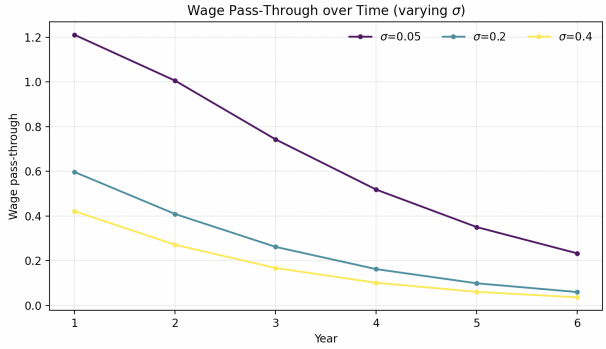
Figure 3: Wage pass-through over years under different parameters



(a) Varying γ



(b) Varying λ_e



(c) Varying σ

Notes: We take the average across all active firms and plot the response from years 1 to years 6 following the shock. Each panel varies one of the parameters of interest and other parameters will take the following values: $\delta = 0.016, \beta = 0.996, \rho = 0.95, \mu = 2, z_{min} = 1, w_R = 1.8, \lambda_u = 0.4, \lambda_e = 0.15, \sigma = 0.2, \gamma = 0$.

in promised utility at each point in time. We accomplish this by solving the firm's contracting problem.

To deepen the intuition for what determines how firms respond to shocks, we simulate the model. Here, we parameterize the worker's utility function as constant relative risk aversion with coefficient γ, the firm permanent productivity distribution as being log-normal with mean μ and standard deviation σ, and we hold the reservation wage constant so changing the extent of risk aversion has no effect on market power.

Search frictions and dispersion in the productivity dispersion both influence firm's market power and in Figure 3, we show that these parameters also affect the wage pass-through. Increasing search frictions, increasing productivity dispersion—and reducing worker risk aversion—all broadly result in a larger wage pass-through. But notably, increasing worker risk aversion results in a more persistent, flatter wage response. This is because when there are costs to adjusting incumbent worker's promised utility, this induces larger promised util-

ity increases in the future, as it is costly to reduce pay back down as the shock is dissipating. This means that the early wage increases are muted and the overall wage effect is flatter over time when worker risk aversion is higher. In other words, insurance induces a more persistent wage response.

Modeling the firm's problem allows us to understand how the value of the contract is changing in the years following the shock. Then, we can look at the labor response that these promised utility changes induce over time. The rate at which workers receive job offers and the productivity distribution both influence how employment adjusts to changes in the value of the contract. We search for the set of parameters where the solution to the firm's problem matches the observed wage changes and where the solution to the worker's problem matches the observed labor changes.

5 Estimating the Model

We have established that worker risk aversion induces a smaller, more persistent wage response. This allows us to use the observed wage changes and infer the contract's value change. Then, the dynamics of how labor responds to these promised utility changes allows us to estimate the wage markdown. We now turn to the data to quantify our model and estimate the prevalence of insurance and market power in the U.S. labor market.

Each period corresponds to one month and we set $\beta = 0.996$ to match an annual interest rate of 4%. We also externally calibrate $\delta = 0.016$. To match the persistence of the productivity shock, we set $\rho = 0.953$[17]. The productivity distribution is log-normal with mean μ and standard deviation σ. It is truncated so that the minimum productivity value is z_{min}. This parameter will influence the share of firms which are below the reservation wage (and therefore inactive) and allow us to better match the unemployment rate. The utility function is CRRA with coefficient γ. Lastly, we have to estimate the arrival rate of job offers to the unemployed, λ_u, and to the employed, λ_e.

We will match these six parameters by using simulated method of moments. In addition to the dynamic response of wages and labor to the productivity shock, we will match the standard deviation of revenue, wages, and employment as well that employment-to-employment transition rate (EE rate) and the unemployment rate. Together, we are estimating these six parameters by matching seventeen moments.

In our model, large firms and small firms behave very differently. Large firms are further up the productivity distribution where the odds of a worker getting an offer better than

[17]We estimate an annual persistence of 0.56 and assuming an AR(1) shock process, this implies that the monthly persistence is 0.953

their current job are smaller—leading them to face a more inelastic labor supply and mark down wages more. Our wage and labor pass-through estimates come from Compustat firms which are substantially larger and more productive than the median American firm. To take this concern seriously, we generate our model wage and labor pass-through measures on the top of the productivity distribution employing 30% of the workforce. This roughly matches the share of workers at publicly traded firms in the United States. Our standard deviation moments were made on the LBD-LEHD sample which is fairly representative of the economy and does not have this concern. Similarly, the unemployment rate and the EE rate are economy-wide moments.

We use simulated method of moments and minimize the loss function

$$L = \sum_i \left(\frac{m_i^{\text{model}} - m_i^{\text{data}}}{m_i^{\text{data}}}\right)^2 \tag{17}$$

where m_i^{model} and m_i^{data} denote the model-implied and empirical values of moment i, respectively.

5.1 Baseline Estimation

Figure 4 plots our simulated wage and labor pass-throughs over time against the empirical analogs. Table 2 reports the fit for cross-sectional moments. Our model fit is broadly good. We match the small wage pass-through, the relatively flat wage response, the accumulation of labor over time, and the cross-sectional dispersion. Notably, our model fails to match the first period wage response. This is because our model features perfect firm commitment; because there is such a large, persistent wage increase in the subsequent years, the firm does not initially raise wages but is successfully able to recruit/retain additional workers. Relaxing the assumption of perfect commitment would likely facilitate our ability to match this moment.

Our estimate of γ is on the large side, but the literature estimating the coefficient of relative risk aversion produces a wide range of values—Gandelman and Hernández-Murillo (2015) note that these estimates range from 0.2 to over 10. This suggests workers strongly value stable pay: a worker would accept approximately 3% lower pay to avoid a 10% symmetric wage gamble[18]. We find a relatively high level of λ_e and a relatively small σ. This suggests that firms face considerable competition—both due to their workers frequently receiving outside offers and because they are relatively close, in terms of productivity, to their competitor firms.

[18]The Arrow-Pratt risk premium approximation is $\pi \approx \frac{1}{2}\gamma \text{Var}(w) = \frac{1}{2} \cdot 5.97 \cdot 0.01 = 2.99$ for a 10% gamble

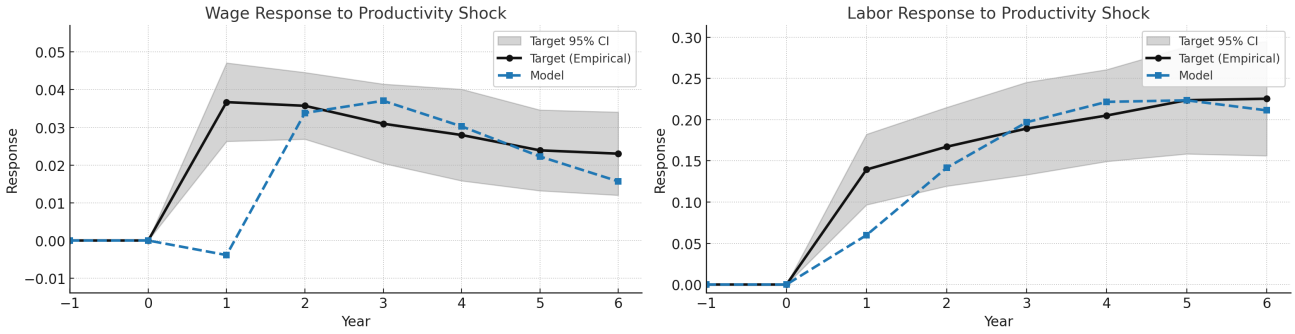
What do our parameter estimates imply about the level of labor market power? Our estimates of σ and λ_e imply that firms exert a relatively small degree of monopsony power. Indeed, we estimate that the average worker's wage is marked down by 8.3%. This is a modest markdown relatively to the literature which typically finds markdowns between 15% and 50% (Azar and Marinescu, 2024).

Table 2: Aggregate Moments and Model Fit

Moment	Target	Model
ee	0.022000	0.024856
sd[log L]	1.309000	1.226063
sd[log Rev]	1.513000	1.289961
sd[log w]	0.442000	0.432393
u	0.043000	0.042354

Notes: "Target" are data moments; "Model" are simulated moments at baseline parameters; dispersion moments from LBD–LEHD, EE and u are economy-wide. Target moments are from FSRDC Project Number 3060 (CBDRB-FY25-P3060-R12493)

Figure 4: Baseline impulse responses of wages and labor



Wage Response Labor Response

Notes: Empirical impulse response to value-added per worker shocks from Eq. (1) and model-simulated paths; horizons are years; Compustat-LEHD-LBD sample; 95% confidence intervals if shown. Target moments are from FSRDC Project Number 3060 (CBDRB-FY25-P3060-R12493).

5.2 Alternate Specifications

Our model features sluggish labor adjustment due to search frictions and the firm provision of wage insurance. How would our estimate of the wage markdown differ if we adopted an approach which did not account for these dynamics? We can compare our estimate of

Table 3: Estimated structural parameters

Parameter	Estimate	Description
γ	5.973	Worker risk aversion
λ_e	0.145	Employed worker job arrival rate
λ_u	0.362	Unemployed worker job arrival rate
μ	0.683	Mean of productivity distribution
σ	0.057	Standard deviation of the productivity distribution
$z_{\min}$	0.682	Truncation point of productivity distribution

Notes: SMM point estimates; $\beta = 0.996$, $\rho = 0.953$, and $\delta = 0.016$ fixed (monthly); CRRA utility; productivity is log-normal truncated at $z_{\min}$.

an 8.3% markdown to what we would get if we inverted our shock-identified labor supply elasticities. If we inverted the year 1 labor supply elasticity we would estimate a 26% markdown, while if we inverted the year 6 labor supply elasticity we would estimate a 10% markdown. Alternatively, if we pooled the average response across the six years, we would estimate a 16% wage markdown. These markdown values are standard for the literature and suggest that modeling dynamics can hugely affect how we interpret shock-identified labor supply elasticities as evidence of market power.

How important is the provision of insurance for matching wage and employment dynamics? To answer this, we shut down the insurance channel in our baseline model by fixing $\gamma = 0$ and estimate this risk-neutral, wage contracting monopsony model.

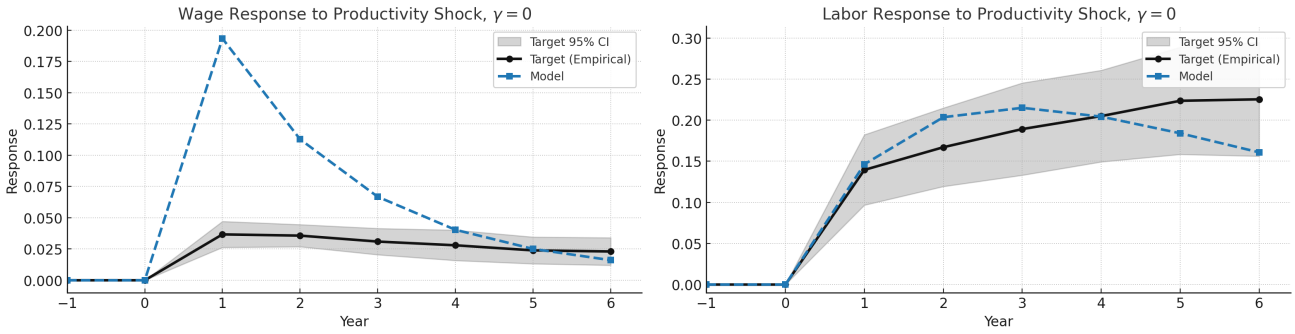
Table 4: Aggregate Moments and Model Fit

Moment	Target	Model
ee	0.022000	0.033129
sd[log L]	1.309000	1.388853
sd[log Rev]	1.513000	1.547781
sd[log w]	0.442000	0.373331
u	0.043000	0.042553

Notes: "Target" are data moments; "Model" are simulated moments at baseline parameters; dispersion moments from LBD–LEHD, EE and u are economy-wide. Target moments are from FSRDC Project Number 3060 (CBDRB-FY25-P3060-R12493).

Figure 5 shows the wage and labor response to the estimated risk-neutral wage contract model. Without insurance, the model is unable to match the small wage pass-through. Additionally, the wage response is far too steep—initially, the model produces a wage increase that is 600% larger than the one year pass-through but by year six the wage response is below the empirical moment. This highlights how insurance is key to generating a small, flat wage

Figure 5: Wage and labor fit with wage contracts and risk neutral workers



Notes: Empirical impulse response to value-added per worker shocks from Eq. (1) and risk-neutral model-simulated paths; horizons are years; Compustat-LEHD-LBD sample; 95% CIs if shown. Target moments are from FSRDC Project Number 3060 (CBDRB-FY25-P3060-R12493).

Table 5: Estimated structural parameters ($\gamma = 0$)

Parameter	**Estimate**	**Description**
γ	Fixed: 0	Worker risk aversion
λ_e	0.280	Employed worker job arrival rate
λ_u	0.360	Unemployed worker job arrival rate
μ	2.298	Mean of productivity distribution
σ	0.166	Standard deviation of productivity distribution
$z_{\min}$	2.608	Truncation point of productivity distribution

Notes: SMM point estimates for model with risk-neutral workers and dynamic wage contracts; $\beta = 0.996$, $\rho = 0.953$, and $\delta = 0.016$ fixed (monthly); CRRA utility; productivity is lognormal truncated at $z_{\min}$.

response to shocks.

Notably, the model almost exclusively struggles to fit the wage response. It roughly matches the aggregate moments, except for the employment-to-employment transition rate which it overshoots, and it matches the labor response.

We can also compare the value of the loss function between the estimation when we model insurance and this version where we set $\gamma = 0$. The error is 0.744 in our baseline, insurance wage contract model and the error is 5.654 in the risk-neutral wage contract model—over 7 times worse.

Clearly, the risk-neutral wage contract model is unable to match the observed wage and labor dynamics. To understand the implications of worker risk aversion for our estimate of labor market power, we will re-estimate the model to match a more limited set of moments. Specifically, we will target the same aggregate moments and, rather than targeting the full path of wages and labor to an idiosyncratic shock, we will match the average labor supply elasticity over the six year horizon. We define the average labor supply elasticity as the

average labor response divided by the average wage response. Empirically, the value of the average labor supply elasticity is 6.44.

We display the estimated parameters in Table 6 and the model fit in Table 7. The parameters imply an average wage markdown of 11%—between the insurance contract model wage markdown (8.3%) and the static model wage markdown estimate that inverts this targeted average labor supply elasticity (16%). Risk aversion causes workers to value temporary wage changes by less than permanent wage changes with the same present dollar value. This is because risk aversion induces an intertemporal smoothing motive where workers would prefer the same dollar change spread out over more years. Assuming that workers are risk neutral would result in overestimating the utility value change associated with a temporary wage change, leading the researcher to infer that the labor response is more inelastic than it truly is. This is reflected in the fact that we estimate wider wage markdowns when we model workers as risk-neutral than when we estimate our full insurance wage contracting model.

Table 6: Estimated structural parameters

Parameter	**Estimate**	**Description**
γ	**Fixed:** 0	Worker risk aversion
μ	12.379	Mean of log productivity
σ	0.060	Std. dev. of log productivity
λ_e	0.222	Employed worker job arrival rate
λ_u	0.302	Unemployed worker job arrival rate
$z_{\min}$	19.980	Truncation point of productivity distribution

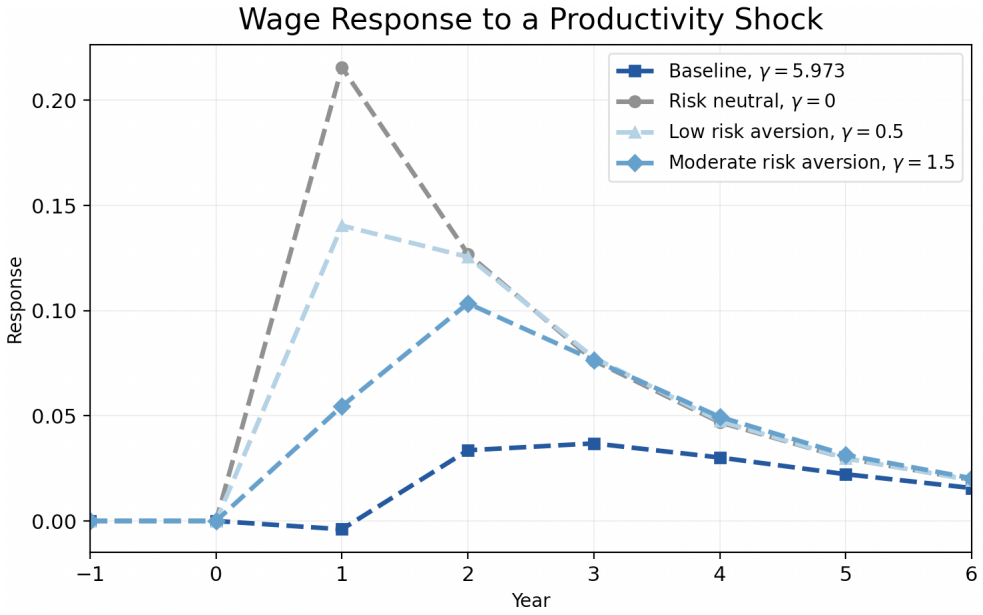
Notes: This table presents the estimated parameters. Parameters are estimated by simulated method of moments.

Table 7: Aggregate moments and model fit

Moment	Target	Model
Average ε_{Lw}	6.448320	6.392164
ee	0.022000	0.030103
sd[log L]	1.309000	1.316793
sd[log Rev]	1.513000	1.374384
sd[log w]	0.442000	0.382863
u	0.043000	0.050342

Notes: "Target" are data moments; "Model" are simulated moments at the estimated parameters. Target moments are from FSRDC Project Number 3060 (CBDRB-FY25-P3060-R12493).

Figure 6



Notes: Wage response to a temporary ($\rho = 0.953$) 1% idiosyncratic shock among the most productive firms employing 30% of workers (as in baseline estimation); the extent of worker risk aversion is varied and all other parameter values are those reported in Table 3.

5.3 Insurance and the Wage Pass-Through

We have shown that the wage patterns we observe are consistent with the firm provision of insurance. Our baseline estimation finds that workers are considerably risk-averse, and when we re-estimate the model under the assumption that workers are risk-neutral we dramatically overshoot the wage pass-through. Now, we ask a related question: how does the firm provision of insurance affect the observed wage pass-through? That is, if workers were less risk averse but we retain the same other structural parameters, how would the simulated wage response to an idiosyncratic shock change?

Figure 6 shows that the firm provision of insurance results in a strikingly smaller and more persistent wage pass-through. If workers were risk neutral, firms would respond to the 1% temporary, idiosyncratic shock by increasing wages 0.21% in the first year and quickly reducing pay closer to baseline in the subsequent years. Across the six year following the shock, the average wage increase would be 0.086% without insurance—much larger than the 0.03% average increase we observe empirically. The wage responses when workers are only somewhat risk averse ($\gamma = 0.5$) or are moderately risk averse ($\gamma = 1.5$) display a similar pattern—worker risk aversion causes firms to offer wage insurance which has a profound effect on the wage pass-through.

How firms adjust wages following shocks is very dependent on the extent of worker risk

aversion. This parallels Kline (2025), which notes that the wage pass-through is sensitive to the functional form of the labor supply curve firms.[19] Empirical estimates of how wages respond to firm performance are important evidence of monopsony power—if labor markets were perfectly competitive, there would be one market wage. But can these wage pass-through measures be used to infer the level of wage-setting power? We have demonstrated that this requires carefully modeling the firm provision of insurance, as well as the persistence of shocks and the dynamics of firm adjustment.

6 The Constrained Efficient Allocation

We have shown that the firm's role as an insurer for the worker results in a small and flatter wage response to productivity shocks. This causes firms to hoard labor, adjusting the size of their workforce by less following shocks. We estimate that workers are considerably risk-averse and that firms' behavior reflects this fact. Now, we turn to ask the efficiency implications of this firm provision of insurance. We will show that the provision of wage insurance distorts the job ladder, leading to fewer job-to-job transitions than would be socially optimal. This means that understanding the extent of wage insurance is not just important for our measurement of market power but also has direct policy implications—policies which reduce the need for firms to insure workers will improve labor market dynamism.

Consider a social planner who is constrained by search frictions but can perform transfers. This planner cannot influence how many job offers a worker receives—they can only decide whether the worker should make a transition. Notably, since the planner can perform transfers this problem is not influenced by worker risk aversion. The planner's solution may differ from the decentralized equilibrium in two ways—in which job offers unemployed workers accept and in which job-to-job moves employed workers make.

First, we will briefly describe the unemployed worker's decision. When the search efficiency of employed and unemployed workers is not equal, unemployed workers who are deciding whether to accept or reject a job offer will trade off staying in unemployment for longer with the opportunity to potentially receive a better job offer more quickly. Wages are marked down relative to productivity, so the worker's choice may not align with the planner's solution (as the planner will maximize total value). When workers are risk-averse, they will be quicker to accept job offers. This suggests that the extent of worker risk aversion is important for the optimal unemployment insurance as the level of b influences the worker's decision over which jobs to accept. This point parallels Acemoglu and Shimer (2000), which

[19]Notably, the wage pass-through can either rise or fall with the labor supply elasticity depending on the firm labor supply curve assumed

shows how unemployment insurance can increase productivity by allowing workers to direct their search to riskier (but more productive) jobs.

Now, we characterize the optimal job-to-job transitions. This will allow us to ask whether firms respond optimally to shocks. We will show that in response to idiosyncratic shocks, firms initially engage in too much labor hoarding and that this is exacerbated by the presence of insurance. Using our parameter estimates from Section 5.1 we will quantify the planner's response to shocks and compare it to the decentralized equilibrium with risk-averse workers *and* the decentralized equilibrium with risk neutral workers.

The planner will move the worker if the social value of the new job, $W(z_1, s_1)$, exceeds the social value of their current job, $W(z_0, s_0)$ where these value functions are defined as follows:

$$\begin{aligned} W(z,s) = (1+s)z + \beta(1-\lambda_e)(1-\delta)\,\mathbb{E}[W(z,s')] \\ + \beta\lambda_e(1-\delta)\,\mathbb{E}\left[\int \max(W(z,s'),W_0)\,dF(W_0)\right] + \delta\beta W_U \end{aligned} \tag{18}$$

We want to know how idiosyncratic risk affects the job ladder. To do this, we will leverage the same first order approximation approach used to characterize the decentralized equilibrium. This will allow us to solve for the efficient firm-level labor response to an idiosyncratic shock and compare that to the decentralized equilibrium.

Since the aggregate distributions are unaffected by the small amount of risk, we know that $F(W(z,0)) = J(z)$. This is because, without risk, firms with a higher permanent productivity z will always have a higher social value $W(z,0)$. Differentiating 18 with respect to the shock s and evaluating at $s = 0$ we can rearrange to show:

$$W_s = \frac{z}{1-\beta(1-\delta)\rho[1-\lambda_e+\lambda_e J(z)]} = z\Phi_0 \tag{19}$$

where ρ is the persistence of the shock. Intuitively, the value of having a worker at a firm with productivity z increases in response to the shock and the extent of this benefit depends on a present-value multiplier style term which reflects the persistence of shock and the odds of the worker leaving.

Labor evolves according to:

$$L^*(z,s_t) = L^*(z,s_{t-1})(1-\delta)(1-\lambda_e(1-F(W(z,s_t)))+\lambda_u u+\lambda_e(1-u)(1-\delta)G(W(z,s_t)) \tag{20}$$

Differentiating this labor flow expression with respect to the shock s_1, we can show that the planner's choice of labor at firm with productivity z will respond by $\frac{dL^*(z,s_t)}{ds_1}$ in period t after

the shock in period 1:

$$\begin{aligned}\frac{dL^*(z,s_t)}{ds_1} = \frac{dL^*(z,s_{t-1})}{ds_1}(1-\delta)\big(1-\lambda_e(1-J(z))\big) \\ + L^*(1-\delta)\lambda_e F'(W)W_s + \lambda_e(1-u)(1-\delta)G'(W)W_s\end{aligned} \tag{21}$$

Proposition 4 (Planner response to permanent shock)**.** *Following a permanent shock, at each point in time the planner's labor response is larger than the decentralized labor response, but the two converge to the same value over time.*

$$\frac{dL^*(z,s_t)}{ds_1} > \frac{dL(z,s_t)}{ds_1}$$

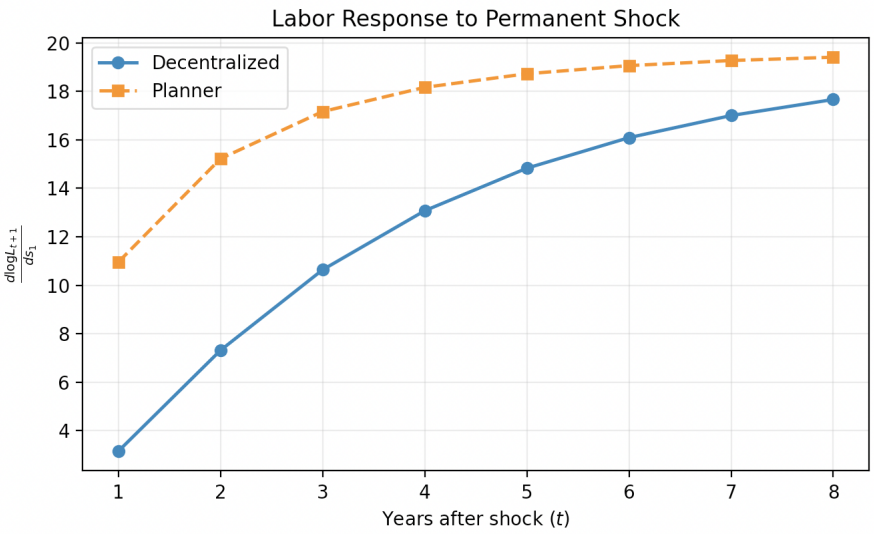
$$\lim_{t\to\infty}\frac{dL^*(z,s_t)}{ds_1} = \frac{dL(z,s_t)}{ds_1} = \frac{2(1-\delta)\lambda_e J'(z)zL}{1-(1-\delta)[1-\lambda_e(1-J(z))]}$$

Proposition 4 reveals that the planner responds more aggressively to permanent shocks than the decentralized equilibrium. The planner ranks firms by the permanent productivity differences (since there is only a small amount of idiosyncratic risk) so if the firm receives a positive shock it will be immediately moved to the new position on the job ladder—where it will stay indefinitely as it continues to accumulate more workers and converges to its new, larger employment level. But in the decentralized equilibrium, firms move slowly up the job ladder following a positive permanent shock. As they accumulate workers this boosts the retention motive, so they increase the promised utility and this is reinforced by the fact the incumbents' promised utility is a state variable—so when today's promised utility is higher it is even more desirable to offer a higher promised utility tomorrow. In the long-run they converge to the same place on the job ladder as the planner's solution, but along the transition path they are slower to accumulate workers.

This highlights an important implication of idiosyncratic shocks for the efficiency properties of models in the style of Burdett and Mortensen (1998). While job-to-job transitions when there are no firm-level shocks *and* the job-to-job transitions with wage contracting over aggregate shocks (see Moscarini and Postel-Vinay (2013)) are constrained efficient, the job ladder is no longer efficient in the presence of idiosyncratic risk.

With this equation 21 in hand, we can simulate the planner labor response and compare it to the decentralized labor response. This allows us to quantify the importance of Proposition 4 given the parameters that we estimated in Section 5. In Figure 7, we plot the decentralized equilibrium and planner response to a permanent productivity shock. As expected, we see that the planner responds far more aggressively to the shock and firms are too slow to accumulate workers. This effect is quantitatively large—in the first year the planner has

Figure 7



Notes: Employment response to a permanent ($\rho = 1$) 1% idiosyncratic shock; decentralized equilibrium vs constrained-efficient planner; parameter values are those reported in Table 3 horizons are years.

a three times larger labor response than the decentralized equilibrium. Even eight years following the shock there is still a meaningful gap between the planner and decentralized responses.

How does the decentralized response to a *temporary* shock compare to the constrained efficient benchmark? While Propostion 4 is specific to permanent shocks, many of the same forces are present in how firms adjust to the temporary shocks. In Figure 8, we plot the decentralized response to a temporary productivity shock that we estimated in Section 5.1.[20] Then, we take these same underlying parameters and plot the planner solution.

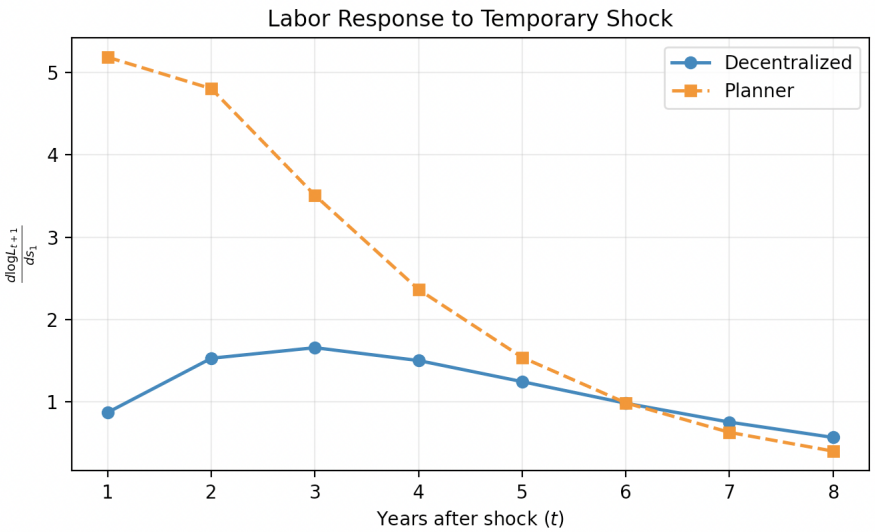
In the first years following the shock, the planner response is strikingly larger. In response to a 1% productivity shock[21], the planner moves more workers to the firm, causing it to grow by 5% one year after impact. In contrast, the decentralized equilibrium features a far more modest labor response, with employment increasing by less than 1%. This suggests that the economy features too much labor hoarding in response to temporary shocks relative to the planner's solution.

Notably, by year 7 this pattern reverses with firms holding onto more workers than the planner would choose. The firm provision of insurance results in a stable wage response which

[20]Note that the labor response plotted is different than what we plot in Figure 5.1. This is because we previously matching the response of firms in the top of the productivity distribution (representing 30% of employment) to the empirical response of Compustat firms. For this exercise, we are plotting the response of all firms.

[21]with persistence $\rho = 0.953$

Figure 8



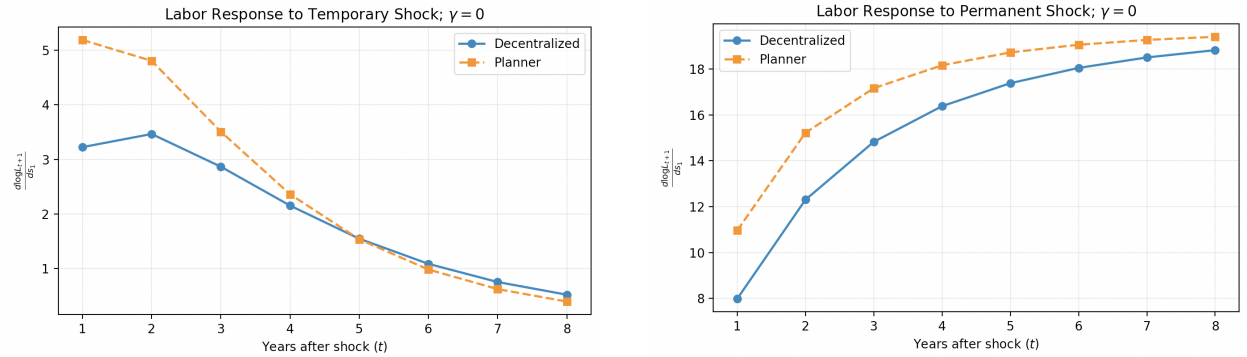
Notes: Employment response to a temporary ($\rho = 0.953$) 1% idiosyncratic shock; decentralized equilibrium vs constrained-efficient planner; parameter values are those reported in Table 3 horizons are years.

causes labor to stay higher for longer. The shock has largely faded away, so the planner would prefer the workers be moved to a more productive firm, but insurance prevents this.

How important is the provision of insurance in distorting the job ladder? Now, we replicate the earlier exercises but change $\gamma = 0$ when solving the decentralized equilibrium. In Figure 9 we plot this risk-neutral decentralized labor response. Directionally, the patterns look similar—firms adjust slower than the planner wants. But the magnitude of this gap is very different—one year after the shock, the response to the temporary shock and the permanent shock when workers are risk neutral is more than *double* than when workers are as risk-averse as we estimate. By year 4, the risk neutral decentralized response is qualitatively similar to the planner and, similarly, the response to the permanent shock converges much quicker to the planner when workers are risk neutral.

Notably, we have shown that a more aggressive labor response is efficient when the planner can perform transfers. In practice such transfers may be unrealistic. Firms offer wage insurance because workers demand it—but it comes at the cost of reducing job-to-job transitions following shocks. This suggests that policies which reduce the workers' demand for wage insurance—expanding workers' access to credit, social insurance, encouraging household savings—would cause firms to adjust labor more aggressively in response to shocks. And by allowing workers to transition to more productive jobs, this would boost output.

Figure 9: Risk neutrality brings the decentralized equilibrium close to the planner solution



(a) Temporary shock

(b) Permanent shock

Notes: Employment response to a temporary ($\rho = 0.953$) and permanent ($\rho = 1$) 1% idiosyncratic shock with risk-neutral workers $\gamma = 0$; decentralized equilibrium vs constrained-efficient planner; other parameter values are those reported in Table 3 horizons are years.

7 Conclusion

How do we interpret empirical estimates of labor market power that are identified off shocks? We write a random search model where firms offer wage contracts over idiosyncratic shocks to risk-averse workers. We estimate that the average worker's wage is marked down 8.3% which is a narrower markdown than the literature and smaller than what we find when estimate markdowns by inverting the shock-identified labor supply elasticity. The firm provision of insurance is critically important to explaining how firms respond to shocks. Insurance leads firms to hoard labor beyond what is socially efficient and reducing worker risk aversion can bring the firm response to shocks closer to the planner's solution.

References

Abowd, John M, Paul Lengermann, and Kevin L McKinney (2003), "The measurement of human capital in the us economy."

Acemoglu, Daron and Robert Shimer (2000), "Productivity gains from unemployment insurance." *European economic review*, 44, 1195–1224.

Agostinelli, Francesco, Domenico Ferraro, Giuseppe Sorrenti, and Leonard Treuren (2025), "Employment relationships, wage setting, and labor market power." Technical report, National Bureau of Economic Research.

Amodio, Francesco and Nicolas De Roux (2024), "Measuring labor market power in developing countries: Evidence from colombian plants." *Journal of Labor Economics*, 42, 949–977.

Azar, José and Ioana Marinescu (2024), "Monopsony power in the labor market." In *Handbook of Labor Economics*, volume 5, 761–827, Elsevier.

Azar, José A, Steven T Berry, and Ioana Marinescu (2022), "Estimating labor market power." Technical report, National Bureau of Economic Research.

Azariadis, Costas (1975), "Implicit contracts and underemployment equilibria." *Journal of political economy*, 83, 1183–1202.

Baily, Martin Neil (1974), "Wages and employment under uncertain demand." *The Review of Economic Studies*, 41, 37–50.

Balke, Neele and Thibaut Lamadon (2022), "Productivity shocks, long-term contracts, and earnings dynamics." *American Economic Review*, 112, 2139–2177.

Berger, David, Kyle Herkenhoff, and Simon Mongey (2022), "Labor market power." *American Economic Review*, 112, 1147–1193.

Bewley, Truman F (1999), *Why wages don't fall during a recession*. Harvard university press.

Bloesch, Justin, Birthe Larsen, and Anders Yding (2024), "Monopsony with recruiting." *Available at SSRN*.

Boal, William M and Michael R Ransom (1997), "Monopsony in the labor market." *Journal of economic literature*, 35, 86–112.

Bontemps, Christian, Jean-Marc Robin, and Gerard J Van den Berg (2000), "Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation." *International Economic Review*, 41, 305–358.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani (2018), "The morale effects of pay inequality." *The Quarterly Journal of Economics*, 133, 611–663.

Burdett, Ken and Melvyn Coles (2003), "Equilibrium wage-tenure contracts." *Econometrica*, 71, 1377–1404.

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

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin (2006), "Wage bargaining with on-the-job search: Theory and evidence." *Econometrica*, 74, 323–364.

Caldwell, Sydnee and Emily Oehlsen (2023), "Gender, outside options, and labor supply: Experimental evidence from the gig economy." *University of California, Berkeley Working Paper*.

Campello, Murillo, John R Graham, and Campbell R Harvey (2010), "The real effects of financial constraints: Evidence from a financial crisis." *Journal of financial Economics*, 97, 470–487.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline (2018), "Firms and labor market inequality: Evidence and some theory." *Journal of Labor Economics*, 36, S13–S70.

Card, David, Ana Rute Cardoso, and Patrick Kline (2016), "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women." *The Quarterly journal of economics*, 131, 633–686.

Chan, Mons, Sergio Salgado, and Ming Xu (2020), "Heterogeneous passthrough from tfp to wages." *Available at SSRN 3538503*.

Crouzet, Nicolas and Neil R Mehrotra (2020), "Small and large firms over the business cycle." *American Economic Review*, 110, 3549–3601.

Donangelo, Andres, Francois Gourio, Matthias Kehrig, and Miguel Palacios (2019), "The cross-section of labor leverage and equity returns." *Journal of Financial Economics*, 132, 497–518.

Ellul, Andrew, Marco Pagano, and Fabiano Schivardi (2018), "Employment and wage insurance within firms: Worldwide evidence." *The Review of Financial Studies*, 31, 1298–1340.

Friedrich, Benjamin, Lisa Laun, Costas Meghir, and Luigi Pistaferri (2019), "Earnings dynamics and firm-level shocks." Technical report, National Bureau of Economic Research.

Fujita, Shigeru, Giuseppe Moscarini, and Fabien Postel-Vinay (2024), "Measuring employer-to-employer reallocation." *American Economic Journal: Macroeconomics*, 16, 1–51.

Fukui, Masao (2020), "A theory of wage rigidity and unemployment fluctuations with on-the-job search." *Job Market Paper, Massachusetts Institute of Technology*.

Gandelman, Nestor and Rubén Hernández-Murillo (2015), "Risk aversion at the country level." *Available at SSRN 2646134*.

Garin, Andrew and Filipe Silvério (2023), "How responsive are wages to firm-specific changes in labour demand? evidence from idiosyncratic export demand shocks." *Review of Economic Studies*, rdad069.

Giroud, Xavier and Holger M Mueller (2017), "Firm leverage, consumer demand, and employment losses during the great recession." *The Quarterly Journal of Economics*, 132, 271–316.

Gottfries, Axel and Gregor Jarosch (2023), *Dynamic monopsony with large firms and noncompetes*. 31965, National Bureau of Economic Research.

Gouin-Bonenfant, Émilien (2022), "Productivity dispersion, between-firm competition, and the labor share." *Econometrica*, 90, 2755–2793.

Guiso, Luigi, Luigi Pistaferri, and Fabiano Schivardi (2005), "Insurance within the firm." *Journal of Political Economy*, 113, 1054–1087.

Hadlock, Charles J and Joshua R Pierce (2010), "New evidence on measuring financial constraints: Moving beyond the kz index." *The review of financial studies*, 23, 1909–1940.

Harris, Milton and Bengt Holmstrom (1982), "A theory of wage dynamics." *The Review of Economic Studies*, 49, 315–333.

Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin (2024), "Granular search, market structure, and wages." *Review of Economic Studies*, rdae004.

Jordà, Òscar (2005), "Estimation and inference of impulse responses by local projections." *American economic review*, 95, 161–182.

Kalemli-Özcan, Şebnem, Luc Laeven, and David Moreno (2022), "Debt overhang, rollover risk, and corporate investment: Evidence from the european crisis." *Journal of the European Economic Association*, 20, 2353–2395.

Kaplan, Greg, Giovanni L Violante, and Justin Weidner (2014), "The wealthy hand-to-mouth." Technical report, National Bureau of Economic Research.

Kline, Patrick (2025), "Labor market monopsony: Fundamentals and frontiers." *Handbook of Labor Economics*, 6, 655–728.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar (2019), "Who profits from patents? rent-sharing at innovative firms." *The quarterly journal of economics*, 134, 1343–1404.

Knight, Frank Hyneman (1921), *Risk, uncertainty and profit*, volume 31. Houghton Mifflin.

Kroft, Kory, Yao Luo, Magne Mogstad, and Bradley Setzler (2025), "Imperfect competition and rents in labor and product markets: The case of the construction industry." *American Economic Review*, 115, 2926–2969.

Kudlyak, Marianna (2014), "The cyclicality of the user cost of labor." *Journal of Monetary Economics*, 68, 53–67.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler (2022), "Imperfect competition, compensating differentials, and rent sharing in the us labor market." *American Economic Review*, 112, 169–212.

Manning, Alan (2003), *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press.

Manning, Alan (2021), "Monopsony in labor markets: A review." *ILR Review*, 74, 3–26.

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

Moscarini, Giuseppe and Fabien Postel-Vinay (2013), "Stochastic search equilibrium." *Review of Economic Studies*, 80, 1545–1581.

Moscarini, Giuseppe and Fabien Postel-Vinay (2016), "Wage posting and business cycles: A quantitative exploration." *Review of Economic Dynamics*, 19, 135–160.

Rajan, Raghuram G and Luigi Zingales (1995), "What do we know about capital structure? some evidence from international data." *The journal of Finance*, 50, 1421–1460.

Rudanko, Leena (2009), "Labor market dynamics under long-term wage contracting." *Journal of monetary Economics*, 56, 170–183.

Rudanko, Leena (2011), "Aggregate and idiosyncratic risk in a frictional labor market." *American Economic Review*, 101, 2823–2843.

Seegmiller, Bryan (2021), "Valuing labor market power: The role of productivity advantages." Technical report, Working paper.

Sharpe, Steven A (1994), "Financial market imperfections, firm leverage, and the cyclicality of employment." *The American Economic Review*, 84, 1060–1074.

Sorkin, Isaac (2018), "Ranking firms using revealed preference." *The quarterly journal of economics*, 133, 1331–1393.

Souchier, Martin (2022), "The pass-through of productivity shocks to wages and the cyclical competition for workers." Technical report, Working Paper.

Souchier, Martin (2025), "Insurance inside and outside the firm." *Available at SSRN 5137109*.

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

Tian, Shaonan, Yan Yu, and Hui Guo (2015), "Variable selection and corporate bankruptcy forecasts." *Journal of Banking & Finance*, 52, 89–100.

Van Reenen, John (1996), "The creation and capture of rents: wages and innovation in a panel of uk companies." *The quarterly journal of economics*, 111, 195–226.

Appendix

A1 Data

We construct our earnings data to adjust for the duration of time a worker was employed at a firm and ultimately create an annual measure of earnings. This follows the approach in Abowd et al. (2003) and Sorkin (2018). Specifically, we define a worker to be full-quarter employed if in the preceding and subsequent quarters they also received earnings from the firm. Additionally, we define the worker as continuously employed if they received earnings in *either* the preceding or subsequent quarters. Lastly, the worker is discontinuously employed if they received earnings in neither the preceding nor the subsequent quarters. We can now aggregate the quarterly earnings data into an annual measure as follows:

1. If the worker has any full-quarter employment, we take the average over those full quarters and multiply by 4

2. If the worker has no full-quarter employment but does have at least one continuous quarter of employment, we take the average over the continuous employment quarters and multiply this number by 8

3. If the worker only has discontinuous employment in that year, we take the average over the discontinuous employment quarters and multiply this number by 12

A2 Estimation

We estimate $\theta = (\mu, \sigma, \lambda_e, \gamma, \lambda_u, z_{min})$ by simulated method of moments:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, L(\theta) \equiv \sum_i \left(\frac{m_i^{\text{model}}(\theta) - m_i^{\text{data}}}{m_i^{\text{data}}}\right)^2 \tag{22}$$

We use a global-to-local search strategy. First, we generate 48 starting values by Latin hypercube sampling over the parameter bounds. Then, we keep the 6 best starts and select the best local optimum. Finally, we increase the simulation accuracy (by increasing the number of quadrature nodes) and run a final local optimization starting from the best candidate.

To report standard errors, we recast the simulated method of moments approach in terms of GMM and, assuming the standard regularity conditions, we know that the estimator is asymptotically normal with variance given by:

$$Var(\hat{\theta}) = (G'WG)^{-1}G'WSWG(G'WG)^{-1} \tag{23}$$

where S is the variance-covariance matrix of the empirical moments, G is the gradient of the model moments (which we compute numerically), and where the weighting matrix W is a diagonal matrix with values scaled by $\frac{1}{m_i^{data}}^2$. As is common practice, we set the off-diagonal elements of S to zero. The empirical variances for our Census moments are straight-forward, but for the two non-Census moments, we construct conservative approximate variances and describe the assumptions used to do so.

For the employment-to-employment transition rate, we use the average monthly Fujita, Moscarini, and Postel-Vinay E2E transition probability from 1997 to 2022. In the Appendix of Fujita et al. (2024), they report the sample size for five different respondent groups used in calculating the EE rate from September 1995 to May 2020. The total sample size is 11664941 which implies an average sample of 39,000 per month. This suggests that the underlying sample used in generating our average EE rate from 1997 to 2022 is 12.2 million. Since the EE rate is the mean of an indicator variable, we can approximate the variance as follows[22]

$$Var(\bar{EE}) \approx \frac{\bar{EE}(1-\bar{EE})}{N} = 1.89 \times 10^{-9} \tag{24}$$

We multiply this standard error approximation by 10 in order to generate a conservative benchmark: 1.89×10^{-8}.

The second non-Census moment we use is the average unemployment rate from 1997 to 2022. There are approximately 110,000 individuals surveyed each month.[23] This means that the sample underlying our average unemployment rate is approximately 34.3 million. And similarly, we can approximate the variance as $Var(\bar{u}) = \frac{\bar{u}(1-\bar{u})}{N} =$

A3 Proofs

A3.1 Proofs for Section 4

A3.1.1 Proposition 1

We know that the first order conditions for the firm's problem satisfy:

[22]This is an approximation because our EE rate is an average across time and is not weighted by the number of observations in each month, which is not reported. This also does not account for error in imputation process that Fujita et al. (2024) use in generating the underlying observations.

[23]see: `https://www.bls.gov/cps/cps_htgm.htm`

$$\mu_{t+1}(V_{t+1}, L_{t+1}(L_t, V_{t+1}), V_{t+2}(L_{t+1}, V_{t+1}, s_{t+2})) = \mu_t(V_t, L_t, V_{t+1})\left[(1-\lambda_e)(1-\delta) + \lambda_e(1-\delta)F(V_{t+1})\right] + \Pi_L(V_{t+1}, L_{t+1}(L_t, V_{t+1}), s_{t+1})L_V(V_{t+1}, L_t) \tag{25}$$

$$\mu_t(V_t, L_t, V_{t+1}) = \frac{L_t}{u'(w_t(V_t, V_{t+1}))} \tag{26}$$

To first order in s, we have shown that the equilibrium is the same as the zero-risk equilibrium. So that means we have a steady state. Rearranging and dropping time subscripts:

$$\frac{L}{u'(w)} = \frac{L}{u'(w)}(1-\delta)[1-\lambda_e(1-F(V))]) + \Pi_L L_V \tag{27}$$

We can differentiate Π with respect to L to show:

$$\Pi_L = \frac{z-w}{1-\beta(1-\delta)[1-\lambda_e(1-F(V))]} \tag{28}$$

Now, plugging in 28 to 27 and rearranging:

$$\frac{z-w}{w} = \frac{1-(1-\delta)[1-\lambda_e(1-F(V))]}{u'(w)\frac{wL_V}{L}}\left(1-\beta(1-\delta)[1-\lambda_e(1-F(V))]\right) \tag{29}$$

Note that this is the flow labor response to a change in V. If we wanted to convert this to a permanent response, we can use that $L_{stock,V} = \frac{L_V}{1-(1-\delta)[1-\lambda_e(1-F(V))]}$. So we have:

$$\frac{z-w}{w} = \frac{1-\beta(1-\delta)[1-\lambda_e(1-F(V))]}{u'(w)\frac{wL_{stock,V}}{L}} \tag{30}$$

Lastly, we might want to have this expression in terms of an elasticity of stock labor with respect to wages, since the utility of a contract is not observable. There are many ways one could adjust the wage to achieve the same increase in promised utility, but since we are consider the stock (long-run) elasticity we can focus on what permanent wage increase achieve a given increase in promised utility. To achieve this, we can write the promise-keep constraint as follows (constant V):

$$V = u(w) + \lambda_e(1-\delta)\beta\int \max(V, V_0)dF(V_0) + (1-\lambda_e)(1-\delta)\beta V + \delta\beta V_U \tag{31}$$

implicitly differentiating with respect to V shows that:

$$w_V = \frac{1-(1-\delta)\beta[1-\lambda_e+\lambda_e F(V)]}{u'(w)} \tag{32}$$

So, we can write $L_{stock,V} = \frac{dL_{stock}}{dw} w_V = \frac{dL_{stock}}{dw} \frac{1-(1-\delta)\beta[1-\lambda_e+\lambda_e F(V)]}{u'(w)}$. Plugging this in and letting $\varepsilon_{stock,Lw}$ be the elasticity of stock labor with respect to a permanent increase in the wage:

$$\frac{z-w}{w} = \frac{1}{\varepsilon_{stock,Lw}} \tag{33}$$

Now, turning to solving for the wages and labor. To first order in s, we have shown that the equilibrium is the same as the zero-risk equilibrium. This is similar to Bontemps et al. (2000) in that it is a Burdett and Mortensen (1998) model with heterogeneous productivity drawn from a continuous distribution, but also features risk-averse workers.

First, we will solve for the steady state labor at a promised utility V. The labor flow equation in steady state must satisfy:

$$L(V) = L(V)(1-\delta)(1-\lambda_e F(V)) + \lambda_u u + \lambda_e(1-u)(1-\delta)G(V) \tag{34}$$

We can rearrange to show that:

$$L(V) = \frac{\lambda_u u + \lambda_e(1-u)(1-\delta)G(V)}{1-(1-\delta)(1-\lambda_e)F(V)} \tag{35}$$

Now we need to solve for u, $G(V)$, and $F(V)$. We know that we have a rank-preserving equilibrium, so $F(V) = J(z)$. The unemployment flow equation in steady state satisfies:

$$u = \delta(1-u) + (1-\lambda_u \int_{V_R} dF(V))u \tag{36}$$

where V_R is the reservation promised utility (which can also be formulated as a reservation wage since wages are constant in the first order approximation).

$$\implies u = \frac{\delta}{\delta + \int_{V_R} dF(V)} \tag{37}$$

The reservation promised utility is equal to V_U:

$$V_U = u(b) + \lambda_u \beta \int_{V_U} V_0 dF(V_0) + (1-\lambda_u[1-F(V_U)])\beta V_U \tag{38}$$

When the wage is constant, we can solve also for a reservation wage which will satisfy:

$$V_U = u(w_R) + \lambda_e(1-\delta)\beta \int \max(V_U, V_0) dF(V_0) + \beta(1-\lambda_e)(1-\delta)V_U + \beta\delta V_U \tag{39}$$

Now we must solve for $G(V)$. For this, we can consider the flows into and out of employment below the promised utility V. Let $\hat{G}(V)$ be the mass of workers with a promised utility below V:

$$\hat{G}(V) = \lambda_u F(V)u + [(1-\lambda_e)(1-\delta) + (1-\delta)\lambda_e F(V)]\hat{G}(V) \tag{40}$$

Rearranging and using that $\frac{\hat{G}(V)}{1-u} = G(V)$ and $F(V) = J(z)$:

$$G(V) = \frac{u}{1-u}\frac{\lambda_u F(V)}{1-(1-\delta)[1-\lambda_e+\lambda_e F(V)]} \tag{41}$$

Now we have all the requisite components to solve for $L(V)$ in closed form.

Now, we can take the wage solution, $w(z) = z - \int_{w^R}^{z} \frac{\left[\delta+(1-\delta)\lambda_e(1-J(z))\right]^2}{\left[\delta+(1-\delta)\lambda_e(1-J(x))\right]^2}\,dx$, and calculate $m(z) = \frac{z-w}{w}$:

$$m(z) = \frac{\int_{w^R}^{z} \frac{\left[\delta+(1-\delta)\lambda_e(1-J(z))\right]^2}{\left[\delta+(1-\delta)\lambda_e(1-J(x))\right]^2}\,dx}{z - \int_{w^R}^{z} \frac{\left[\delta+(1-\delta)\lambda_e(1-J(z))\right]^2}{\left[\delta+(1-\delta)\lambda_e(1-J(x))\right]^2}\,dx} \tag{42}$$

A3.1.2 Proposition 2

We know that $w(z) = z - \int_{w^R}^{z} \frac{\left[\delta+(1-\delta)\lambda_e(1-J(z))\right]^2}{\left[\delta+(1-\delta)\lambda_e(1-J(x))\right]^2}\,dx$.

And also:

$$m(z) = \frac{\int_{w^R}^{z} \frac{\left[\delta+(1-\delta)\lambda_e(1-J(z))\right]^2}{\left[\delta+(1-\delta)\lambda_e(1-J(x))\right]^2}\,dx}{z - \int_{w^R}^{z} \frac{\left[\delta+(1-\delta)\lambda_e(1-J(z))\right]^2}{\left[\delta+(1-\delta)\lambda_e(1-J(x))\right]^2}\,dx} \tag{43}$$

Clearly, this expression only depends on the wage through w^R. We know that the reservation wage satisfies the following equation:

$$V_U = u(w_R) + \lambda_e(1-\delta)\beta \int \max(V_U, V_0)dF(V_0) + \beta(1-\lambda_e)(1-\delta)V_U + \beta\delta V_U \tag{44}$$

where V_U is determined by:

$$V_U = u(b) + \lambda_u\beta \int_{V_U} \max(V_U, V_0)dF(V_0) + (1-\lambda_u)\beta V_U \tag{45}$$

We can rearrange these two expression to write the reservation wage as follow:

$$u(w_R) = u(b) + \beta[\lambda_e(1-\delta) - \lambda_u]\left(\int_{V_U} \max(V_U, V_0)dF(V_0) + V_U\right) \tag{46}$$

To see that this implies that w_R is decreasing in γ, consider the difference between the utility reservation wage and a hypothetical wage at a new job, $u(w) - u(w_R) = \frac{w^{1-\gamma} - w_R^{1-\gamma}}{1-\gamma}$. Differentiating with respect to γ gives us:

$$= \frac{(1-\gamma)\left[-\ln(w)w^{1-\gamma} + \ln(w_R)w_R^{1-\gamma}\right] + (w^{1-\gamma} - w_R^{1-\gamma})}{(1-\gamma)^2} \tag{47}$$

This is negative when $w > w_R$. So the gap in utility decreases between the reservation wage and a future higher wage. This means that the right-hand side of 46 is smaller (because $\lambda_e(1-\delta) - \lambda_u > 0$ in the proposition), so the reservation wage must fall as γ rises.

A3.1.3 Derivation of the wage and labor response to idiosyncratic shocks:

Set Up Reminder

Firm solves:

$$\Pi(V_t, L_t, s_t, \bar{z}) = \max_{w_t, V_{t+1}(s_{t+1})}((1+s_t)\bar{z} - w_t)L_t + \beta\mathbb{E}_{s_{t+1}}\left[\Pi(V_{t+1}, L_{t+1}, s_{t+1}, \bar{z})\right] \tag{48}$$

subject to a promise-keeping constraint:

$$V_t = u(w_t) + \beta(1-\lambda_e)(1-\delta)\mathbb{E}_{s_{t+1}}\left[V_{t+1}\right] + \beta\lambda_e(1-\delta)\mathbb{E}_{s_{t+1}}\left[\int \max(V_0, V_{t+1})dF(V_0)\right] + \beta\delta V_U \tag{49}$$

50 and 51 are the promised utility and wage FOCs respectively with every relevant dependency written out explicitly to facilitate differentiation.

$$\mu_{t+1}(V_{t+1}, L_{t+1}(L_t, V_{t+1}), V_{t+2}(L_{t+1}, V_{t+1}, s_{t+2})) = \mu_t(V_t, L_t, V_{t+1})\left[(1-\lambda_e)(1-\delta) + \lambda_e(1-\delta)F(V_{t+1})\right] + \Pi_L(V_{t+1}, L_{t+1}(L_t, V_{t+1}), s_{t+1})L_V(V_{t+1}, L_t) \tag{50}$$

$$\mu_t(V_t, L_t, V_{t+1}) = \frac{L_t}{u'(w_t(V_t, V_{t+1}))} \tag{51}$$

First Order Approximation for Wage/Labor Pass-Throughs

We start in the zero-risk steady state in period 0. The firm learns that in period 1 there will be a shock that is either a small positive or small negative s_1. Now we will solve for the entire path of promised utility, wages, and labor in response to this shock. Consider the first order conditions for wages and promised utility for the firm at time t which is choosing which promised utility to offer in $t+1$—these are 50 and 51 shown above. We can implicitly differentiate these equations with respect to the initial shock s_1 to ask how promised utility, wages, and labor adjust on impact ($t = 1$) and over time in response to this idiosyncratic shock.

First, note that the response of promised utility in period $t + 1$ depends on how the firm state variables—incumbent labor, incumbent promised utility, and the shock itself—have adjusted due to the shock. 52 decomposes the total derivative into these three partial derivatives. Note that the small idiosyncratic risk has no effect on any equilibrium objects to first order, so we have omitted the time subscripts when writing these partial derivatives. Once we solve for the partial derivatives, we will be able to evaluate them at the closed form zero-risk steady state equilibrium values (for labor, wages, etc.).

$$\frac{dV_{t+1}}{ds_1} = V_s \frac{ds_{t+1}}{ds_1} + V_L \frac{dL_t}{ds_1} + V_V \frac{dV_t}{ds_1} \tag{52}$$

In period $t = 1$, right when the shock hits, we will have $\frac{dL_0}{ds_1} = 0$ and $\frac{dV_0}{ds_1} = 0$. For $t > 0$, $\frac{dL_t}{ds_1}$ and $\frac{dV_t}{ds_1}$ vary as the firm's labor responds and their existing promised utility reflects their previous decisions.

The shock has persistence ρ—which implies 53.

$$\frac{ds_{t+1}}{ds_1} = \rho^t \tag{53}$$

We can differentiate the flow labor equation 54 with respect to L_t and with respect to V_{t+1} to recursively define the $\frac{dL_t}{ds_1}$ based on the current and prior promised utility responses. This is shown in 55.

$$\begin{aligned} L_{t+1,flow}(V_{t+1}, L_t) = L_t(1-\delta)\left(1 - \lambda_e(1 - F(V_{t+1})) \times \mathbb{I}[V_{t+1} > V_U]\right) \\ + \lambda_u u \times \mathbb{I}[V_{t+1} > V_U] + \lambda_e(1-u)(1-\delta)G(V_{t+1}) \end{aligned} \tag{54}$$

$$\frac{dL_t}{ds_1} = \left[(1-\lambda_e)(1-\delta) + \lambda_e(1-\delta)F(V_t)\right] \frac{dL_{t-1}}{ds_1} + L_V \frac{dV_t}{ds_1} \tag{55}$$

Stacking these equations, the first-order responses satisfy the linear system

$$x_{t+1} = Ax_t, \qquad x_t \equiv \begin{pmatrix} \frac{dL_t}{ds_1} \\ \frac{dV_t}{ds_1} \\ \frac{ds_t}{ds_1} \end{pmatrix} \tag{56}$$

where:

$$A \equiv \begin{pmatrix} C + L_V V_L & L_V V_V & \rho L_V V_s \\ V_L & V_V & \rho V_s \\ 0 & 0 & \rho \end{pmatrix} \tag{57}$$

Note that A is upper block-triangular, so its eigenvalues are the AR(1) root ρ and the two eigenvalues of the 2×2 endogenous block. The latter solve:

$$\lambda^2 - (C + L_V V_L + V_V)\lambda + C V_V = 0 \tag{58}$$

Given one predetermined state (L_t) and one forward-looking variable (V_t), the Blanchard–Kahn condition implies that this quadratic has one root inside and one root outside the unit circle. We will solve for the different endogenous terms of A and then numerically compute these eigenvalues and select the stable root (the one with $|\lambda| < 1$). This pins down the response over time of promised utility and labor to the shock.

Let us use the FOC 50 to define $\mathcal{H}(V_{t+1}, L_t, V_t, s_t)$:

$$\begin{aligned} \mathcal{H}(V_{t+1}, L_t, V_t, s_t) \equiv\ & \mu_{t+1}(V_{t+1}, L_{t+1}(L_t, V_{t+1}), V_{t+2}(L_{t+1}, V_{t+1}, s_{t+2})) \\ & - \mu_t(V_t, L_t, V_{t+1}) \left[(1-\lambda_e)(1-\delta) + \lambda_e(1-\delta)F(V_{t+1})\right] \\ & - \Pi_L(V_{t+1}, L_{t+1}(L_t, V_{t+1}), s_{t+1}) L_V(V_{t+1}, L_t) \end{aligned} \tag{59}$$

Note that we have the following three identities by the implicit function theorem:

$$V_L = -\frac{\mathcal{H}_L}{\mathcal{H}_{V_{t+1}}} \tag{60}$$

$$V_s = -\frac{\mathcal{H}_s}{\mathcal{H}_{V_{t+1}}} \tag{61}$$

$$V_{V_t} = -\frac{\mathcal{H}_{V_t}}{\mathcal{H}_{V_{t+1}}} \tag{62}$$

Starting with solving for $\mathcal{H}_{V_{t+1}}$, we will let $C(V) \equiv [(1-\lambda_e)(1-\delta) + \lambda_e(1-\delta)F(V)]$

and differentiate with respect to V_{t+1} to show: [24]

$$\mathcal{H}_{V_{t+1}} = \mu_V + \mu_L L_V + \mu_{V'}(V_V + V_L L_V) - \mu C_V - \mu_{V'} C - \Pi_{LV} L_V - \Pi_{LL} L_V^2 - \Pi_L L_{VV} \quad (63)$$

$$\mathcal{H}_L = \mu_L L_L + \mu_{V'} V_L L_L - \mu_L C - \Pi_{LL} L_L L_V - \Pi_L L_{VL} \quad (64)$$

$$\mathcal{H}_s = \mu_{V'} V_s \frac{ds_{t+2}}{ds_{t+1}} - \Pi_{Ls} L_V \quad (65)$$

$$\mathcal{H}_{V_t} = -\mu_V C \quad (66)$$

Now we begin deriving the term in these equations to be solely in terms of zero risk equilibrium objects. First, we will solve for Π_L and its subsequent derivatives.

$$\begin{aligned}\Pi_L(V_{t+1}, L_{t+1}(L_t, V_{t+1}), s_{t+1}) = (1 + s_{t+1})z - w_{t+1}(V_{t+1}, V_{t+2}(L_{t+1}, s_{t+2})) \\ + \beta(1-\delta)\Big[(1-\lambda) + \lambda F(V_{t+2}(L_{t+1}, s_{t+2}))\Big] \\ \times \Pi_L(V_{t+2}(L_{t+1}, s_{t+2}), L_{t+2}(L_{t+1}, V_{t+2}(L_{t+1}, s_{t+2})), s_{t+2})\end{aligned} \quad (67)$$

Note that $\Pi_V = -\mu(V_{t+1}, L_{t+1}, V_{t+2}(L_{t+1}, V_{t+1}, s_{t+2}))$, so we have that:

$$\Pi_{LV} = \Pi_{VL} = -\mu_L - \mu_{V'} V_L \quad (68)$$

Next, solving for Π_{LL}

$$\Pi_{LL} = -\frac{dw_{t+1}}{dV_{t+2}} V_L + \beta C_V V_L \Pi_L + \beta C \Pi_{LL} (\frac{dL_{t+2}}{dL_{t+1}}) + \beta C \Pi_{LV} V_L \quad (69)$$

We can solve for w_V from the promise-keeping constraint (70):

$$V_t = u(w_t) + \beta(1-\lambda_e)(1-\delta)V_{t+1} + \beta\lambda_e(1-\delta) \int \max(V_0, V_{t+1})\, dF(V_0) + \beta\delta V_U \quad (70)$$

Differentiating with respect to V_{t+1}:

[24] sometimes we omit time subscripts for notational simplicity because we are evaluating at steady state. Additionally, let $\frac{\partial \mu(V_t, L_t, V_{t+1})}{\partial V_{t+1}} \equiv \mu_{V'}$

$$\frac{dw_t}{dV_{t+1}} = \frac{1}{u'(w)}[-\beta(1-\lambda_e)(1-\delta) - \beta\lambda_e(1-\delta)F(V)] = \frac{-\beta C}{u'(w)} \tag{71}$$

Now simplifying 70 using 71 and 68 and rearranging:

$$\Pi_{LL} = \frac{\beta C_V V_L \Pi_L + \frac{L\gamma}{u'(w)^2 w}(\beta C)^2 V_L^2}{1 - \beta C(\frac{dL_{t+2}}{dL_{t+1}})} \tag{72}$$

Now turning to Π_{Ls}:

$$\Pi_{Ls} = z - \frac{dw_{t+1}}{dV_{t+2}}\frac{ds_{t+2}}{ds_{t+1}}V_s + \beta C_V \frac{ds_{t+2}}{ds_{t+1}}V_s\Pi_L + \beta C\frac{ds_{t+2}}{ds_{t+1}}\Pi_{Ls} + \beta C\Pi_{LV}\frac{ds_{t+2}}{ds_{t+1}}V_s + \beta C\Pi_{LL}\frac{ds_{t+2}}{ds_{t+1}}L_V V_s \tag{73}$$

Using 53, 68, and 71 and letting $\Phi_0 \equiv \frac{1}{1-\beta\rho C}$, we can simplify:

$$\Pi_{Ls} = \Phi_0 z + \beta\Phi_0\rho C_V V_s \Pi_L + \beta\Phi_0\rho C\Pi_{LL}L_V V_s + (\beta C)^2\rho\Phi_0\frac{L\gamma}{u'(w)^2 w}V_L V_s \tag{74}$$

Next, we solve for labor related objects.

$$L_L(V_{t+1}, L_t) = [(1-\lambda_e)(1-\delta) + \lambda_e(1-\delta)F(V)] = C \tag{75}$$

$$L_V(V_{t+1}, L_t) = (1-\delta)\lambda_e F'(V)L + \lambda_e(1-u)(1-\delta)G'(V) \tag{76}$$

Combining 76 and 75 we have:

$$\frac{dL_{t+1}}{dL_t} = L_L + L_V V_L = C + (1-\delta)\lambda_e F'(V)V_L L + \lambda_e(1-u)(1-\delta)G'(V)V_L \tag{77}$$

Additionally, we have:

$$L_{VV} = (1-\delta)\lambda_e F''(V)L + \lambda_e(1-u)(1-\delta)G''(V) \tag{78}$$

also:

$$L_{LV} = \lambda_e(1-\delta)F'(V) = C_V \tag{79}$$

Finally, we need to solve for the derivatives of the Lagrange multiplier:

$$\mu_V = -\frac{L\frac{\partial w_{t+1}}{\partial V_{t+1}}u''(w)}{u'(w)^2} = \frac{\gamma L}{u'(w)w}\frac{\partial w_{t+1}}{\partial V_{t+1}} \tag{80}$$

$$\mu_L = \frac{1}{u'(w)} \tag{81}$$

$$\mu_{V'} = \frac{\gamma L}{u'(w)w} \frac{\partial w_{t+1}}{\partial V_{t+2}} \tag{82}$$

We can solve for $\frac{dw_{t+1}}{dV_{t+1}}$ from the promise-keeping constraint 70 (holding V_{t+2} constant)

$$\frac{\partial w_{t+1}}{\partial V_{t+1}} = \frac{1}{u'(w)} \tag{83}$$

Now for $\frac{dw_{t+1}}{dV_{t+2}}$

$$\frac{dw_{t+1}}{dV_{t+2}} = \frac{-\beta C}{u'(w)} \tag{84}$$

To have completely defined the system of equations which pins down the response to promised utility (and therefore wages and labor) to the shock, we now need some objects from the zero-risk steady state solution. First, how does F and G change with V? We have a rank-preserving equilibrium, which implies that:

$$F(V(z)) = J(z) \tag{85}$$

The following two equations follow:

$$F'(V(z))V'(\bar{z}) = J'(z) \tag{86}$$

$$F''(V) = \frac{J''(z) - F'(V)V''(z)}{V'(z)^2} \tag{87}$$

We can also solve for G' and G'' in terms of F, F', and F''.

$$G'(V) = \frac{(\delta + (1-\delta)\lambda_e)\delta F'(V)}{(\delta + (1-\delta)\lambda_e(1 - F(V)))^2} \tag{88}$$

$$G''(V) = (\delta + (1-\delta)\lambda_e)\delta \frac{F''(V)\left(\delta + (1-\delta)\lambda_e(1 - F(V))\right) + 2F'(V)^2(1-\delta)\lambda_e}{(\delta + (1-\delta)\lambda_e(1 - F(V)))^3} \tag{89}$$

Also note that we have the following two steady state equations:

$$L(z) = \frac{\frac{\lambda_u \delta}{\lambda_u + \delta}\left[\delta + (1-\delta)\lambda_e\right]}{\left[\delta + (1-\delta)\lambda_e\left(1 - J(z)\right)\right]^2} \tag{90}$$

$$u = \frac{\delta}{\lambda_u + \delta} \tag{91}$$

Next, we can plug 88 and 86 into 76 and 78: Plugging in 88 and 86 into **??** and then simplifying using 90 and 91 we have:

$$L_V = 2(1-\delta)\lambda_e F'(V)L = 2(1-\delta)\lambda_e \frac{J'(\bar{z})}{V'(\bar{z})}L \tag{92}$$

Now, we plug in 89 into 78 and simplify using the steady-state labor stock 90 [25] and the unemployment rate 91 to get $L''(V)$ in terms of F', F'':

$$L_{VV} = \lambda_e(1-\delta)L\left[2F''(V) + \frac{2\lambda_e F'(V)^2(1-\delta)}{\delta + (1-\delta)\lambda_e(1-F(V))}\right] \tag{93}$$

Using 87 and 86 we can continue to simplify 93:

$$L_{VV} = \frac{2\lambda_e(1-\delta)L}{V'(\bar{z})^2}\left[J''(\bar{z}) - \frac{J'(\bar{z})V''(\bar{z})}{V'(\bar{z})} + \frac{\lambda_e J'(\bar{z})^2(1-\delta)}{\delta + (1-\delta)\lambda_e(1-J(\bar{z}))}\right] \tag{94}$$

Now we need to know $V'(\bar{z})$, and $V''(\bar{z})$ to simplify L_{VV} and L_V. Since we are asking how the promised utility varies with productivity in the zero-risk steady state, we know the wage the firms offer is constant. So we can explicitly map $w(z)$ onto $V(z)$ and so forth. This also allows us to solve for Π_L.

$$V'(z) = \frac{u'(w)w'(z)}{1 - \beta(1-\lambda_e)(1-\delta) - \beta\lambda_e(1-\delta)F(V)} \equiv \Phi_1 u'(w)w'(z) \tag{95}$$

Where Φ_1 is the present-value multiplier associated with permanent differences. It is the same as Φ_0 if $\rho = 1$, but is strictly larger if the shock we are considering is not entirely permanent. Similarly for $V''(\bar{z})$:

$$V''(z) = \left[-\frac{\gamma w'(z)^2}{w(z)} + w''(z) + \beta\lambda_e(1-\delta)J'(\bar{z})w'(z)\Phi_1\right]\Phi_1 u'(w) \tag{96}$$

Additionally, note that we can solve for Π_L:

$$\Pi_L = (z-w)\Phi_1 \tag{97}$$

We can use the steady-state wage equations to continue to simplify:

[25] in contrast to the previous expression 54 which is a flow equation for labor

$$w(\bar{z}) = \bar{z} - \int_{w^R}^{\bar{z}} \frac{\left[\delta + (1-\delta)\lambda_e \left(1 - J(\bar{z})\right)\right]^2}{\left[\delta + (1-\delta)\lambda_e \left(1 - J(x)\right)\right]^2} \, dx \tag{98}$$

$$w'(\bar{z}) = (\bar{z} - w(\bar{z})) \, \frac{2\,(1-\delta)\lambda_e \, J'(\bar{z})}{\delta + (1-\delta)\lambda_e \left(1 - J(\bar{z})\right)} \tag{99}$$

$$w''(\bar{z}) = w'(\bar{z}) \left[\frac{1 - w'(\bar{z})}{\bar{z} - w(\bar{z})} + \frac{J''(\bar{z})}{J'(\bar{z})} - \frac{(1-\delta)\,\lambda_e \, J'(\bar{z})}{\delta + (1-\delta)\lambda_e \big(1 - J(\bar{z})\big)} \right] \tag{100}$$

In order to simplify our expression for L_{VV} 93, the following equation is useful:

$$\frac{V''(\bar{z})}{V'(\bar{z})} = \frac{V''(\bar{z})}{\Phi_1 u'(w) w'(z)} = \left[-\frac{\gamma w'(\bar{z})}{w(z)} + \frac{w''(\bar{z})}{w'(\bar{z})} + \beta\lambda_e(1-\delta) J'(\bar{z}) \Phi_1 \right] \tag{101}$$

Using 101, 100, and 99: we can to continue to simplify 94 and remove all V' and V'' terms:

$$L_{VV} = \frac{2\lambda_e(1-\delta)\, L}{V'(\bar{z})^2} \left[\frac{\gamma \, w'(\bar{z}) \, J'(\bar{z})}{w(\bar{z})} - \frac{J'(\bar{z})(1 - w'(\bar{z}))}{\bar{z} - w(\bar{z})} - \beta\lambda_e(1-\delta) J'(\bar{z})^2 \Phi_1 \right] \tag{102}$$

Now we collect our terms and simplify our expressions for V_s, V_L, and V_V:

$$\begin{aligned} \mathcal{H}_{V_{t+1}} = {} & \frac{2(1-\delta)\lambda_e F'(V) L}{u'(w)} - \frac{L(1-\beta C V_V) u''(w)}{u'(w)^3} - \frac{L\lambda_e(1-\delta) F'(V)}{u'(w)} + \frac{2(1-\delta)\lambda_e F'(V) L}{u'(w)} - \Pi_{LL} L_V^2 \\ & - \frac{2\lambda_e(1-\delta)\, L(z-w)\Phi_1}{V'(\bar{z})^2} \left[\frac{\gamma \, w'(\bar{z}) \, J'(\bar{z})}{w(\bar{z})} - \frac{J'(\bar{z})(1 - w'(\bar{z}))}{\bar{z} - w(\bar{z})} - \beta\lambda_e(1-\delta) J'(\bar{z})^2 \Phi_1 \right] - \frac{2 L_V V_L L \gamma}{u'(w)^2 w} \beta C + \frac{\gamma L \beta C^2}{u'(w)^2 w} \end{aligned} \tag{103}$$

Collecting like terms and simplifying:

$$\begin{aligned} \mathcal{H}_{V_{t+1}} = {} & \frac{(1-\delta)\lambda_e F'(V) L}{u'(w)} - \frac{L(1-\beta C V_V) u''(w)}{u'(w)^3} - \Pi_{LL} L_V^2 \\ & - \frac{2\lambda_e(1-\delta) L(z-w)\Phi_1 J'(z)}{V'(\bar{z})^2} \left[\frac{\gamma \, w'(z)}{w(z)} - \frac{1}{z - w(z)} - \beta\lambda_e(1-\delta) J'(z) \Phi_1 \right] - \frac{2 L_V V_L L \gamma}{u'(w)^2 w} \beta C + \frac{\gamma L \beta C^2}{u'(w)^2 w} \end{aligned} \tag{104}$$

Further simplifying:

$$\mathcal{H}_{V_{t+1}} = \frac{L_V}{2u'(w)} + \frac{\gamma}{u'(w)w}\left(\frac{L(1-\beta C V_V)}{u'(w)} - (z-w)L_V + \frac{L\beta C^2}{u'(w)}\right) - \Pi_{LL}L_V^2 + \frac{L_V}{u'(w)w'(z)} + \frac{\beta\Phi_1 L_V(1-C)}{2u'(w)} - \frac{2L_V V_L L\gamma}{u'(w)^2 w}\beta C \quad (105)$$

Similarly, simplifying $\mathcal{H}_L$:

$$\mathcal{H}_L = -\Pi_{LL}L_V C - \frac{1-C}{2u'(w)} - \frac{L\gamma}{u'(w)^2 w}\beta C^2 V_L \quad (106)$$

$$\mathcal{H}_s = -L_V[\Phi_0 z + \frac{\beta\Phi_0(1-C)\rho V_s}{2u'(w)} + \beta\Phi_0\rho C\Pi_{LL}L_V V_s + (\beta C)^2\rho\Phi_0\frac{L\gamma}{u'(w)^2 w}V_L V_s] - \frac{\gamma L\beta C\rho V_s}{u'(w)^2 w} \quad (107)$$

$$\mathcal{H}_{V_t} = -\frac{CL\gamma}{u'(w)^2 w} \quad (108)$$

This is all the expressions required to solve the system of equations to simulate how promised utility and labor respond to a shock. We do this numerically, picking the stable solution (eigenvalues inside the unit circle).

Now, how do we get the wage?

$$\frac{dV_t}{ds_1} = u'(w)\frac{dw_t}{ds_1} + \beta C\frac{dV_{t+1}}{ds_1} \quad (109)$$

$$\implies \frac{dw_t}{ds_1} = \frac{\frac{dV_t}{ds_1} - \beta C\frac{dV_{t+1}}{ds_1}}{u'(w)} \quad (110)$$

A3.1.4 Planner problem

Here, we focus on the planner's choice of whether to move an employed worker who has received a job offer. The planner will move the worker if the social value of the new job, $W(z_1, s_1)$, exceeeds the social value of their current job, $W(z_0, s_0)$ where these value functions are defined as follows:

$$W(z,s) = (1+s)z + \beta(1-\lambda_e)(1-\delta)\mathbb{E}[W(z,s')] + \beta\lambda_e(1-\delta)\mathbb{E}\left[\int \max(W(z,s'),W_0)dF(W_0)\right] + \delta\beta W_U \quad (111)$$

Now, considering the first order approximation. Since the aggregate distributions are

unaffected by the small amount of risk, we know that $F(W(z,0)) = J(z)$.

$$W_s = z + \beta(1-\lambda_e)(1-\delta)\rho W_s + \beta\lambda_e(1-\delta)J(z)\rho W_s \tag{112}$$

$$W_s = \frac{z}{1-\beta(1-\delta)\rho[1-\lambda_e+\lambda_e J(z)]} = z\Phi_0 \tag{113}$$

Labor evolves according to:

$$L(z,s_t) = L(z,s_{t-1})(1-\delta)(1-\lambda_e(1-F(W(z,s_t)))+\lambda_u u+\lambda_e(1-u)(1-\delta)G(W(z,s_t)) \tag{114}$$

$$G(W(z,0)) = \frac{\delta F(W(z,0))}{\delta+(1-\delta)\lambda_e(1-F(W(z,0))} \tag{115}$$

Note that we have:

$$F'(W)W_z = J'(z) \tag{116}$$

And similarly:

$$G'(W) = \frac{(\delta+(1-\delta)\lambda_e)\delta F'(W)}{(\delta+(1-\delta)\lambda_e(1-F(W)))^2} = \frac{(\delta+(1-\delta)\lambda_e)\delta\frac{J'(z)}{W_z}}{(\delta+(1-\delta)\lambda_e(1-J(z))^2} \tag{117}$$

Now, what is W_z? We can return to 111 to solve for this (again, this is with $s=0$):

$$W_z = 1 + \beta(1-\lambda_e)(1-\delta)W_z + \beta\lambda_e(1-\delta)J(z)W_z \tag{118}$$

Rearranging, we have:

$$W_z = \frac{1}{1-\beta(1-\delta)[1-\lambda_e+\lambda_e J(z)]} = \Phi_1 \tag{119}$$

So we can solve for the socially optimal labor response:

$$\begin{aligned}\frac{dL(z,s_t)}{ds_1} &= \frac{dL(z,s_{t-1})}{ds_1}(1-\delta)\big(1-\lambda_e(1-J(z))\big) \\ &\quad + \rho^{t-1}(1-\delta)\lambda_e\frac{J'(z)z\Phi_0}{\Phi_1}L + \rho^{t-1}(1-\delta)\lambda_e(1-u)\frac{(\delta+(1-\delta)\lambda_e)\,\delta\,\frac{J'(z)z\Phi_0}{\Phi_1}}{\big(\delta+(1-\delta)\lambda_e(1-J(z))\big)^2}.\end{aligned} \tag{120}$$

Simplifying using $u = \frac{\delta}{\lambda_u+\delta}$ and steady state labor:

$$\frac{dL(z,s_t)}{ds_1} = 2\rho^{t-1}(1-\delta)\lambda_e \frac{J'(z)z\Phi_0}{\Phi_1} L + \frac{dL(z,s_{t-1})}{ds_1}(1-\delta)\big(1-\lambda_e(1-J(z))\big) \tag{121}$$

A3.1.5 Proof of Proposition

From the previous section, we know that the planner response in period t following a shock in period 1 is:

$$\frac{dL(z,s_t)}{ds_1} = 2\rho^{t-1}(1-\delta)\lambda_e \frac{J'(z)z\Phi_0}{\Phi_1} L + \frac{dL(z,s_{t-1})}{ds_1}(1-\delta)\big(1-\lambda_e(1-J(z))\big) \tag{122}$$

So when the shock is permanent we have $\Phi_0 = \Phi_1$ and $\rho = 1$ and:

$$\frac{dL(z,s_t)}{ds_1} = 2(1-\delta)\lambda_e J'(z)zL + \frac{dL(z,s_{t-1})}{ds_1}(1-\delta)\big(1-\lambda_e(1-J(z))\big) \tag{123}$$

First, note that this is the cumulative labor change associated with increasing the promised utility by $V'(z)z$ in the decentralized equilibrium. The instantaneous flow labor change associated with this increase is:

$$L_{V,t=1}V'(z)z = 2(1-\delta)\lambda_e J'(z)zL \tag{124}$$

And the increase over time is:

$$L_{V,t}V'(z)z = 2(1-\delta)\lambda_e J'(z)zL + L_{V,t-1}V'(z)z(1-\delta)(1-\lambda_e(1-J(z))) \tag{125}$$

This is analogous to 123, so we know that the planner's labor response to a permanent shock is analogous to the labor response induced by increasing the utility $V'(z)z$ in the decentralized equilibrium. Intuitively, this is because the shock s increased the productivity by z so the firm should move up the job ladder based on that change, $V'(z)z$. Productivity is CRS and there are no other output-related scale effects so this change should happen immediately (in contrast, if there were direct adjustment costs then the planner might slowly adjust over time too).

Next, note that the decentralized response converges to $V'(z)z$ over time. In the long run after the permanent shock, in the new steady state the firm's productivity has increased by z and so its promised utility will have increased by $V'(z)z$. But what happens along the transition path? It is always lower than this value. To see this recall that, the promised utility response is:

$$\frac{dV_{t+1}}{ds_1} = V_s \frac{ds_{t+1}}{ds_1} + V_L \frac{dL_t}{ds_1} + V_V \frac{dV_t}{ds_1} \tag{126}$$

In response to a permanent shock, the V_s channel is the same over time because $\frac{ds_{t+1}}{ds_1} = 1$. The V_L and V_V channels both change over time as $\frac{dL_t}{ds_1}$ and $\frac{dV_t}{ds_1}$ adjust following the permanent shock. Both V_L and V_V are positive. Accumulating workers boosts the retention margin and offering a higher promised utility makes it more costly to reduce the promised utility later (except when workers are risk neutral).

So $\frac{dL_t}{ds_1}$ and $\frac{dV_t}{ds_1}$ both are initially zero (when $t = 0$) and then positive which increases the promised utility until it converges to $zV'(z)$. What is the labor increase associated with this? Solving for steady state in labor:

$$\frac{dL^*}{ds_1} = \frac{2(1-\delta)\lambda_e J'(z) z L}{1-(1-\delta)[1-\lambda_e(1-J(z))]} \tag{127}$$