The Effects of Moral Hazard on Wage Inequality in a Frictional Labor Market*

Árpád Ábrahám[†] Fernando Álvarez-Parra[‡] Susanne Forstner[§]

First version: September 2012

This version: February 2015

Abstract

Performance-dependent pay is widely observed in labor contracts. In this paper we study the impact of performance pay, due to moral hazard, on the cross-sectional wage distribution. Our analysis builds on a search model with job-to-job mobility and competition among firms for workers. Firms offer long-term contracts to risk-averse workers in the presence of repeated moral hazard and two-sided limited commitment. In order to provide incentives to workers, wage payments need to depend on realized match output. This direct effect of moral hazard increases wage inequality by inducing wage dispersion among workers with otherwise identical histories. In the presence of on-the-job search and limited commitment, however, moral hazard also affects the wage distribution through several indirect channels, as incentive provision through wage variation increases the costs of worker effort to firms. For a quantitative analysis, we calibrate the model to match characteristics of the U.S. labor market in the mid-2000s. We find that, overall, the presence of moral hazard increases wage inequality by around ten percent. Within the lower half of the distribution, however, the increase in inequality is much larger. A decomposition of effects shows that limits to incentive provision in low-wage states significantly contribute to the increase in effort costs to firms. As a consequence, firms decrease effort levels prescribed and wages paid to their workers, and the wage distribution expands substantially at the lower end.

Keywords: Wage Dispersion, Job-to-Job Transitions, Repeated Moral Hazard, Incentive Contracts

JEL Codes: J31, J62, D31

*We thank Tobias Broer, Tom Cunningham, John Hassler, Hans Holter, Sebastian Koehne, Per Krusell, Moritz Kuhn, Ludo Visschers, and seminar and conference participants at Bocconi, IIES, Uppsala University, the Annual Conference of the Hungarian Economic Association 2012, ENSAI Economic Day on Consumption and Labor Supply 2013, Mannheim Workshop in Quantitative Macroeconomics 2013, SED Annual Meeting 2013, NBER Summer Institute 2013, Nordic Summer Symposium in Macroeconomics 2013, EEA Annual Congress 2013, and the Cologne Workshop on Macroeconomics 2013 for useful comments.

[†]European University Institute, Florence. Email: arpad.abraham@eui.eu

[‡]CAF Development Bank of Latin America, Caracas.

[§]Chair of Applied Economics, RWTH Aachen. Email: susanne.forstner@ewifo.rwth-aachen.de

1 Introduction

Why do workers with similar characteristics earn very different hourly wages? Since labor earnings are the main source of income for many households, answering this question is central to understanding the determinants of income and consumption inequality in a broader context. Previous studies have focused on the role of search frictions in the labor market in explaining wage dispersion. In particular, recent contributions to the literature have emphasized the importance of differences in "luck" among workers with respect to receiving outside job offers and moving to better-quality matches over time.[1] In the present paper we explore an additional, very intuitive source of wage dispersion: Performance pay due to moral hazard in the worker-firm relationship. If a worker's effort on the job is private information, wage payments need to vary across realizations of match output in order to provide incentives. Moral hazard thus leads to wage differences among workers with otherwise identical labor market histories and thereby contributes to wage inequality.

Performance-dependent pay is a widely observed feature of labor contracts.[2] However, to the best of our knowledge, it has not been taken into account in the analysis of wage inequality based on quantitative-theoretic frameworks of frictional labor markets. We analyze the effects of moral hazard – as a source of performance pay – on the wage distribution in such a framework. In particular, our paper addresses the following questions: How does moral hazard in employment relations interact with other frictions in the labor market? Through which channels does it affect the wage distribution? And finally, what is the resulting impact of moral hazard on the extent of wage inequality?

Our analysis builds on a search model with on-the-job search and competition among firms for workers.[3] Firms offer long-term contracts to risk-averse workers in the presence of repeated moral hazard and two-sided limited commitment. Combining on-the-job search and employer competition with a moral hazard problem in the worker-firm relation is of interest for at least two reasons. First, including both mechanisms of wage determination in a search model allows for a rich structure of individual wage dynamics, both within and across jobs, that successfully reproduces empirical observations. Second, and more importantly, we show that the two mechanisms influence each other: Incentive provision is affected by the presence of on-the-job search, and the outcomes of on-the-job search and employer competition are affected by the fact that incentive provision through wage variation is costly. In addition, the basic outside options of unemployment for the worker and of inactivity for the firm constrain the possibilities of providing incentives. These constraints increase the costs of incentive provision and, in consequence, affect the outcomes of on-the-job search and employer competition.

[1] See Hornstein et al. (2011) for a comprehensive survey and evaluation of this literature.

[2] For example, Lemieux et al. (2009) estimate that, between 1976 and 1998, around 40% of male workers in the U.S. received part of their compensation in the form of bonuses, commissions, or piece rates, and that this fraction is much higher in specific industries and occupations, reaching nearly 60% for wholesale trade and 80% for sales.

[3] Job-to-job mobility is a quantitatively important component of worker turnover. For example, Fallick and Fleischman (2004) estimate that, between 1994 and 2003, on average each month 2.6% of employed U.S. workers moved to a new employer, and that these job-to-job transitions made up around two-fifths of total monthly separations. Hornstein et al. (2011) argue that search models featuring job-to-job mobility and employer competition are among the most promising frameworks for reconciling observed worker turnover rates with the observed extent of wage dispersion. See Postel-Vinay and Robin (2002) for a seminal exposition and analysis of this type of search model.

In our quantitative analysis, based on a calibration of the model to U.S. data, we find that the presence of moral hazard in employment relations increases residual wage inequality in total by around 10%. Incentive provision interacts with search frictions and limited commitment to create quantitatively substantial effects on the wage distribution that partly offset each other. More importantly, these indirect effects have a particularly strong impact on the lower parts of the wage distribution. At the very bottom, wages paid and effort levels prescribed to workers decrease significantly. As a result, the increase in wage inequality is much larger in the lower half of the distribution, amounting to nearly 30%. This also suggests that the evaluation of policy instruments that mainly affect low-wage matches, such as the level of unemployment benefits or minimum wage restrictions, may change substantially when the need for incentive provision is taken into account.[4]

In terms of modelling, the present paper contributes to the literature by incorporating dynamic moral hazard into an equilibrium analysis of job-to-job mobility and firm competition for workers. In our model, risk-neutral firms offer long-term contracts to risk-averse, ex-ante identical workers in a labor market characterized by search frictions. Output of a worker-firm pair is a function of match-specific productivity and a stochastic productivity component that depends on the worker's unobservable effort. Commitment to contracts is limited on both the worker's and the firm's side, as firms can commit to wage payments only as long as profits are non-negative, and workers may quit to unemployment or report an outside job offer to their employer. As soon as an offer is disclosed, the current and the potential future employer start competing for the worker by offering new contracts. Bidding takes place in the form of Bertrand competition in terms of lifetime utility levels that the contracts promise to the worker.

In our framework, wage differences between ex-ante identical workers originate from search frictions and from the informational friction specific to the moral hazard problem. First, search frictions lead to different wages for workers with different histories of unemployment spells and job offers. For instance, wage levels differ between employed workers who have received outside job offers and those who have not. Likewise, job offers that are associated with different levels of match productivity imply different levels of wages. Second, when effort is not observable, firms need to provide incentives to their workers by making wage payments depend on output realizations. Moral hazard thus leads to wage differences between workers who received similar job offers, but experienced different histories of output realizations. In other words, the direct effect of moral hazard is to induce additional wage dispersion within groups of workers with the same job offer histories.

The latter argument suggests that the presence of moral hazard in employment relations unambiguously increases wage inequality. This would indeed be the case if the only effect of moral hazard were a mean preserving spread of wages, for each possible job-offer history, around the level determined by a model with observable effort. However, in the presence of on-the-job search and limited commitment, moral hazard affects the wage distribution also indirectly through a number of channels. All of these indirect effects are related to the fact that providing incentives through wage

[4] For studies that address such policy issues without considering incentive contracts, see, for instance, Flinn and Mabli (2009) on minimum wage restrictions and Lise et al. (2013) on the optimal level of unemployment benefits.

variation makes worker effort more expensive for firms. One reason is that workers need to be compensated not only for the disutility from spending effort but also for the variation in income arising from incentive provision. Another reason is that, since neither worker nor firm can fully commit to a contract, incentive provision is constrained at particular stages of an employment relation, which creates additional effort costs. Some of the indirect effects associated with increased effort costs counteract the direct effect of moral hazard on wage inequality whereas others reinforce it.

The impact of the different indirect effects can be outlined in the following way: First, in order to compensate risk-averse workers for income variation, higher wages have to be paid at all levels of effort and lifetime utility. Wage levels therefore increase throughout the wage distribution. Because required worker effort is higher at low levels of lifetime utility, the corresponding wage levels, however, increase more strongly. As a result, the wage distribution is compressed upwards from the lower end. Second, higher effort costs affect the maximum lifetime utility levels that firms are willing to offer. Increased costs of effort imply lower profits for firms from any match, and hence, lower levels of utility that firms can offer to workers while just breaking even. In consequence, outside job offers lead to lower gains for workers in terms of utility and wages. As the maximum utility levels attained by workers decrease, the wage distribution is compressed downwards from the upper end. At the same time, workers at low utility levels need to be compensated for lower expected continuation values, associated with outside job offers, by higher contemporaneous wages. As a result, the wage distribution is also compressed upwards from the lower end. Third, in reaction to higher effort costs, firms prescribe lower levels of effort to workers at all values of lifetime utility. This leads to lower wage levels throughout the distribution for two reasons: Contemporaneous wages are lower because workers need to be compensated less for disutility from effort. In addition, effort levels directly affect the dynamics of wages within a job. Lower effort levels imply lower probabilities of high output, and hence, lower probabilities of wage increases in the next period.

Finally, limited commitment of workers and firms creates additional costs of effort at very low levels of lifetime utility and at levels close to a firm's break-even value. On the one hand, since workers are free to quit to unemployment at any time, contracts ensure that workers are at least weakly better off working than being unemployed. Thus, when lifetime utility of a worker is close to the value of unemployment, incentive provision through utility variation is constrained from below. As a consequence, firms need to rely more heavily than they would like to on rewarding workers for high output through wage increases in order to provide incentives. These additional effort costs, arising from the worker's participation constraint, lead to further decreases in prescribed effort at low levels of lifetime utility. Wages in this range of utility decrease because of lower effort compensation and because workers expect wage increases in the future. Moreover, the probability of a wage raise declines, so that workers remain longer at low wage levels. As a result, the wage distribution expands at the lower end. On the other hand, contracts ensure that profits to firms remain non-negative. Therefore, incentive provision is constrained from above when a worker's lifetime utility is close to the break-even value of the firm and needs to take primarily the form of wage decrease as punishment for low output. The firm's participation constraint thus creates similar additional effort costs, and effort levels also decrease further towards the match-specific upper bound on utility. Since these costs arise at the break-even utilities of all match types, they impact on wages throughout the distribution. Moreover, lower effort compensation costs and lower probabilities of a wage raise act to decrease wage levels, while the need to compensate workers for expected wage decreases in the future acts to increase contemporaneous wages.

The goals of our quantitative analysis are to assess the overall impact of moral hazard on the wage distribution and to disentangle, characterize, and measure the direct and indirect effects. For this purpose, we calibrate our model to key moments of individual wage dynamics in the U.S. labor market, obtained from micro data from the Survey of Income and Program Participation, and perform a series of counterfactual experiments.

We find that, overall, the presence of moral hazard in employment relations increases residual wage inequality – in terms of the standard deviation of log wages – by around 10%. The strength of impact, however, is not even across different parts of the wage distribution. For instance, the ratio between the top and the bottom 5th-percentile wages increases by around 28%, and nearly all of this increase comes from larger inequality within the lower half of the distribution.

Our decomposition exercises show that the direct and indirect effects at work behind the overall change are quantitatively substantial and partially counteract each other. More precisely, the direct effect of incentive provision would by itself increase wage inequality by over 15%. Compensation of risk-averse workers for income variation together with high wage growth caused by the workers' participation constraint lead to a strong upward compression of the wage distribution from the bottom, and thus, to a large decrease in inequality. The decline in the value of outside job offers due to lower firm profits compresses the wage distribution at both ends, but leads only to a moderate decrease in wage inequality. By contrast, changes in effort levels prescribed to workers in response to increased effort costs result in a large increase in wage inequality. In particular, the additional costs due to the workers' participation constraint lead to large decreases in effort at low levels of lifetime utility, and, in consequence, to a strong expansion of the wage distribution at the lower end.

The decomposition also reveals that constraints on incentive provision due to workers' limited commitment play a quantitatively important role for the overall effect on wage inequality. Through the corresponding adjustments of effort and wage levels, the presence of moral hazard has a particularly strong impact at low levels of lifetime utility. This turns out to be the main force behind the large inequality increase in the lower half of the wage distribution.

The remainder of the paper is organized as follows: After a review of related literature, Section 2 presents the model framework, states the optimal contracting problem, and provides two theoretical results. In Section 3 we consider a simplified version of the model and illustrate analytically key parts of the channels through which moral hazard impacts on the wage distribution. Section 4 returns to the full model and describes its calibration, while Section 5 presents the results of the quantitative analysis together with a discussion of the underlying mechanisms. Section 6 concludes the paper. Some analytical results and a more extensive data description are presented in the appendices.

1.1 Related literature

The present paper closely relates to at least two strands of the literature. First, it contributes to the study of optimal dynamic contracts in the context of labor markets. Thomas and Worrall (1988) provide an early application of dynamic contracts to employment relations under limited commitment, a feature that is present in our model too. Furthermore, our analysis of optimal wage contracts in a setting with repeated moral hazard builds on the seminal work of Rogerson (1985a) and Spear and Srivastava (1987). To the best of our knowledge, there are only a few other papers that study a dynamic moral hazard problem in a labor market context with on-the-job search. Manoli and Sannikov (2005) analyze properties of the optimal contract in a continuous-time environment where bidding-strategies of firms are themselves a device for incentive provision. The authors focus on the analytical characterization of bidding strategies and job changes as well as on ex-post inefficiencies. However, they neither investigate the role of moral hazard in shaping the cross-sectional wage distribution, nor do they provide a quantitative framework for analysis. Tsuyuhara (2013) and Lentz (2013) also combine dynamic moral hazard with an on-the-job search environment. In the former paper, in contrast to our framework, search is directed and employers are assumed not to react to outside job offers received by their workers. Moreover, low productivity realizations always lead to the dissolution of a match, therefore the model cannot reproduce negative wage changes within a job. The main focus of the study is on the longitudinal characteristics of the optimal wage contract. In the latter paper, the moral hazard problem is assumed to arise with respect to workers' search effort on the job, and not with respect to work effort as in our analysis.

Second, our paper contributes to the analysis of sources of wage inequality. In a recent article, Hornstein et al. (2011) assess the degrees to which different versions of job search models can account for empirically observed levels of wage dispersion. They find that, in this context, models of on-the-job search with employer competition as in Postel-Vinay and Robin (2002) are among the most promising approaches. In the present paper, we extend this framework by introducing an informational friction leading to moral hazard as an additional source of wage inequality. Another set of related papers is Low et al. (2010) and Altonji et al. (2012). Similar to the present analysis, these two papers emphasize that individuals' earnings dynamics are composed of changes within and between jobs. However, they do not associate within-job wage dynamics with incentive provision. Finally, Lemieux et al. (2009) empirically study the relation between performance pay and wage inequality. Their analysis is based on the assumption that performance pay arises because of ex-ante uncertainty about a worker's skills, and not because of asymmetric information.

2 The Model

This section presents the model framework for our analysis. After outlining the basic environment, we describe the optimal contract design problem. With a view to the quantitative analysis in Section 5, we then state two theoretical results and provide a definition of equilibrium.

2.1 The environment

We consider an economy that is populated by a unit mass of ex-ante identical workers and a continuum of ex-ante identical firms. Time is discrete with an infinite horizon and is indexed by $t = 0, 1, 2, \ldots$. Workers are risk-averse, whereas firms are risk-neutral, and each firm can employ only one worker. The number of workers and of firms is invariant over time as a consequence of the following replacement rule: When a worker dies, he is replaced by a newborn worker who is unemployed, and when a firm goes out of business, it is replaced by a newly established firm. Whenever a new firm is established, it can hire a worker out of unemployment or of employment with another firm. If it fails to do so, it disappears. Finally, workers and firms discount the future with a common discount factor $\beta \in (0, 1)$.

2.1.1 Firms

When a firm is matched with a worker, the pair is characterized by a match-specific productivity level $z \in \mathcal{Z} = \{z_1, z_2, \ldots, z_N\}$ where $z_n < z_{n+1}$. The value is drawn from a distribution with cumulative distribution function $F(\cdot)$ at the time when the firm and the worker meet and remains constant over time. Firm output y is a function of both the fixed match-specific productivity level z and a stochastic worker-specific productivity factor A which depends on the effort ϵ spent by the worker in a given period. Firms use a production technology according to which output is given by

$$y = Y(z, A) = zA \tag{1}$$

The worker-productivity factor A can take on two different values, depending stochastically on the worker's effort level ϵ:

$$A = \begin{cases} A^+ & \text{with probability} \quad \pi(\epsilon) \\ A^- & \text{with probability} \quad 1 - \pi(\epsilon) \end{cases} \tag{2}$$

where $A^+ > A^-$ and the function $\pi(\cdot)$ is continuous, strictly increasing and strictly concave $(\pi'(\cdot) > 0, \pi''(\cdot) < 0)$ with $\pi(\epsilon) \in [0, 1]$ for $\epsilon \in [0, \bar{\epsilon}]$. Firms maximize their expected present-value profits. Worker-firm matches are exogenously destroyed with probability δ at the end of a period.

2.1.2 Workers

A worker can be either unemployed or employed by one of the firms active in a given period. Each period, worker i derives utility from consumption c_{it} and suffers disutility from spending effort ϵ_{it} when working. The period utility $u(c)$ from consumption is assumed to be a continuous, strictly increasing and strictly concave function $(u'(\cdot) > 0, u''(\cdot) < 0)$ which is bounded from above by zero. By contrast, the period disutility $g(\epsilon)$ from effort is assumed to be a continuous, strictly increasing and convex function $(g'(\cdot) > 0, g''(\cdot) \geq 0)$ with $g(\epsilon) \geq 0$ for $\epsilon \in [0, \bar{\epsilon}]$. The probability for a worker to die within a period is $(1 - \psi)$. Workers maximize their expected lifetime utility.

While unemployed, a worker enjoys a level of consumption $b > 0$ which is the same for all workers and also invariant over time. Assuming that workers do not save, an unemployed worker's period utility thus amounts to $u(b)$. In unemployment, the probability of receiving exactly one job offer within a given period is λ_u. While employed, a worker spends effort ϵ_{it} and receives wage w_{it} within a time period, implying a period utility of $u(w_{it}) - g(\epsilon_{it})$. For an employed worker, the probability of receiving exactly one outside job offer within a given period is λ_e. Finally, a worker whose current match is exogenously destroyed immediately receives a new job offer with probability λ_r. The match productivity associated with any job offer is a random draw from the distribution $F(\cdot)$.

2.1.3 Employment contracts and firm competition

When a worker and a firm meet, they draw their potential match productivity. The firm then decides whether to make a take-it-or-leave-it offer to the worker in terms of a long-term employment contract.

We assume that the contractual relationship between a worker and a firm has a number of particular features. First, the firm cannot cannot observe the level of effort spent by its worker in a given period, but only the realization of the worker-specific productivity A_{it}. Second, commitment to the contractual arrangement is asymmetric: While a worker may walk away from the contract at any time, a firm can commit to a compensation plan as long as expected profits are non-negative. Third, a firm cannot observe an outside job offer to its employee unless he reports it. And, fourth, if a worker reports an outside offer, the current and the potential future employers start competing for the worker by offering new contracts. A firm's strategy for such competition is not part of the labor contract.[5]

In the present framework, there are two ways in which a worker can break up an employment contract: On the one hand, he may quit to unemployment at any time, in which case his current employer goes out of business. On the other hand, he may report an outside job offer he has received and thereby trigger firm competition. In this case, the original contract becomes void and the two firms enter a Bertrand competition in terms of expected liftime utility the respective contracts offer to the worker.

2.1.4 Timing of events

The timing of events within a model period is as follows: A worker i employed in a match of productivity z receives wage w_{it} and spends effort ϵ_{it}. Then output y_{it} is produced, revealing the period realization A_{it}. With probability λ_e, the worker meets another firm, and their potential match productivity $\tilde{z}$ is drawn from $F(\cdot)$. If he reports the outside job offer to his current employer, the two firms start competing by offering new contracts which are to start at the beginning of the

[5] See Manoli and Sannikov (2005) for a continuous-time framework where firms can commit ex-ante to bidding strategies. In our discrete-time setup, this would unduly complicate the model. Moreover, we believe that ex-ante commitment to bidding strategies is much less plausible than ex-ante commitment to a wage path conditional on non-negative profits.

next period. The worker chooses which firm to work for in the future. Within the same time, an unemployed worker consumes b. With probability λ_u, he is contacted by a firm, and their potential match productivity is drawn from $F(\cdot)$. The firm decides whether to offer a contract to the worker. If the match is formed, the employment contract starts with the beginning of the next period. At the end of the period, existing worker-firm matches are exogenously destroyed with probability δ. A laid-off worker immediately meets a new firm with probability λ_r, the associated match productivity is drawn from $F(\cdot)$, and production starts with the beginning of the next period if the match is formed. Finally, workers die with probability $(1-\psi)$ and are replaced by newborn workers who are unemployed. Firms that have lost their workers or that have failed to recruit are replaced by newly established firms.

2.2 Optimal contract design

Firms offer to workers long-term contracts that are optimally designed with respect to the non-observability of worker effort and with respect to the facts that workers may quit and firms cannot commit to wage payments associated with negative expected profits. In equilibrium, firms offer wages which depend on the history of a worker's output realizations in order to provide optimal incentives for effort. Moreover, contracts ensure that a worker never wants to quit to unemployment and that firms make non-negative profits at any point in time. When designing a contract, a firm takes as given both its own and a potential competitor's ex-post optimal bidding strategies together with the optimal strategies of workers for reporting outside offers.

2.2.1 Towards a recursive formulation

Similar to other dynamic contracting environments, an employment contract in our setup is an infinite-dimensional object. Consider a worker in period t who is employed in a match of productivity z_t and is offered an alternative job with match productivity $\tilde{z}_t$. For any period $\tau = t+1, t+2, \ldots$ in the future, denote the continuation history of the worker's history of productivity realizations up to date τ by $\mathbf{A}^{\tau-1} \equiv \{A_j\}_{j=t+1}^{\tau-1}$ (and define $\mathbf{A}^t \equiv 0$). A contract offered by the outside competitor specifies, conditional on potential match productivity $\tilde{z}_t$ and the productivity of the worker's current match z_t, for all future dates $\tau > t$ and all possible histories $\mathbf{A}^{\tau-1}$, a period wage w_τ and a period effort level ϵ_τ, thus consisting of a sequence of functions $\left\{w_\tau(\mathbf{A}^{\tau-1}, \tilde{z}_t, z_t), \epsilon_\tau(\mathbf{A}^{\tau-1}, \tilde{z}_t, z_t)\right\}_{\tau=t+1}^{\infty}$.[6] Analogously, a contract offered by the current employer can be written as $\left\{w_\tau(\mathbf{A}^{\tau-1}, z_t, \tilde{z}_t), \epsilon_\tau(\mathbf{A}^{\tau-1}, z_t, \tilde{z}_t)\right\}_{\tau=t+1}^{\infty}$, where the roles of the current and the alternative match productivities are interchanged. For the case of an unemployed worker in period t who is offered a job with match productivity $\tilde{z}_t$, defining his current match productivity as $z_t \equiv 0$ one can write a contract offered by the potential employer as $\left\{w_\tau(\mathbf{A}^{\tau-1}, \tilde{z}_t, 0), \epsilon_\tau(\mathbf{A}^{\tau-1}, \tilde{z}_t, 0)\right\}_{\tau=t+1}^{\infty}$.

[6] Given our assumptions on the timing of events, wage and effort levels specified by a contract for a given period depend on the worker's productivity history with the firm up to one period before. Wage and effort levels in the first period of a contract are therefore functions of $(\tilde{z}_t, z_t)$ only. For notational convenience, we define $\mathbf{A}^t \equiv 0$, which yields $w_{t+1} = w_{t+1}(0, \tilde{z}_t, z_t)$ and $\epsilon_{t+1} = \epsilon_{t+1}(0, \tilde{z}_t, z_t)$.

In all cases, within a contract, the actions prescribed to the worker and the payoffs to both the worker and the firm at any point in time thus depend on the whole history of previous actions, payoffs, and productivity realizations. We use the *promised utility approach*, developed, among others, by Spear and Srivastava (1987), to formulate the firm's problem of designing an optimal contract under repeated moral hazard recursively. In this approach, all relevant aspects of history are summarized in a single variable, namely, the expected lifetime utility for the worker implied by the contract. This utility level, denoted by U, together with current match productivity z, fully characterizes an employed worker's state.

Before writing down a full recursive formulation of an optimal long-term contract between a worker and a firm, we need to develop some notation and introduce a number of additional variables related to expected lifetime utility or profits. Starting from an unemployed worker's perspective, we denote his expected lifetime utility by U^n. For an employed worker, $U^i(U,z)$ denotes the continuation value of the current labor contract in a z-type match under the conditions that the worker's current productivity realization is A^i, with $i \in \{+,-\}$, and that he did not receive an outside job offer. For notational convenience, we will sometimes use U^i instead of $U^i(U,z)$. By contrast, $U_o(U^i,z,\tilde{z})$ represents the continuation value to the worker when his productivity realization is A^i and he receives an outside offer associated with match productivity $\tilde{z}$. The value of $U_o(U^i,z,\tilde{z})$ is equal to the continuation value of the current labor contract if the worker decides not to report the outside offer. If he does, however, report the offer, $U_o(U^i,z,\tilde{z})$ takes on the value of the expected lifetime utility offered by a new contract which is the outcome of Bertrand competition between the two firms.

Switching to the viewpoint of a firm, the value of a contract, when match productivity is z, of a contract delivering lifetime utility U in an optimal way is denoted by $V(U,z)$. The function $V_o(U^i,z,\tilde{z})$ represents the continuation value to the firm when the worker's current productivity realization is A^i and he has received an outside offer associated with match productivity $\tilde{z}$. In analogy to utility U_o, the value of V_o is the continuation value to the firm of the current labor contract for the case that the worker does not report the outside offer he has received. If he does, however, report the offer, $V_o(U^i,z,\tilde{z})$ assumes the value to the firm of the outcome of the Bertrand competition with the outside competitor.

With the notation developed up to this point, the value $V(U,z)$ of a contract to a firm can be expressed as

$$\begin{aligned} V(U,z) &= z\Big[A^+\pi(\epsilon)+A^-(1-\pi(\epsilon))\Big]-w \\ &\quad +\beta\psi(1-\delta)\Bigg\{(1-\lambda_e)\Big[V(U^+,z)\pi(\epsilon)+V(U^-,z)(1-\pi(\epsilon))\Big] \\ &\quad +\lambda_e\sum_{\tilde{z}\in\mathcal{Z}}\Big[V_o(U^+,z,\tilde{z})\pi(\epsilon)+V_o(U^-,z,\tilde{z})(1-\pi(\epsilon))\Big]f(\tilde{z})\Bigg\} \end{aligned} \tag{3}$$

while the value of a contract to the worker can be written as

$$u(w) - g(\epsilon) + \beta\psi\delta U^n + \beta\psi(1-\delta)\Bigg\{(1-\lambda_e)\Big[U^+\pi(\epsilon) + U^-(1-\pi(\epsilon))\Big]$$
$$+\lambda_e \sum_{\tilde{z}\in\mathcal{Z}} \Big[U_o(U^+, z, \tilde{z})\pi(\epsilon) + U_o(U^-, z, \tilde{z})(1-\pi(\epsilon))\Big] f(\tilde{z})\Bigg\} \tag{4}$$

A recursive contract in the present framework is defined as follows:

Definition 1. *A recursive contract $\mathcal{C}$ is a collection of functions that, for each pair (U, z) of promised utility U and match productivity level z, specify a prescribed worker effort $\epsilon(U, z)$, a current period wage $w(U, z)$, and continuation values $\{U^+(U, z), U^-(U, z)\}$ for the worker who attains productivity realization $A \in \{A^+, A^-\}$ and does not receive an outside job offer.*

2.2.2 Continuation values in the case of an outside offer

As specified before, when an employed worker reports an outside offer to his firm, the current and the potential future employers enter a Bertrand competition for the worker. In this competition, the critical utility level $U^*(z)$, defined by the condition of zero expected profits

$$V(U^*(z), z) = 0 \tag{5}$$

plays a decisive role: A firm with match productivity z is in general willing to offer a level of lifetime utility U up to a maximum equal to the break-even value of $U^*(z)$.[7] Therefore, if the lifetime utility promised by the current contract is lower than the break-even utilities of both firms, a worker triggering Bertrand competition will be offered a new contract which promises him the lower of the two critical utility levels.

Based on the above arguments and the fact that, due to a firms' limited commitment, a worker's lifetime utility from a given contract will never exceed the current employer's break-even value $U^*(z)$, we can distinguish four different cases of relationships between values of U^i, $U^*(z)$, and $U^*(\tilde{z})$. The resulting classification is presented below. It includes specifications of the worker's decisions on whether to trigger firm competition and on whether to stay with his current employer or move to the job offered by the competing firm. It also states the corresponding continuation values for the worker and the current employer.

The first case reflects the situation of the competitor firm having a higher match productivity than the current employer:

Case 1: $U^i \leq U^*(z) < U^*(\tilde{z})$

The worker discloses the offer, moves to the new employer and gets $U_o = U^*(z)$. The firm currently employing the worker disappears ($V_o = 0$).

[7] Since $V(U, z)$ is strictly decreasing in U and continuous on the relevant range of the domain, for a given z, the value $U^*(z)$ is unique and well-defined by equation (5).

The second case covers the situation where both firms have equal match productivities:

Case 2: $U^i \leq U^*(z) = U^*(\tilde{z})$

The worker discloses the offer and gets $U_o = U^*(z)$. With probability 1/2 he stays with his current employer who gets $V_o = V(U^*(z), z) = 0$. With probability 1/2 the worker moves to the new firm, in which case the incumbent firm disappears ($V_o = 0$).

The last two cases reflect the situation where the competitor firm has a lower match productivity than the current employer:

Case 3: $U^i \leq U^*(\tilde{z}) < U^*(z)$

The worker discloses the offer, stays with his current employer, and gets $U_o = U^*(\tilde{z})$. The current employer gets $V_o = V(U^*(\tilde{z}), z)$.

Case 4: $U^*(\tilde{z}) < U^*(z)$ and $U^*(\tilde{z}) < U^i$

The worker does not disclose the offer and stays with his current employer. The worker gets $U_o = U^i$, and the current employer gets $V_o = V(U^i, z)$.

The continuation value U_o to a worker employed in a match of productivity z, who would get utility U^i under his current labor contract, and who has received an outside offer associated with a match productivity $\tilde{z}$, can be summarized in the following expression:

$$U_o(U^i, z, \tilde{z}) = \max\left\{ U^i(U, z), \min\left[U^*(z), U^*(\tilde{z}) \right] \right\} \quad . \tag{6}$$

The corresponding continuation value to the current employer is given by

$$V_o(U^i, z, \tilde{z}) = \begin{cases} V(U^i(U, z), z) & \text{if} \quad U^*(\tilde{z}) < U^i(U, z) \\ V(U^*(\tilde{z}), z) & \text{if} \quad U^i(U, z) \leq U^*(\tilde{z}) < U^*(z) \\ 0 & \text{otherwise} \end{cases} \quad . \tag{7}$$

2.2.3 Constraints on the contract

The contract offered by a firm to a worker is designed optimally, given the particular features of the current setup. In response to the informational friction of non-observability of worker effort, a firm prescribes effort levels that the worker would optimally choose under the conditions of the contract. This requirement is introduced into the firm's optimization problem in the form of an incentive-compatibility constraint.

An incentive-compatible effort level ϵ maximizes a worker's expected lifetime utility associated with the contract, that is,

$$\begin{aligned} \epsilon \in \underset{\hat{\epsilon} \in [0, \bar{\epsilon}]}{\operatorname{argmax}} \quad & u(w) - g(\hat{\epsilon}) + \beta\psi\delta U^n \\ & + \beta\psi(1-\delta)\left\{ (1-\lambda_e)\left[U^+ \pi(\hat{\epsilon}) + U^-(1-\pi(\hat{\epsilon})) \right] \right. \\ & \left. + \lambda_e \sum_{\tilde{z} \in \mathcal{Z}} \left[U_o(U^+, z, \tilde{z})\pi(\hat{\epsilon}) + U_o(U^-, z, \tilde{z})(1-\pi(\hat{\epsilon})) \right] f(\tilde{z}) \right\} \end{aligned} \tag{8}$$

The first-order condition corresponding to an interior solution of this maximization problem is

$$\begin{aligned} g'(\epsilon) &= \pi'(\epsilon)\beta\psi(1-\delta)\Bigg\{(1-\lambda_e)\Big[U^+ - U^-\Big] \\ &\quad + \lambda_e \sum_{\tilde{z}\in\mathcal{Z}} \Big[U_o(U^+, z, \tilde{z}) - U_o(U^-, z, \tilde{z})\Big] f(\tilde{z})\Bigg\} \end{aligned} \tag{9}$$

We use equation (9) as the incentive-compatibility constraint (ICC).[8] This relationship, characterizing the set of incentive-compatible effort levels, expresses the standard condition that at the optimum marginal benefit and marginal cost of effort have to be equal. In addition, equation (9) indicates that potential outside offers make the provision of incentives more expensive relative to a situation in which workers cannot search on the job. In other words, the higher the probability of outside job offers (λ_e), the higher is the spread in continuation values from the current contract ($U^+ - U^-$) that is needed to extract a certain level of effort. The reason is that, when the worker receives a relevant outside offer, the firm has only little control over his continuation utility.

Limited commitment on both the worker's and the firm's side is taken into account in the optimal contract, too. On the one hand, firms want to keep workers from quitting to unemployment, and therefore always promise their workers at least as much utility from the contract as they would when unemployed.[9] Hence, the worker's participation constraint (WPC) requires that the value of a contract to the worker at any point in time weakly exceeds the value of unemployment U^n. On the other hand, firms take into account that they cannot commit to wage payments such that profits are negative. The firm's participation constraint (FPC) thus requires that the value of a contract to the worker does not exceed the firm's break-even level of utility $U^*(z)$ given the current match productivity.

Furthermore, the contract has to deliver at least the expected lifetime utility promised to the worker. This is captured by the following promise-keeping constraint (PKC):

$$\begin{aligned} U &\leq u(w) - g(\epsilon) + \beta\psi\delta U^n \\ &\quad + \beta\psi(1-\delta)\Bigg\{(1-\lambda_e)\Big[U^+\pi(\epsilon) + U^-(1-\pi(\epsilon))\Big] \\ &\quad + \underbrace{\lambda_e \sum_{\tilde{z}\in\mathcal{Z}} \Big[U_o(U^+, z, \tilde{z})\pi(\epsilon) + U_o(U^-, z, \tilde{z})(1-\pi(\epsilon))\Big] f(\tilde{z})}_{\text{expected continuation value in case of outside offer}}\Bigg\} \end{aligned} \tag{10}$$

The possibility of exogenous job destruction and of outside job offers triggering firm competition are the sources of special features of the promise-keeping constraint in the present setup. First, the value of unemployment U^n enters the constraint as a component which the firm cannot influence at all. Second, and as mentioned above, the expected continuation value for a worker who has

[8] The first-order approach is valid under the standard conditions provided by Rogerson (1985b). In particular, given our assumptions on the properties of $\pi(\epsilon)$, the worker's problem of effort choice satisfies the monotone likelihood-ratio and the convex distribution function conditions.

[9] In the present setup, firms whose productivity is too low to make positive profits at U^n will never form a match with a worker. In consequence, in equilibrium all operating firms make positive profits at U^n, and a worker's endogenous quit to unemployment is always inefficient.

received an outside offer is subject to only little control by the firm. Moreover, this component of the promise-keeping constraint varies between firms with different levels of match productivity. The reason is that the productivity level of a worker's current match determines the upper bound for the increase in lifetime utility the worker can obtain from firm competition. In particular, the higher the current employer's match productivity level, the higher is the worker's expected continuation value in case an outside offer arrives. Consequently, a firm with high match productivity faces a more relaxed promise-keeping constraint than one with a low-productivity match in the sense that, keeping everything else equal, it can satisfy the constraint by paying a lower wage.

Since the value of unemployment U^n affects the promise-keeping constraint and the worker's participation constraint, the term needs to be specified in order to be able to fully describe the contractual problem. Under our assumption that workers have no bargaining power, a firm offers to an unemployed worker the value of unemployment. Therefore U^n satisfies

$$U^n = u(b) + \beta\psi\Big[(1-\lambda_u)U^n + \lambda_u U^n\Big] \tag{11}$$

and the value of unemployment can be expressed as

$$U^n = \frac{u(b)}{1-\beta\psi} \tag{12}$$

As the utility promised to a worker has to satisfy the participation constraint, the above expression is also a lower bound for U.

Finally, the contract must guarantee feasibility of delivering the utility promised to the worker. Given that the period utility of an employed worker, $u(c)-g(\epsilon)$, is bounded from above by zero, this value must also be an upper bound on the expected lifetime utility U. In addition, the presence of exogenous separations imposes an even tighter upper bound $\bar{U}$ on promised utility. It can be derived from the expression of a worker's utility from a contract, given by (4), in the following way:

$$\begin{aligned}\underbrace{u(w)-g(\epsilon)}_{\leq 0} &+\beta\psi\delta U^n + \beta\psi(1-\delta)\Bigg\{(1-\lambda_e)\Big[\underbrace{U^+}_{\leq\bar{U}}\pi(\epsilon) + \underbrace{U^-}_{\leq\bar{U}}(1-\pi(\epsilon))\Big] \\ &+\lambda_e\sum_{\tilde{z}\in\mathcal{Z}}\Big[\underbrace{U_o(U^+,z,\tilde{z})}_{\leq\bar{U}}\pi(\epsilon) + \underbrace{U_o(U^-,z,\tilde{z})}_{\leq\bar{U}}(1-\pi(\epsilon))\Big]f(\tilde{z})\Bigg\}\end{aligned} \tag{13}$$

Given that the whole expression (13) is also bounded from above by $\bar{U}$, this leads to the equation

$$\bar{U} = \beta\psi\delta U^n + \beta\psi(1-\delta)\bar{U} \tag{14}$$

Substituting (12) then yields the expression

$$\bar{U} = \frac{\beta\psi\delta}{1-\beta\psi(1-\delta)}\frac{u(b)}{1-\beta\psi} < 0 \tag{15}$$

The corresponding feasibility constraint (FC) requires that the value of a contract to the worker at any point in time does not exceed the upper bound $\bar{U}$.

2.2.4 The optimization problem

Using the components defined and the functions outlined previously, the contractual problem can be stated as the following functional equation problem the optimal contract $\mathcal{C}^*$ has to solve:

$$\begin{aligned} V(U,z) \;=\; & \max_{\{w,\epsilon,U^+,U^-\}} z\Big[A^+\pi(\epsilon) + A^-(1-\pi(\epsilon))\Big] - w \\ & +\beta\psi(1-\delta)\Bigg\{(1-\lambda_e)\Big[V(U^+,z)\pi(\epsilon) + V(U^-,z)(1-\pi(\epsilon))\Big] \\ & +\lambda_e \sum_{\tilde{z}\in\mathcal{Z}} \Big[V_o(U^+,z,\tilde{z})\pi(\epsilon) + V_o(U^-,z,\tilde{z})(1-\pi(\epsilon))\Big] f(\tilde{z})\Bigg\} \end{aligned} \tag{16}$$

subject to the promise-keeping constraint (PKC)

$$\begin{aligned} U \;\leq\; & u(w) - g(\epsilon) + \beta\psi\delta U^n \\ & +\beta\psi(1-\delta)\Bigg\{(1-\lambda_e)\Big[U^+\pi(\epsilon) + U^-(1-\pi(\epsilon))\Big] \\ & +\lambda_e \sum_{\tilde{z}\in\mathcal{Z}} \Big[U_o(U^+,z,\tilde{z})\pi(\epsilon) + U_o(U^-,z,\tilde{z})(1-\pi(\epsilon))\Big] f(\tilde{z})\Bigg\} \quad , \end{aligned} \tag{17}$$

the incentive-compatibility constraint (ICC)

$$\begin{aligned} g'(\epsilon) \;=\; & \pi'(\epsilon)\beta\psi(1-\delta)\Bigg\{(1-\lambda_e)\Big[U^+ - U^-\Big] \\ & +\lambda_e \sum_{\tilde{z}\in\mathcal{Z}} \Big[U_o(U^+,z,\tilde{z}) - U_o(U^-,z,\tilde{z})\Big] f(\tilde{z})\Bigg\} \quad , \end{aligned} \tag{18}$$

the participation constraints of the worker (WPC) and the firm (FPC)

$$U^n \leq U^i \leq U^*(z) \quad \text{for} \quad i \in \{+,-\} \quad , \tag{19}$$

the feasibility constraints (FC)

$$U^i \leq \bar{U} \quad \text{for} \quad i \in \{+,-\} \quad , \tag{20}$$

and

$$w \geq 0 \tag{21}$$

$$0 \leq \epsilon \leq \bar{\epsilon} \quad . \tag{22}$$

2.3 On continuation values in the absence of outside offers

The following two lemmas provide results on the relationship between continuation values U^i, $i \in \{+,-\}$, for the employed worker under his current labor contract, given that he has not received any outside job offer. Lemma 1 establishes that under moral hazard a firm can only extract positive effort from its worker by promising him a strictly higher continuation utility after a high output realization than after a low one. In other words, positive worker effort requires a spread in wages between the two possible output realizations, and therefore a worker's wage depends on the history of productivity shocks.

Lemma 1. *Under the assumption that worker effort is not observable, that is, for the optimization problem (16) with constraints (17) to (22), the following result on continuation values $U^i(U,z)$ holds: If at a given state (U,z) the level of worker effort $\epsilon(U,z)$ is positive, then the worker's continuation value under high productivity realization $U^+(U,z)$ exceeds the corresponding value $U^-(U,z)$ associated with low realization, in short: $\epsilon(U,z) > 0 \Rightarrow U^+(U,z) > U^-(U,z)$.*

Proof. If effort is not observable, the incentive-compatibility constraint (18) must hold. Since $g'(\epsilon) > 0$, the right-hand side of (18) must be positive. As, for any pair $(z,\tilde{z})$, the function $U_o(U^i,z,\tilde{z})$ is non-decreasing in U^i, this requires that $U^+(U,z) > U^-(U,z)$. □

By contrast, Lemma 2 shows that under observable effort, a worker's continuation utility is independent of the output realization, that is, the risk-neutral firm fully insures the risk-averse worker against the uncertainty of effort-dependent productivity.

Lemma 2. *Under the assumption that worker effort is observable, that is, for the optimization problem (16) with constraints (17) and (19) to (22) – and without the incentive-compatibility constraint (18) – the following result on continuation values $U^i(U,z)$ holds: If the firm's value function $V(U,z)$ is concave in U, then the function $U^+(U,z)$, representing the worker's continuation value under high productivity realization, is identical with the corresponding function $U^-(U,z)$ associated with low realization, in short: $V(U,z)$ concave in $U \Rightarrow U^+(U,z) = U^-(U,z) \equiv U'(U,z)$.*

Proof. See Appendix A.1. □

2.4 Stationary equilibrium

In the quantitative analysis to follow, we focus on the properties of stationary equilibria of the present model. Let l denote the labor market status of a worker in a given period, where $l = 1$ if he is employed and $l = 0$ if he is unemployed. Further, let $\mu(l,U,z)$ denote the distribution of workers in a given period over labor market statuses, expected lifetime utilities, and firm productivities. A stationary equilibrium in the present framework can then be defined as follows:

Definition 2. *A stationary equilibrium consists of a value function $V(U,z)$, policy functions $w(U,z)$, $\epsilon(U,z)$, $U^+(U,z)$, and $U^-(U,z)$, functions $U_o(U^i,z,\tilde{z})$ and $V_o(U^i,z,\tilde{z})$ specifying continuation values* *to employed workers and firms for the case of an outside offer, laws of motion* $M : \mu(l, U, z) \rightarrow \mu'(l, U, z)$*, and a distribution* $\mu_s(l, U, z)$ *such that:*

(i) Given $U_o(U^i, z, \tilde{z})$ *and* $V_o(U^i, z, \tilde{z})$*, the functions* $V(U, z)$*,* $w(U, z)$*,* $\epsilon(U, z)$*,* $U^+(U, z)$*, and* $U^-(U, z)$ *solve the contractual problem (16) subject to constraints (17) to (22).*

(ii) The functions $U_o(U^i, z, \tilde{z})$ *and* $V_o(U^i, z, \tilde{z})$ *are consistent with firms' optimal bidding behavior in the case of firm competition, with workers' optimal reporting of outside offers, and with workers' optimal decisions whether to stay with the current employer or move to the new job, and are given by equations (6) and (7).*

(iii) The laws of motion M *are generated by firms' and workers' optimal decision rules and the specification of exogenous shocks.*

(iv) The distribution $\mu_s(l, U, z)$ *is consistent with the laws of motion* M *and is stationary, that is,* $\mu_s(l, U, z) = M(\mu_s(l, U, z))$*.*

3 Analysis of a Simplified Model

In this section, we consider a simplified version of the model in which we can illustrate analytically key parts of the channels through which moral hazard affects the wage distribution. We make the following simplifying assumptions: First, all firms have the same productivity level $z = 1$, hence period output is either $A^+ > 1$ or $A^- < 1$. Second, we assume that effort can only take one out of two values $\{0, \epsilon^*\}$. We set the disutility from effort to $g(0) = 0$ and $g(\epsilon^*) = v > 0$, and the probability of a high worker productivity realization to $\pi(0) = 0$ and $\pi(\epsilon^*) = p > 0$. Third, we assume that firms can only commit to one-period contracts. Fourth, we change the timing within the period so that wages are paid to workers at the end of the period after the output realization has been observed. Finally, to render the analysis pertinent, we assume parameter values to be such that in all cases firms prescribe the high level of effort ϵ^* to their workers.

These changes imply that only two types of job offer histories for workers are relevant for employment contracts: A worker has either received an outside job offer since his last unemployment spell or he has not. Conditional on a worker's job offer history, a one-period contract specifies a wage level for each possible output realization. As a result, there are four equilibrium levels of wages under moral hazard: w_u^+ and w_u^- for workers leaving unemployment, and w_o^+ and w_o^- for workers who have already received an outside offer. When effort is observable, there are only two equilibrium levels of wages, $\hat{w}_u$ and $\hat{w}_o$. As before, U^n denotes the value to a worker of being unemployed. To denote a worker's lifetime utility after having received an outside offer, we use U_o for the case of moral hazard, and $\hat{U}_o$ for that of observable effort. A firm's period expected profits after a worker has received an outside offer is denoted by Π_o under moral hazard, and by $\hat{\Pi}_o$ under observable effort.

We start the exposition by analyzing the problem of a firm whose worker has received an outside offer. Since, in this case, the level of lifetime utility promised to the worker is determined by firm competition, a firm makes zero expected period profits on the contract. Under moral hazard, this condition reads as

$$\Pi_o = p\big[A^+ - w_o^+\big] + (1-p)\big[A^- - w_o^-\big] = 0 \tag{23}$$

and implies that the mean wage is equal to mean output, $pw_o^+ + (1-p)w_o^- = pA^+ + (1-p)A^-$. It should be noted that the same argument implies that the wage under observable effort is equal to mean output, $\hat{w}_o = pA^+ + (1-p)A^-$. In the case of moral hazard, wages moreover need to satisfy the incentive-compatibility constraint

$$\begin{aligned} pu(w_o^+) + (1-p)u(w_o^-) - v + \beta\psi\Big\{\delta U^n + (1-\delta)U_o\Big\} \quad &\geq \\ u(w_o^-) + \beta\psi\Big\{\delta U^n + (1-\delta)U_o\Big\} & \end{aligned} \tag{24}$$

which, in equilibrium, will be satisfied with equality, and can then be reduced to

$$p\big[u(w_o^+) - u(w_o^-)\big] = v \tag{25}$$

These observations lead to the following results:

Lemma 3. *(i) At the top of the wage distribution, moral hazard leads to a mean-preserving spread in wages. (ii) Workers are unambiguously worse off under moral hazard than under observable effort after having received an outside offer.*

Proof. (i) We have shown above that $pw_o^+ + (1-p)w_o^- = \hat{w}_o$. Moreover, (25) implies that $w_o^+ > w_o^-$. (ii) Given that workers face a mean-preserving spread in period wages and the same prospect in case they lose their job, this result is a trivial consequence of risk aversion. More formally, the value of employment after having received an outside offer, under moral hazard and under observable effort, respectively, are given by

$$U_o = \frac{pu(w_o^+) + (1-p)u(w_o^-) - v + \beta\psi\delta U^n}{1 - \beta\psi(1-\delta)} \quad \text{and} \quad \hat{U}_o = \frac{u(\hat{w}_o) - v + \beta\psi\delta U^n}{1 - \beta\psi(1-\delta)} \tag{26}$$

Part (*i*) and strict concavity of $u(\cdot)$ imply that $U_o < \hat{U}_o$. □

The intuition behind these results is simple. At the top of the wage distribution, firms need to break even; therefore, the expected wage is equal to expected output in both scenarios. Given that the level of effort is the same, average output and average wage are the same in both economies. Moreover, as long as effort is constant across the two scenarios, these results for workers who have received an outside offer can easily be generalized to a setting with dynamic contracts, different timing of wage payments, and multiple firm productivity levels.

We now turn to the problem of a firm employing a worker who has not yet received an outside offer. Due to the assumption that workers have no bargaining power, the value of unemployment U^n is the same in both scenarios. Under moral hazard, the promise-keeping constraint is given by

$$U^n = pu(w_u^+) + (1-p)u(w_u^-) - v + \beta\psi\Big\{\delta U^n + (1-\delta)\big[\lambda_e U_o + (1-\lambda_e)U^n\big]\Big\} \tag{27}$$

which can be rearranged to yield

$$U^n = \frac{pu(w_u^+) + (1-p)u(w_u^-) - v + \beta\psi(1-\delta)\lambda_e U_o}{1 - \beta\psi\big[\delta + (1-\delta)(1-\lambda_e)\big]} \tag{28}$$

Similar algebra leads to the following expression for U^n under observable effort:

$$U^n = \frac{u(\hat{w}_u) - v + \beta\psi(1-\delta)\lambda_e \hat{U}_o}{1 - \beta\psi\big[\delta + (1-\delta)(1-\lambda_e)\big]} \tag{29}$$

Equations (28) and (29) inform us about the following implication of moral hazard for wages at the bottom of the distribution:

Lemma 4. *The expected wage of workers who have not yet received an outside offer is higher under moral hazard than under observable effort.*

Proof. Since $U_o < \hat{U}_o$, (28) and (29) imply that $pu(w_u^+) + (1-p)u(w_u^-) > u(\hat{w}_u)$. From concavity of $u(\cdot)$ it follows that $pw_u^+ + (1-p)w_u^- > \hat{w}_u$. □

It should be noted that the expected wage of workers leaving unemployment is higher under moral hazard for two reasons. First, as stated in Lemma 3 above, workers face worse future prospects under moral hazard than when effort is observable, that is, $U_o < \hat{U}_o$. In consequence, they require higher contemporaneous utility and therefore a higher expected contemporaneous wage in order to accept a job out of unemployment. However, the average wage at the bottom of the distribution would still be higher under moral hazard, even if it were the case that $U_o = \hat{U}_o$. The reason is that wages need to satisfy an incentive-compatibility constraint corresponding to equation (25), which implies that $w_u^+ > w_u^-$. Due to concavity of utility, $pw_u^+ + (1-p)w_u^- > \hat{w}_u$ must then hold, even if $pu(w_u^+) + (1-p)u(w_u^-) = u(\hat{w}_u)$.

The above results show that introducing moral hazard affects the wage distribution in the simplified model both directly and indirectly. The direct effect consists of the fact that, both for workers who have received an outside offer and for those who have not, wages need to vary with output realizations in order to provide incentives. This wage variation obviously increases wage dispersion relative to the scenario with observable effort. At the top of the distribution, this is the only effect of moral hazard, since the average wage is the same in both scenarios. At the bottom of the distribution, however, additional channels lead to an increase in the average wage of workers who have not yet received an outside offer. This indirect effect reduces the difference in average wages between the two groups of workers. As a result, the overall impact of moral hazard on wage dispersion depends on the relative magnitude of the two effects.

In other words, the indirect effect of moral hazard leads to an upward compression of the wage distribution. This compression ultimately stems from the need to compensate risk-averse workers for wage variation. Interestingly, it is accompanied by a downward compression of lifetime utilities at the top of the distribution. While workers' expected lifetime utility U^n when leaving unemployment is the same in both scenarios, the maximal lifetime utility (after having received an outside offer) is strictly lower in the case of moral hazard. Given the set-up of firm competition, workers at the top of the distribution cannot be compensated for earnings variation by a higher expected wage. Instead, compensation takes place when workers leave unemployment, so that the top-down compression in lifetime utilities translates into an upward compression of wages at the bottom of the distribution.

The potentially most restrictive feature of the present analysis is the assumption of discrete and constant effort levels, both across job offer histories and across scenarios. It is straightforward to see that, if effort were continuous, moral hazard would lead to lower levels of effort, since incentive provision increases the costs of effort to firms. This would decrease the effort compensation component of wages, and thus impact on wage levels. Depending on whether the reduction in effort levels is larger at the top or the bottom of the distribution, this could either reinforce or partially offset the indirect compression effect of moral hazard on the wage distribution.

All the channels of impact of moral hazard described in the present simplified framework appear in the analysis of the full model too. In addition, the assumption of long-term contracts introduces more intricate dynamics of wages within a job even in the absence of outside offers. In the full model we moreover allow for continuous effort choice. This results in differences between the moral hazard and the observable effort scenarios in terms of implemented effort levels, which will in turn lead to differences in wage levels and in the dynamics of wages within a job.

4 Calibration

We now return to the full model of Section 2 and describe its calibration to U.S. data. Starting from some basic specifications, we move on to describing the data background to our calibration. The final part explains the approach to parameter selection and the choice of targets to match, presents the parameter values obtained, and discusses the fit of statistics between the model and the data.

4.1 Functional forms and distributions

We assume a worker's period utility from consumption to be a function with constant relative risk aversion,

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma} \quad , \tag{30}$$

where $\sigma > 1$. The period disutility from spending effort is given by a power function

$$g(\epsilon) = \epsilon^{\gamma} \tag{31}$$

with exponent $\gamma > 0$, while the probability of a high output realization as a function of effort is given by an exponential function

$$\pi(\epsilon) = 1 - \exp\{-\rho\epsilon\} \tag{32}$$

with parameter $\rho > 0$. We normalize the levels of effort-dependent worker productivity as follows:

$$\begin{aligned} A^- &= 1 \\ A^+ &= 1 + \Delta A \end{aligned} \tag{33}$$

where $\Delta A > 0$. Regarding match-specific productivity, we assume that values are drawn from a three-parameter Weibull distribution

$$z \sim WB(\zeta_0, \zeta_1, \zeta_2) \tag{34}$$

with shift parameter $\zeta_0 > 0$, shape parameter $\zeta_1 > 0$, and scale parameter $\zeta_2 > 0$. The corresponding cumulative distribution function is

$$F(z; \zeta_0, \zeta_1, \zeta_2) = \begin{cases} 1 - \exp\left\{ - \left(\frac{z - \zeta_0}{\zeta_2} \right)^{\zeta_1} \right\} & \text{if} \quad z \geq \zeta_0 \\ 0 & \text{if} \quad z < \zeta_0 \end{cases} . \tag{35}$$

For the numerical solution, we discretize the distribution $F(\cdot)$ with a total of 15 levels of match-specific productivity levels.

4.2 Data background

We obtain empirical statistics for the calibration from the 2004 panel of the Survey of Income and Program Participation (SIPP). The data set contains information at high frequencies of individuals' labor market histories and income sources for a representative sample of households in the United States. In particular, the SIPP collects detailed job-specific information for up to two wage and salary jobs per person in a given period (wave). This information allows for a distinction between different jobs that a person has held with different employers over the time span of the panel. In consequence, it is possible to quite reliably set apart job-to-job transitions from other types of labor market transitions. The structure of the data thus allows for a quantitative assessment of the dynamics of wages, both with regard to within-job changes and to changes between jobs upon job-to-job transition.

In the present exercise, we use data covering the period from January 2004 to December 2006 and restrict our attention to information on male workers between 20 and 65 years of age. A further restriction of our sample is that to individuals who have held at least one non-contingent, non-self-employed paid job during the time span of the panel. We classify individuals as employed or unemployed in a given month according to the labor market status they report in the second week of that month. Based on this classification, we compute monthly rates of transition between the states of employment and unemployment. In our measure of job-to-job transitions, we include workers who are classified as employed with two different main employers in two consecutive months, do not return to a previous employer, and have not been without a job and searching or on layoff in between.

For the estimation of residual wages, we further restrict the sample to individuals who are working full-time. Based on this sample, we run a pooled regression of log real hourly wages on the following variables: five educational groups, four region groups, a dummy for being non-white, and year dummies. The residuals from this regression are used to compute statistics of the cross-sectional wage distribution, as well as statistics on wage changes within jobs and between jobs upon job-to-job transitions. Appendix A.2 provides more detailed information on the data underlying our calibration.

4.3 Calibration strategy and results

Given the above assumptions on functional forms and on the distribution of match-specific productivity levels, values for the following fourteen parameters need to be selected:

1. β, the discount factor,
2. σ, the workers' coefficient of relative risk aversion,
3. ψ, the probability of survival of a worker,
4. δ, the probability of exogenous destruction of a worker-firm match,
5. λ_u, the probability for unemployed workers of receiving a job offer,
6. λ_e, the probability for employed workers of receiving an outside job offer,
7. λ_r, the probability for a worker of receiving a job offer immediately after his match has been destroyed,
8. γ, the power parameter in the function for disutility from spending effort,
9. ρ, the coefficient in the expression for the probability of high worker productivity realization as a function of effort,
10. ΔA, the difference between high and low realizations of worker productivity,
11. b, the level of consumption while unemployed,
12. and the shift, shape, and scale parameters $(\zeta_0, \zeta_1, \zeta_2)$ of the sampling distribution for match-specific productivity.

We choose the length of a time period in the model to be one quarter.[10] The discount factor (β) is set at 0.99, a value which corresponds to an annual interest rate of 4%, and the coefficient of workers' relative risk aversion (σ) at the value 2, which is standard in the quantitative macroeconomic literature. The survival probability of workers (ψ) is set at 0.994, a value which corresponds to an expected length of an individual's working life of forty years. The remaining parameters are selected with the objective of matching empirical statistics on the frequency of labor market transitions, individuals' wage changes both within and between jobs, and the mean of the cross-sectional residual wage distribution, by corresponding stationary equilibrium statistics of the model. In the following paragraphs, we discuss our choice of target statistics and their relation to model parameters.

[10] This is a compromise between using short time periods which correspond more closely to the high frequency of workers' labor market transitions and computational concerns in the numerical solution of the model.

Our first set of statistics comprises the frequencies of worker's transitions between the states of unemployment and employment, and between jobs. The model parameter most closely linked to the rate of employment-to-unemployment flows (τ_{eu}) is the probability of match destruction (δ), while the rate of unemployment-to-employment transitions (τ_{ue}) is mainly determined by the probability for an unemployed worker to receive a job offer (λ_u). Finally, the rate of job-to-job transitions (τ_{ee}) in the model is closely connected to the probability for an employed worker to receive an outside job offer (λ_e). With a view to matching these transition rates, we set δ at the value of 0.028, λ_u at the value of 0.62, and λ_e at the value of 0.095.

The second set of empirical targets consists of statistics on the distribution of individuals' log wage changes upon job-to-job transitions. In our model, workers who obtain a job offer immediately after match destruction typically receive a lower wage than in their previous match. We therefore include the fraction of job-to-job transitions associated with negative log wage changes ($\varpi_{-}^{bet}(\Delta \ln w) = P[\Delta \ln w^{bet} < 0]$) among the targets with a view to determining the probability of receiving a job offer immediately after layoff (λ_r). The mean and standard deviation of log wage changes upon job-to-job transitions ($\mu^{bet}(\Delta \ln w) = \mathbb{E}[\Delta \ln w^{bet}]$ and $\sigma^{bet}(\Delta \ln w) = SD[\Delta \ln w^{bet}]$) in the model are mainly determined by the distribution from which match productivities associated with job offers are drawn ($F(z)$ with parameters ζ_0, ζ_1, and ζ_2). We fix the shape parameter of the distribution ζ_1 at the value of 2, yielding a shape that is close to that of a log-normal distribution. Conditional on this, we set ζ_0 to 0.5 and ζ_2 to 0.4 with a view to matching the mean and standard deviation of wage changes associated with job-to-job transitions.[11]

Our third set of targets contains statistics on the distribution of log wage changes within a job. As discussed before in detail, incentive provision under moral hazard leads to output-dependent variation of wages within an employment contract. The model parameters that most directly affect the costs and benefits of incentive provision, and therefore wage dynamics within a job, too, are the power parameter in the disutility from effort (γ), the parameter governing the relation between effort and the probability of high worker productivity (ρ), and the difference between high and low worker productivity in the scale of match output (ΔA). In order to pin down values for these parameters, we include the fraction of negative log wage changes within a job ($\varpi_{-}^{win}(\Delta \ln w) = P[\Delta \ln w^{win} < 0]$) as well as the mean and standard deviation of wage changes ($\mu^{win}(\Delta \ln w) = \mathbb{E}[\Delta \ln w^{win}]$ and $\sigma^{win}(\Delta \ln w) = SD[\Delta \ln w^{win}]$) among our targets. Since it is, however, impossible to separate the effects of γ and of ρ on statistics of within-job wage changes, we fix γ at the value of 2. Conditional on this, we set ρ to the value of 3, and ΔA to the value of 1, with a view to matching the above mentioned statistics on within-job wage changes.

Finally, in order to select a value for the level of consumption during unemployment (b) we include the cross-sectional mean of residual wages ($\mu(w)$) in our list of targets. The reason is that b determines the value of unemployment, that is, of a worker's general outside option, and therefore is closely connected to the overall scale of wages in the model economy. We accordingly set b to the

[11] The fifteen match productivity levels $\{z_1, z_2, \ldots z_{15}\}$ in the calibration are distributed over the interval $[0.5, 1.5]$ with equal spacing. The probability mass that is cut off by setting the top of the interval at 1.5 amounts to 0.19 percentage points. In the calibrated economy, all but the two lowest productivity levels generate positive profits for firms when hiring an unemployed worker.

value of 0.9.

Table 4.1 summarizes the parameter values of our calibration. The simulated model statistics together with the empirical targets are reported in Table 4.2. From the latter one can see that the calibrated economy reproduces the empirical rates of labor market transitions very accurately. Regarding statistics on between-job wage changes, the model fit is good with respect to the mean and the fraction of negative wage changes. However, the standard deviation of wage changes upon job-to-job transitions in the model is too large. This is mainly due to the large decline in wages that workers experience in the current model when they are immediately re-employed after a layoff. When looking only at positive between-job wage changes, the model produces a standard deviation of 0.27 that is very close to the corresponding figure of 0.26 in the data. Regarding statistics on within-job wage changes, the model closely matches the fraction of negative wage changes. Within the framework of our analysis, this means that it captures well the relative probabilities of workers' high and low output realizations. However, the model exhibits too high a mean and too low a standard deviation for within-job wage changes. The reason is that wage gains within a job due to outside offers are slightly too large relative to the scale of wage variation due to incentive provision. Finally, the model does well in matching the scale of the wage distribution as captured by the mean wage.

Table 4.1: Values of model parameters

Parameter	Description	Value
β	Discount factor	0.99
σ	CRRA parameter in $u(c)$	2
ψ	Workers' prob. of survival	0.994
δ	Prob. of match destruction	0.028
λ_u	Prob. of job offer, unemployed	0.62
λ_e	Prob. of job offer, employed	0.095
λ_r	Prob. of job offer, laid-off	0.4
γ	Power parameter in $g(\epsilon)$	2
ρ	Coefficient in $\pi(\epsilon)$	3
ΔA	Difference, worker productivity	1
ζ_0	Shift parameter in $F(\cdot)$	0.5
ζ_1	Shape parameter in $F(\cdot)$	2
ζ_2	Scale parameter in $F(\cdot)$	0.4
b	Consumption, unemployed	0.9

Table 4.2: Simulated and empirical statistics

Statistic		Model	Data
Labor market transitions:			
E-U flows	τ^{eu}	0.024	0.024
U-E flows	τ^{ue}	0.566	0.548
E-E flows	τ^{ee}	0.040	0.039
Log wage changes between jobs:			
Mean	$\mu^{bet}(\Delta \ln w)$	0.038	0.029
Std.	$\sigma^{bet}(\Delta \ln w)$	0.529	0.364
Frac. neg.	$\varpi_{-}^{bet}(\Delta \ln w)$	0.356	0.384
Log wage changes within a job:			
Mean	$\mu^{win}(\Delta \ln w)$	0.017	0.007
Std.	$\sigma^{win}(\Delta \ln w)$	0.105	0.142
Frac. neg.	$\varpi_{-}^{win}(\Delta \ln w)$	0.347	0.335
Cross-sectional wages:			
Mean	$\mu(w)$	1.167	1.146
Log Std.	$\sigma(\ln w)$	0.325	0.505

In addition to the targeted statistics, the last row in Table 4.2 reports the standard deviation of log wages in the model and the data. The figures show that the model economy exhibits a substantial amount of wage dispersion, reproducing around two-thirds of the empirically observed residual wage inequality.

5 Quantitative Analysis

The goal of the analysis presented in this section is to provide a quantitative assessment of the overall impact of moral hazard on wage inequality, as well as to identify and characterize the various channels through which the presence of moral hazard affects the wage distribution. The first part presents quantitative results on the overall impact of moral hazard on wage inequality, based on a comparison of the stationary equilibrium of the full model outlined in Section 2 with a scenario where worker effort is observable. In the second part, we explore differences in solutions between the moral hazard (MH) and the observable effort (OE) scenarios and outline how they translate into differences in the cross-sectional wage distributions through diverse channels. The third part contains a collection of counterfactual experiments that decompose the overall impact of moral hazard into its various channels. We first provide a measure of the direct effect of incentive provision and compare it to the overall impact. We then present a sequential decomposition of the overall difference in wage inequality between the OE and the MH scenarios, which show the direction of impact and allow for an assessment of the relative quantitative importance of the indirect effects at work. In the last part of the section, we summarize and discuss our results.

5.1 The overall impact of moral hazard on wage inequality

We first provide a quantitative answer to one of the central questions of this paper: What is the overall impact of moral hazard on wage inequality when employed workers engage in job search and firms compete for workers? Our assessment is based on the comparison between two scenarios which differ only with respect to the assumptions made about observability of worker effort. The benchmark scenario is the full model, described in Section 2 and parametrized according to the calibration procedure outlined in Section 4, where worker effort is not observable, and incentives need to be provided through variation in future utility. The contractual problem in this case takes the form of the optimization problem (16) subject to constraints (17) to (22). We label this the moral hazard (MH) scenario.

For the comparison scenario, we assume that worker effort is observable and therefore becomes an explicit part of the contractual arrangement between firms and workers. In this observable effort (OE) scenario, since incentive-compatibility is not required, an optimal contract is a solution to the optimization problem (16) subject to constraints (17) and (19) to (22). As stated in Lemma 2 (Section 2.3), in this environment a firm fully insures its worker against effort-dependent productivity shocks, that is, $U'(U,z) \equiv U^-(U,z) = U^+(U,z)$. When solving the model under observable effort, we keep parameters fixed at the calibrated values of the moral hazard scenario.

Table 5.1 presents a comparison of the two stationary equilibria corresponding to the above scenarios in terms of various measures of wage inequality. The last two rows of the table report the median and mean of the wage distributions for reference. The difference in statistics – given in the last column of the table – is stated as the percentage change associated with a transition from the OE to the MH scenario. All inequality measures shown in the table indicate that wage inequality is larger when a moral hazard problem is present in labor contracts. The difference in inequality between the two scenarios lies between 6% and 11% for comprehensive measures such as the standard deviation of log wages, the coefficient of variation, and the Gini coefficient.

In addition, a closer look at changes in wage percentiles reveals that the impact of the moral hazard problem is not even across different parts of the wage distribution. The percentile ratios presented in Table 5.1 show that inequality between the top and bottom five percent of the wage distribution is around 28% larger in the MH scenario. Moreover, the greater part of this difference comes from larger inequality within the lower half of the distribution, as measured by the 50th-to-5th percentile ratio. The difference in mean-to-min ratios between scenarios by around 55% points to a similar conclusion, namely, that the presence of moral hazard has a particularly strong impact on the lowest parts of the wage distribution.

As shown in the analysis of the simplified model, moral hazard shapes the wage distribution through a variety of effects. In the next two sections, we attempt to disentangle and characterize the

Table 5.1: Measures of wage inequality, OE vs. MH

	OE	MH	Δ%
Std(ln w)	0.293	0.325	10.8 %
CV(w)	0.285	0.302	6.0 %
Gini(w)	0.163	0.173	6.5 %
95/05(w)	2.330	2.988	28.3 %
95/50(w)	1.485	1.498	0.9 %
50/05(w)	1.569	1.995	27.1 %
Mean/min(w)	1.585	2.459	55.1 %
p50(w)	1.177	1.158	-1.6 %
Mean(w)	1.187	1.167	-1.7 %

different channels of impact in the context of the full model, which will shed light on the mechanisms underlying the present results.

5.2 Differences in optimal contracts, firm profits, and wage dynamics

In the following paragraphs, we discuss differences between the OE and the MH scenarios in the solutions to a firm's optimal contract design problem, as well as the resulting differences in firm profits. In the course of outlining their relevance for differences in wage dynamics and in the cross-sectional wage distributions, we describe the direct as well as the various indirect effects of the presence of moral hazard.

As a point of departure, it is useful to compare the components that determine the levels and the dynamics of workers' wages between the two scenarios. These components can best be seen from the expressions of a worker's value of an employment contract in a match of productivity z. Under moral hazard, this value satisfies

$$
\begin{aligned}
U &= u(w_{MH}) - g(\epsilon_{MH}) + \beta\psi\delta U^n + \\
&+ \beta\psi(1-\delta)\Big\{(1-\lambda_e)\Big[U^+_{MH}\pi(\epsilon_{MH}) + U^-_{MH}\big(1-\pi(\epsilon_{MH})\big)\Big] + \\
&+ \lambda_e \sum_{\tilde{z}\in\mathcal{Z}} \Big[U_o(U^+_{MH}, z, \tilde{z})\pi(\epsilon_{MH}) + U_o(U^-_{MH}, z, \tilde{z})(1-\pi\big(\epsilon_{MH})\big)\Big] f(\tilde{z})\Big\} \quad , \qquad (36)
\end{aligned}
$$

where w_{MH}, ϵ_{MH}, U^+_{MH} and U^-_{MH} are the firm's optimal policies as functions of (U, z). The worker's current wage level is given by $w_{MH}(U, z)$. In the case of no outside offer, next period's wage level is determined by the spread in continuation values $\big[U^+_{MH}(U, z) - U^-_{MH}(U, z)\big]$ necessary to induce effort $\epsilon_{MH}(U, z)$, and by the relative probability of high output as a function of effort: With probability

$\pi(\epsilon_{MH})$ the worker will obtain a wage raise, otherwise his wage will decrease. In the case of a relevant outside offer, next period's wage will, in addition, be determined by the two firms' break-even utility levels, $U^*_{MH}(z)$ and $U^*_{MH}(\tilde{z})$, which depend on the firms' profit functions and enter the term $U_o(U^i_{MH}, z, \tilde{z})$, $i \in \{+, -\}$.

When effort is observable, the corresponding expression becomes much simpler, since continuation utilities within a contract are independent of the worker's output realization. Moreover, lifetime utility within a contract also remains constant over time, that is, $U'_{OE}(U, z) = U$. Under observable effort, a worker's value of a contract in match z thus satisfies

$$U = u(w_{OE}) - g(\epsilon_{OE}) + \beta\psi\delta U^n + \beta\psi(1-\delta)\left\{(1-\lambda_e)U + \lambda_e \sum_{\tilde{z}\in\mathcal{Z}} U_o(U, z, \tilde{z}) f(\tilde{z})\right\} \quad . \tag{37}$$

The current wage level is given by $w_{OE}(U, z)$. Over time, a worker's wage only changes if he receives a relevant outside offer, in which case the wage level will be determined by the two competing firms' break-even utility values, $U^*_{OE}(z)$ and $U^*_{OE}(\tilde{z})$, which enter the term $U_o(U, z, \tilde{z})$. Effort $\epsilon_{OE}(U, z)$ affects wage levels and wage dynamics only indirectly by influencing, on the one hand, the effort compensation part of wages, and, on the other hand, firms' profit functions and break-even utility levels.

Figure 5.1 shows policy functions of firms, for a match of productivity level z_7, for continuation utilities U^i. As mentioned above, in the OE scenario the lifetime utility of a worker within his current labor contract is non-stochastic and constant over time. By contrast, in the MH scenario, incentive-compatibility requires that workers are – within their current labor contract – rewarded for high output realizations with an increase in utility and punished for low output realizations with a utility decrease ($U^+_{MH} > U^-_{MH}$). Accordingly, among workers who start working at the same type of firm, lifetime utilities drift apart over time, independently of whether or not they receive outside job offers, merely as a result of differences in their histories of stochastic-output realizations. By introducing stochastic dynamics in lifetime utility into labor contracts, the presence of moral hazard thus adds a specific source of wage dispersion within groups of workers with the same histories of job offers to the environment.

This direct effect of incentive provision is complemented by a number of additional, indirect effects. All of these indirect effects are related to the cost of extracting the required level of effort from workers. A comparison between the MH and the OE scenarios shows that the need to provide incentives makes worker effort more expensive for firms in the former scenario than in the latter. In both scenarios workers have to be compensated for the disutility from spending effort. Under moral hazard, however, risk-averse workers need, in addition, to be compensated for the variation in income arising from incentive provision: Higher average wages are needed to provide the same level of lifetime utility. These differences in effort costs translate into differences between the wage distributions of the two scenarios through various channels.

One of the effects of increased effort costs works through changes in critical utility levels. Relatively higher effort costs under moral hazard lead to a relatively lower value to the firm of any worker-firm match. Thus, for all match types the value function of the MH scenario is below that of

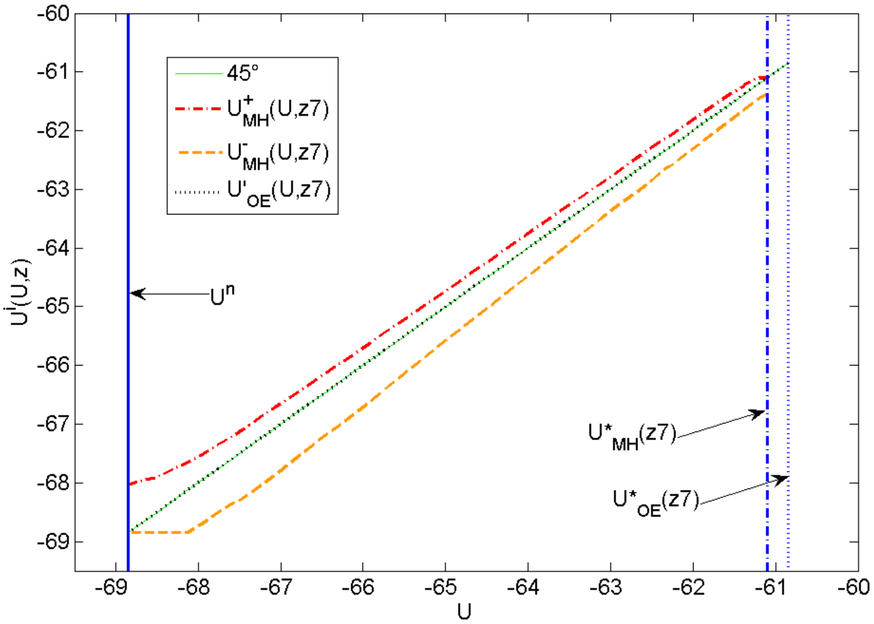

Figure 5.1: Policy functions for continuation utilities, MH vs. OE (match z_7)

the OE scenario, which implies that critical utility levels $U^*(z)$ are lower too. Figure 5.2 illustrates the differences between the value functions and critical utilities of firms with match type z_7 corresponding to the two scenarios. Table 5.2 reports values and differences in maximum profit levels of firms, $V(U^n, z)$, as well as in critical utility levels, $U^*(z)$, for each level of match productivity that makes positive profits. The figures show that maximum profits as well as critical utility levels are indeed lower in the MH scenario for all match types active in the present calibration.

Regarding the impact on the wage distribution, the differences in $U^*(z)$-levels are relevant for differences in the levels of wages associated with outside job offers. On the one hand, a decrease in critical utility levels leads to lower utility gains that workers can achieve through outside job offers, either with their current or with a new employer. It therefore translates into lower wage gains workers can attain through on-the-job search. On the other hand, as demonstrated in the simplified model above, workers need to be compensated for lower continuation values associated with outside offers by higher expected wages when they leave unemployment. To give an impression of the distribution of critical utility levels in the calibrated model, Figure 5.3 shows $U^*(z)$-levels of all match types in the MH scenario against the background of policies for continuation values of a firm with the highest match productivity (z_{15}). Relative to the OE scenario, all the critical utility levels are shifted to the left, so outside job offers lead to lower utility gains for the worker.

Another effect of increased effort costs is associated with the levels of effort that firms choose to extract from their workers. In contrast to the simplified model, worker effort in the present framework is a continuous choice variable. As argued above, due to workers' risk aversion, incentive provision increases the costs of effort at all levels of lifetime utility promised to the worker. Firms react to the increased costs by adjusting downwards the effort levels prescribed to their workers. Figure 5.4

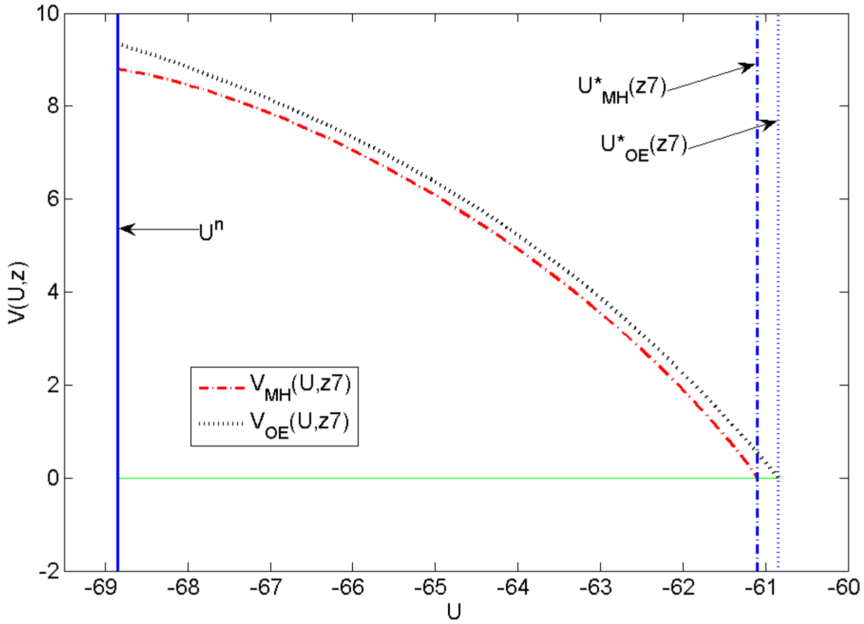

Figure 5.2: Value functions and critical utility levels, MH vs. OE case (match z_7)

Table 5.2: Maximum firm profits and critical utility levels, OE vs. MH

	$V(U^n, z)$:			$U^*(z)$:		
	OE	MH	Δ%	OE	MH	Δ%
$z_3 = 0.67$	1.81	1.24	-31.7 %	-66.98	-67.50	-0.8 %
$z_4 = 0.73$	3.72	3.22	-13.4 %	-65.05	-65.43	-0.6 %
$z_5 = 0.80$	5.54	5.03	-9.3 %	-63.43	-63.76	-0.5 %
$z_6 = 0.87$	7.39	6.86	-7.2 %	-62.04	-62.33	-0.5 %
$z_7 = 0.93$	9.35	8.80	-5.9 %	-60.84	-61.10	-0.4 %
$z_8 = 1.00$	11.44	10.86	-5.0 %	-59.80	-60.02	-0.4 %
$z_9 = 1.07$	13.66	13.06	-4.4 %	-58.88	-59.08	-0.3 %
$z_{10} = 1.13$	16.00	15.36	-4.0 %	-58.06	-58.24	-0.3 %
$z_{11} = 1.20$	18.43	17.75	-3.7 %	-57.33	-57.49	-0.3 %
$z_{12} = 1.27$	20.93	20.21	-3.4 %	-56.67	-56.82	-0.3 %
$z_{13} = 1.33$	23.48	22.70	-3.3 %	-56.07	-56.22	-0.3 %
$z_{14} = 1.40$	26.06	25.23	-3.2 %	-55.53	-55.67	-0.2 %
$z_{15} = 1.47$	28.66	27.77	-3.1 %	-55.04	-55.17	-0.2 %

shows – again for a match of type z_7 – the policy functions for effort in the MH and the OE case. As expected, effort levels are generally lower under moral hazard. However, the figure exhibits a remarkable difference in shape between the two policy functions: While in the OE scenario effort levels are monotonically decreasing in lifetime utility throughout, in the MH case effort is increasing over a short interval of low utility values and decreasing over the rest of the domain. Moreover,

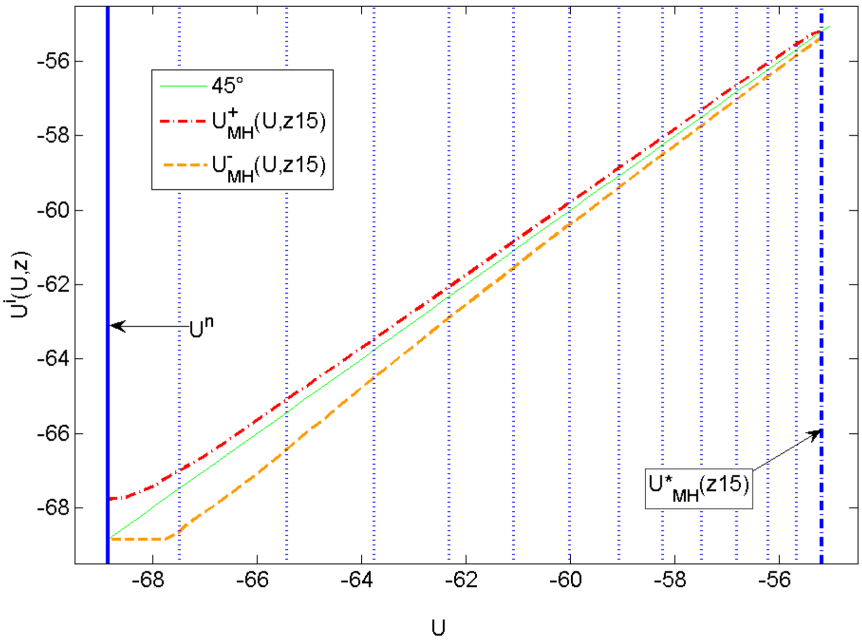

Figure 5.3: Policy functions for continuation utilities, MH (match z_{15}), and critical utility levels of all match types (dotted vertical lines)

effort in the MH scenario decreases at a higher rate towards the upper bound on lifetime utility in the current match.

The initial increase in the effort function under moral hazard arises from specific costs of incentive provision at low levels of lifetime utility. For values close to U^n firms have only little scope to punish workers for low output realizations through utility cuts because in this situation a worker's participation constraint becomes binding easily. Consequently, incentives to extract effort have primarily to take the form of rewarding high output realizations by utility increase. The lower bound on continuation utility gives rise to additional costs of effort, and firms decrease effort levels by a larger amount relative to the OE scenario. As lifetime utility increases, the cost associated with the worker's participation constraint declines so that initially effort levels increase in utility. The range of lifetime utility over which the participation constraint is binding when worker output is low can be clearly seen in Figure 5.4. This region, where the function U^-_{MH} is flat, coincides with the interval over which the effort function is increasing.

Similarly, incentive provision at high values of lifetime utility is restrained by the firm's participation constraint. In the uppermost range of utility levels, incentives need to be provided primarily through punishments for low output via a utility decrease. The upper bound on continuation utility again gives rise to additional costs of effort, leading to a stronger decrease in effort levels relative to the OE scenario.

Effort levels affect the wage distribution in two ways. First, leaving everything unchanged, lower effort levels under moral hazard translate into lower wages at a given level of lifetime utility, since workers have to be compensated less for disutility from effort. Second, as pointed out in the discussion of equation (36), under moral hazard effort levels have a direct impact on the dynamics of lifetime utility of workers within a job. Lower effort levels lead to a lower probability of utility increase in the next period, and thus to a lower probability of a wage raise.

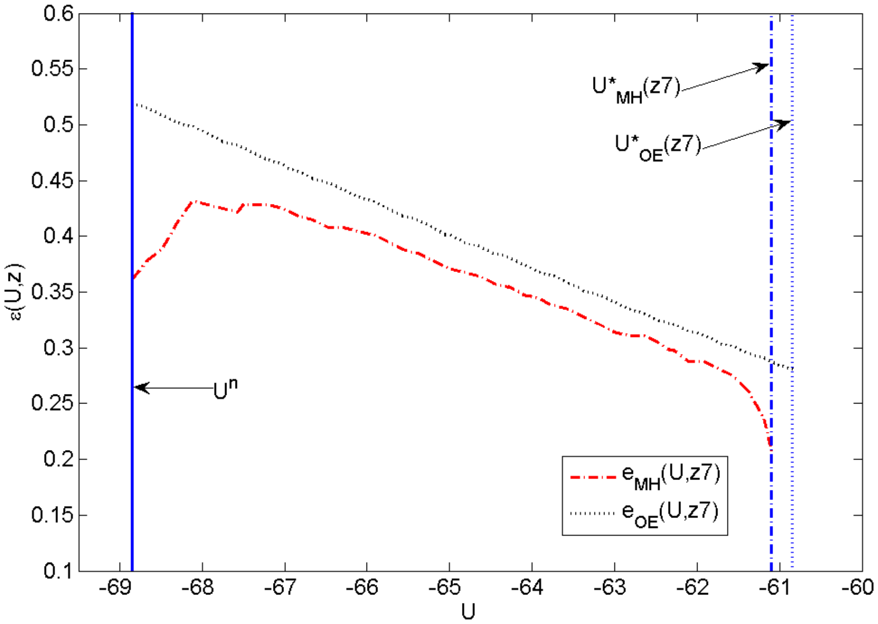

Figure 5.4: Policy functions for effort, MH vs. OE (match z_7)

Finally, differences between the two scenarios in firms' policy functions for wages are related to all of the indirect effects of incentive provision. In the following, we describe the various partial channels of impact on wage levels associated with a change from the OE to the MH scenario. For the description of each channel, we maintain the hypothesis of leaving everything else unchanged.

Most immediately, at a given level of lifetime utility, wages increase in order to compensate workers for the additional variation in earnings due to incentive provision. At low values of lifetime utility, wages also increase in order to compensate workers for lower continuation values associated with outside offers due to lower firm profits. In contrast, wages decrease at all levels of utility as a consequence of lower effort compensation costs due to lower effort levels prescribed to workers. Furthermore, wages also adjust to changes in the dynamics of a worker's lifetime utility within a job. As outlined above, at very low levels of lifetime utility, where the worker's participation constraint restricts incentive provision from below, expected continuation utility is high. In this range of utility, contemporaneous wages decrease, since workers expect a wage raise in the future. Analogously, at levels of lifetime utility close to the firm's participation constraint, expected continuation utility is low. In consequence, contemporaneous wages increase in this range of utility to compensate workers for expected wage loss in the future.

Figure 5.5 presents, again for the case of z_7-type matches, the policy functions for wages in the two scenarios. It clearly illustrates how the strong decrease in effort levels at low values of lifetime utility, together with the expected future utility increase, translate into significantly lower starting wages under moral hazard than under observable effort. The figure also clearly shows how the expected decline in continuation utility leads to an increase in wages at high levels of lifetime utility.

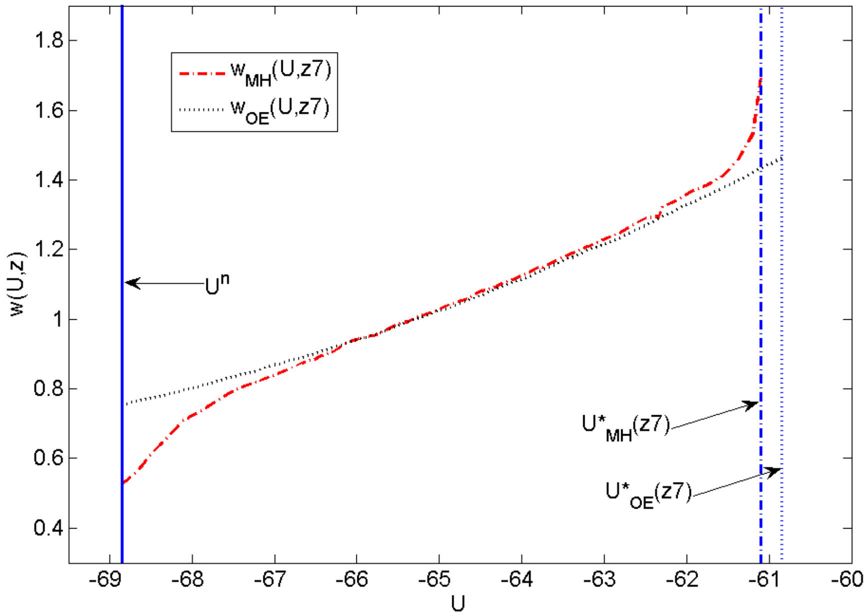

Figure 5.5: Policy functions for wage, MH vs. OE (match z_7)

5.3 Decomposition of effects on the wage distribution

In this part, we carry out a series of counterfactual experiments that aim to decompose the overall impact of moral hazard on the wage distribution into its different channels. We first provide a measure of the direct effect of incentive provision. Comparing it with the results on the overall impact of moral hazard, we document that the direct effect is in fact quantitatively larger. We then move to a sequential decomposition of the overall difference in wage inequality by adding to the observable effort scenario, one by one, the components of the optimal contract and firms' break-even levels of lifetime utility under moral hazard. In the course of this decomposition, we can assess the direction of impact as well as the quantitative importance of the various indirect effects at work.

5.3.1 Measuring the direct effect of incentive provision

We first present a measure of the direct effect of moral hazard that arises from the need to provide incentives to a worker by making wage payments depend on his output realizations. For this end, we construct a counterfactual scenario that is closely related to the analysis of the simplified model: Starting from the scenario under observable effort, we introduce, for each group of possible job offer histories, a simple mean-preserving spread of wages that reproduces the within-group wage dispersion of the moral hazard scenario.

More specifically, denote a particular offer-history group by h. For each of these groups, we keep effort constant at the value of the OE scenario, $\epsilon_{OE}(h)$, and calculate two counterfactual wage levels

$w^{+}_{CF}(h) > w^{-}_{CF}(h)$ such that

$$\pi(\epsilon_{OE}(h))w^{+}_{CF}(h) + \left[1 - \pi(\epsilon_{OE}(h))\right] w^{-}_{CF}(h) \quad = \quad w_{OE}(h) \tag{38}$$

and

$$\begin{aligned} \pi(\epsilon_{OE}(h)) \left[w^{+}_{CF}(h) - w_{OE}(h)\right]^2 & + \\ + \left[1 - \pi(\epsilon_{OE}(h))\right] \left[w^{-}_{CF}(h) - w_{OE}(h)\right]^2 & = \operatorname{Var}\left[w_{MH}|h\right] \quad . \end{aligned} \tag{39}$$

Table 5.3 presents changes in inequality measures in the counterfactual (CF1) relative to the OE scenario. What is striking is that the increase in all comprehensive measures is larger than the overall effects reported in Table 5.1. In particular, the standard deviation of log wages increases by nearly 16%, as opposed to an increase in around 11% for the overall impact. The increase in the ratio of the top and the bottom fifth-percentile wages is slightly smaller than the one found for the overall effect. Moreover, inequality in the upper half of the distribution now decreases slightly, while in the lower half it increases more strongly than before. Finally, the increase in the mean-to-min ratio of 131% is more than double the size of the figure reported in Table 5.1.

Table 5.3: Changes in measures of wage inequality, direct effect

	OE	CF1	Δ%
Std($\ln w$)	0.293	0.340	15.8 %
CV(w)	0.285	0.306	7.1 %
Gini(w)	0.163	0.174	6.9 %
95/05(w)	2.330	2.959	27.0 %
95/50(w)	1.485	1.472	-0.9 %
50/05(w)	1.569	2.010	28.1 %
Mean/min(w)	1.585	3.664	131.2 %

These results indicate that the direct effect of incentive provision is quantitatively substantial, and is in fact larger than the overall impact of moral hazard. Hence, at least some of the indirect effects must work as counteracting forces that compress the wage distribution.

5.3.2 Assessing the indirect effects

For the following sequence of counterfactual experiments, we start from the scenario with observable effort and add, one by one, the components of the model solution under moral hazard. The steps of the decomposition show in which direction the various indirect effects of moral hazard affect wage inequality. Moreover, the series of experiments demonstrates the relative quantitative importance of the effects as well as their particular impact on different parts of the wage distributions.

Step 1: Incentive-compatibility

In the first step, we introduce the incentive-compatibility requirement on workers' continuation values. More precisely, we assume that firms prescribe the same levels of effort to workers as in the observable effort scenario, but need to provide incentives through variation in continuation utility. Moreover, firms are not allowed to adjust contemporaneous wage levels, and we assume that critical utility levels are the same as in the observable effort case. Technically, we keep the policy functions for wage and effort as well as the critical utility levels fixed at the OE solution, so that $w_{CF} = w_{OE}(U, z)$, $\epsilon_{CF} = \epsilon_{OE}(U, z)$, and $U^*_{CF} = U^*_{OE}(z)$. We solve (ICC) and (PKC) for the counterfactual policy functions $U^+_{CF}(U, z)$ and $U^-_{CF}(U, z)$, subject to the worker's and the firms participation constraints, (WPC) and (FPC).

Table 5.4 reports changes in statistics on the wage distribution that are associated with a transition from the OE scenario to the present counterfactual (CF2). Given the setup of the experiment, these changes obviously partially contain the direct effect of providing incentives through wage variation within a job. Indeed, the changes in wage percentiles illustrate that both at the bottom (between the first and the 25th percentiles) and in the upper part (between the 50th and the 75th percentiles) of the distribution the dispersion of wages increases.

However, changes in the lowest percentiles also reveal a strong upward compression of wages from the lower end towards the middle of the distribution. This upward compression arises from two indirect forces acting at low levels of lifetime utility. First, in analogy to the simplified model, risk-averse workers need to be compensated for wage variation by higher expected future wages. Second, incentive provision through variation in future utility is restricted by the worker's participation constraint. Since wage punishments for low output are constrained from below, firms need to reward high output with very large wage raises in order to implement the given levels of effort, thereby increasing a worker's expected future wage. Both forces thus cause wages of workers at low levels of lifetime utility to increase more quickly than in the OE scenario.

A similar restriction on incentive provision is at work at the very top of the distribution: Rewards to the worker through future utility are restricted from above by the firm's participation constraint, so that incentives need to be provided primarily through punishments for low output. This indirect effect leads to a mild downward compression of wages, as indicated by the decrease in the 95th percentile.

The indirect effects, arising from the incentive-compatibility requirement together with the workers' and the firms' participation constraint, lead to a strong compression of the wage distribution. They vastly dominate the expansion of within-group wage differences, due to wage variation within a job, in the present experiment. As a result, overall wage inequality in terms of the standard deviation of log wages decreases by over 16%.

Step 2: Critical utility levels

The setting of the next experiment (CF3) is similar to the previous one, except that we change the critical utility levels to the values of the moral hazard scenario. This means that we add the reduction in lifetime utility levels associated with outside job offers. Technically, we set $w_{CF} = w_{OE}(U, z)$,

Table 5.4: Decomposition step 1, incentive-compatibility

	OE	CF2	Δ%
Std($\ln w$)	0.293	0.245	-16.4 %
p01(w)	0.749	0.752	0.3 %
p05(w)	0.750	0.821	9.4 %
p25(w)	0.826	0.951	15.2 %
p50(w)	1.177	1.189	1.0 %
p75(w)	1.439	1.463	1.7 %
p95(w)	1.748	1.736	-0.7 %

$\epsilon_{CF} = \epsilon_{OE}(U, z)$, $U^*_{CF} = U^*_{MH}(z)$, and obtain the counterfactual policy functions $U^+_{CF}(U, z)$ and $U^-_{CF}(U, z)$ by solving (PKC) and (ICC) subject to (WPC) and (FPC).

Table 5.5 reports changes in wage statistics associated with a transition from the previous counterfactual (CF2) to the present scenario (CF3). The changes in percentiles show that the reduction in critical utility levels affects both the bottom and the upper half of the wage distribution. First, similar to the mechanism demonstrated in the simplified model, workers at low levels of lifetime utility need to be compensated for the relatively worse prospects associated with outside job offers by higher expected continuation values. This again causes low wage workers to move up faster and leads to a mild upward compression of wages from the bottom of the distribution. Second, workers who have received outside offers attain lower levels of lifetime utility and thus lower wages. Consequently, wages are compressed downwards in the upper half of the distribution.

Both indirect effects at work in the present experiment thus counteract the direct effect of moral hazard. The overall reduction in wage inequality associated with the decrease in critical utility levels is, however, modest: The standard deviation of log wages decreases by around 2%.

Step 3: Effort levels

In the third step, we add the change in effort levels between the OE and the MH scenarios. Since we still keep firms' policy functions for wages fixed at the observable effort solution, the present counterfactual scenario (CF4) captures mainly the effect of lower effort levels on the wage distribution that operates through lower probabilities of wage increases within a job. Technically, we impose $w_{CF} = w_{OE}(U, z)$, $\epsilon_{CF} = \epsilon_{MH}(U, z)$, $U^*_{CF} = U^*_{MH}(z)$, and obtain the counterfactual policy functions $U^+_{CF}(U, z)$ and $U^-_{CF}(U, z)$ by solving (PKC) and (ICC) subject to (WPC) and (FPC).

Table 5.6 reports changes in wage statistics relative to the previous counterfactual scenario. As can be seen from the changes in wage percentiles, the indirect effects working through changes in wage dynamics caused by lower effort levels again affect both the top and the bottom parts of the wage distribution. As was illustrated in Figure 5.4, the reduction in effort levels in response to moral hazard is particularly large in two regions of lifetime utility: On the one hand, at high values

Table 5.5: Decomposition step 2, critical utility levels

	CF2	CF3	Δ%
Std(ln w)	0.245	0.240	-2.2 %
p01(w)	0.752	0.752	0.0 %
p05(w)	0.821	0.822	0.1 %
p25(w)	0.951	0.952	0.1 %
p50(w)	1.189	1.171	-1.4 %
p75(w)	1.463	1.439	-1.6 %
p95(w)	1.736	1.716	-1.1 %

close to the firm's participation constraint, and, on the other hand, at low values near the worker's participation constraint. In consequence, when a worker has reached a high level of lifetime utility, he faces a high probability of a wage loss in the next period. This leads to a slight downward compression of wages from the top of the distribution, as illustrated by the decrease in the 95th percentile wage. Analogously, workers at low levels of lifetime utility face a much lower probability of a wage raise. For instance, workers who have recently left unemployment and have not yet received an outside job offer, on average, stay much longer at low wage levels than in the previous scenario.

The changes in wage dynamics at high values of lifetime utility further compress the wage distribution. In contrast, the changes in dynamics at low utility levels partly reverse the upward compression of wages observed in the first two experiments. Moreover, the strength of impact on the lower parts of the wage distribution is clearly dominant in the present experiment. Overall, the changes in workers' dynamics of lifetime utility associated with lower effort levels lead to an increase in wage inequality by over 8% in terms of the standard deviation of log wages.

Table 5.6: Decomposition step 3, effort levels

	CF3	CF4	Δ%
Std(ln w)	0.240	0.261	8.7 %
p01(w)	0.752	0.750	-0.3 %
p05(w)	0.822	0.780	-5.2 %
p25(w)	0.952	0.893	-6.2 %
p50(w)	1.171	1.149	-1.9 %
p75(w)	1.439	1.428	-0.8 %
p95(w)	1.716	1.704	-0.7 %

Step 4: Wage levels

The last step completes the transition from the OE to the MH scenario by adjusting firms' wage policies to the solution under moral hazard. Table 5.7 reports changes in wage statistics associated with this step. The changes in wage percentiles show that the adjustment of wage levels leads to an expansion of the distribution both at the bottom (between the first and the 25th percentile) and in the upper part (between the 50th and the 95th percentiles).

As discussed in Section 5.2, changes in wage policies between the OE and the MH scenarios are associated with a variety of indirect effects of moral hazard. In particular, Figure 5.5 illustrated that, similar to the changes in effort levels, the predominant changes in wage policies occur at levels of lifetime utility close to either the worker's or the firm's participation constraint: In the former case, wages are adjusted downwards due to lower effort compensation needs and in reaction to higher expected future wage gains. In the latter case, contemporaneous wages are increased to compensate workers for expected future wage losses.

According to the figures in Table 5.7, increased wage levels at high lifetime utility lead to a modest expansion at the top of the distribution. In contrast, the downward adjustment of wage levels at low utility values leads to a large expansion at the bottom, reinforcing the inequality-increasing effect of reduced effort levels in the previous scenario. In total, changes in wage levels increase the standard deviation of log wages by nearly 25%.

Table 5.7: Decomposition step 4, wage levels

	CF4	MH	Δ%
Std($\ln w$)	0.261	0.325	24.7 %
p01(w)	0.750	0.500	-33.4 %
p05(w)	0.780	0.581	-25.5 %
p25(w)	0.893	0.880	-1.5 %
p50(w)	1.149	1.158	0.8 %
p75(w)	1.428	1.442	1.0 %
p95(w)	1.704	1.735	1.8 %

Table 5.8 gives an overview of the sequential changes in measures of wage inequality in the four decomposition steps. The first three rows confirm that the direction and size of inequality change as measured by the coefficient of variation and the Gini coefficient are similar to those found for the standard deviation of log wages. In fact, the percentage changes in the first three steps are remarkably close across the three inequality measures. Only in the fourth step, associated with the adjustment of wage levels, the changes in coefficient of variation and the Gini coefficient are significantly lower than the one in the standard deviation of log wages.

The changes in percentile ratios reported in the last three rows of Table 5.8 point to two additional

Table 5.8: Overview of changes in wage inequality in the sequential decomposition

	Incentive-compatibility	Critical utility levels	Effort levels	Wage levels
Std(ln w)	-16.4 %	-2.2 %	8.7 %	24.7 %
CV(w)	-14.0 %	-1.8 %	8.3 %	15.9 %
Gini(w)	-14.1 %	-2.0 %	8.5 %	16.5 %
95/05(w)	-9.3 %	-1.3 %	4.7 %	36.7 %
95/50(w)	-1.6 %	0.3 %	1.3 %	1.0 %
50/05(w)	-7.7 %	-1.6 %	3.4 %	35.4 %

observations: First, inequality between the top and the bottom 5th-percentile wages changes in line with the comprehensive measures of inequality in each step. Second, the larger part of inequality change in each decomposition step takes place between the median and the lowest 5th-percentile wages. This indicates that the indirect effects of moral hazard have a particularly strong impact on the lower part of the wage distribution. For instance, the upward compression of wages in the first step, due to compensation for risky wages and high expected continuation utilities as a consequence of the workers' participation constraint leads to a substantial reduction in the 50th-to-5th percentile ratio by nearly 8%. The overwhelming change in the lower part of the distribution arises from the adjustment of wage levels in the last decomposition step. The decrease in wages at low lifetime utility, in reaction to lower effort compensation costs and to high expected continuation utilities, leads to an increase in the 50th-to-5th percentile ratio by over 35%.

6 Concluding Remarks

In the present paper we study how moral hazard in employment relations impacts on the cross-sectional wage distribution and, in particular, on the extent of residual wage inequality in a labor market characterized by search frictions. Our analysis builds on a search model featuring job-to-job mobility and firm competition for workers. In our framework, firms offer long-term contracts to risk-averse workers in the presence of repeated moral hazard due to a worker's effort on the job not being observable to his employer. Commitment to contracts is limited on both the worker's and the firm's side.

The informational friction – internal to the worker-firm relation – of unobservable effort gives rise to performance-dependent pay: In order to provide incentives to workers, wage payments need to vary across different levels of effort-dependent match output. This direct effect of moral hazard leads to wage dispersion among workers with the same histories of job offers, but different histories of output realizations, and thus increases wage inequality. We show that, in the presence of on-the-job search and limited commitment, moral hazard also affects the wage distribution through several indirect channels. All of these indirect effects are related to the fact that incentive provision through wage variation increases the costs of worker effort to firms. Some of the indirect effects due to increased effort costs counteract the direct effect on wage inequality, while others reinforce it.

For a quantitative analysis, we calibrate the model to key moments of individual wage dynamics in the U.S. labor market obtained from micro data from the mid-2000s. We find that, overall, the presence of moral hazard in employment relations increases residual wage inequality by around ten percent. The strength of impact, however, is not even across different parts of the wage distribution: Inequality increases by much larger amounts within the lower half of the distribution. Decomposition exercises show that the direct and indirect effects at work behind the overall change are quantitatively substantial. Moreover, indirect effects due to increased effort costs have a particularly strong impact on the lower parts of the wage distribution. In particular, constraints on incentive provision at low levels of lifetime utility – due to worker's limited commitment – play an important role for the increase in effort costs to firms. The associated decrease in effort levels prescribed and wages paid to workers turns out to be the main force behind the large inequality increase in the lower half of the wage distribution.

Based on the present study, an immediate area for further research is a comparative analysis between different industries and occupations. Our framework generates predictions for the extent of wage inequality depending on the importance of moral hazard and the degree of job-to-job mobility and firm competition in a particular labor market. We are currently working on testing how well our framework can replicate observed differences in wage inequality between industries and occupations that differ along these dimensions.

Our analysis and current findings also point to a wider area for future research, namely, the study of different policy instruments in labor markets where incentive contracts are prominent. As we have shown, increased effort costs to firms arising from incentive provision play an important role in determining wage levels and wage dynamics as well as worker and match productivity. The costs of incentive provision are, for instance, substantially affected by the level and progressivity of workers' income taxes. Moreover, our finding that incentive provision is particularly costly at low levels of lifetime utility and wages suggests that policy instruments such as the level of unemployment benefits or minimum wage restrictions have comparatively larger effects on industries and occupations where incentive contracts are important. The evaluation of policy instruments may therefore change substantially when taking the need for incentive provision into account.

References

Altonji, J. G., Smith, A. A., and Vidangos, I. (2012). Modeling Earnings Dynamics. Yale University, mimeo.

Eckstein, Z. and Nagypal, E. (2004). The Evolution of U.S. Earnings Inequality: 1961-2002. *Federal Reserve Bank of Minneapolis Quarterly Review*, 28(2):10–29.

Fallick, B. and Fleischman, C. A. (2004). Employer-to-Employer Flows in the U.S. Labor Market: The Complete Picture of Gross Worker Flows. Finance and Economics Discussion Series 2004-34, Board of Governors of the Federal Reserve System (U.S.).

Flinn, C. and Mabli, J. (2009). On-the-Job Search, Minimum Wages, and Labor Market Outcomes in an Equilibrium Bargaining Framework. New York University, mimeo.

Hornstein, A., Krusell, P., and Violante, G. L. (2011). Frictional Wage Dispersion in Search Models: A Quantitative Assessment. *American Economic Review*, 101(7):2873–98.

Katz, L. F. and Autor, D. H. (1999). Changes in the Wage Structure and Earnings Inequality. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3A. North Holland, Amsterdam.

Lemieux, T., MacLeod, W. B., and Parent, D. (2009). Performance Pay and Wage Inequality. *Quarterly Journal of Economics*, 124(1):1–49.

Lentz, R. (2013). Optimal Wage-Tenure Contracts without Search Intensity Commitment. University of Wisconsin-Madison, mimeo.

Lise, J., Meghir, C., and Robin, J.-M. (2013). Mismatch, Sorting and Wage Dynamics. University College London, mimeo.

Low, H., Meghir, C., and Pistaferri, L. (2010). Wage Risk and Employment Risk over the Life Cycle. *American Economic Review*, 100(4):1432–67.

Manoli, D. and Sannikov, Y. (2005). Competitive Contracting and Employment Dynamics. University of California at Berkeley, mimeo.

Menzio, G., Telyukova, I. A., and Visschers, L. (2012). Directed Search over the Life Cycle. University of Pennsylvania, mimeo.

Nagypal, E. (2008). Worker Reallocation over the Business Cycle: The Importance of Employer-to-Employer Transitions. Northwestern University, mimeo.

Postel-Vinay, F. and Robin, J.-M. (2002). Equilibrium Wage Dispersion with Worker and Employer Heterogeneity. *Econometrica*, 70(6):2295–2350.

Rogerson, W. P. (1985a). Repeated Moral Hazard. *Econometrica*, 53(1):69–76.

Rogerson, W. P. (1985b). The First-Order Approach to Principal-Agent Problems. *Econometrica*, 53(6):1357–1367.

Spear, S. E. and Srivastava, S. (1987). On Repeated Moral Hazard with Discounting. *The Review of Economic Studies*, 54(4):599–617.

Thomas, J. and Worrall, T. (1988). Self-Enforcing Wage Contracts. *The Review of Economic Studies*, 55(4):541–553.

Tsuyuhara, K. (2013). Dynamic Contracts with Worker Mobility via Directed On-the-Job Search. University of Calgary, mimeo.

A Appendix

A.1 Proof of Lemma 2

Suppose there is a state (U, z) for which the optimal solution (w, ϵ, U^+, U^-) satisfies $U^+ \neq U^-$. Consider the case in which $U^+ > U^-$ (the opposite case follows the same arguments). One of the following mutually exclusive cases must apply:

1. $U^+ \notin \{U^*(z_n)\}_{n=1}^N$ and $U^- \notin \{U^*(z_n)\}_{n=1}^N$
2. $U^+ \in \{U^*(z_n)\}_{n=1}^N$ and $U^- \notin \{U^*(z_n)\}_{n=1}^N$
3. $U^+ \notin \{U^*(z_n)\}_{n=1}^N$ and $U^- \in \{U^*(z_n)\}_{n=1}^N$
4. $U^+ \in \{U^*(z_n)\}_{n=1}^N$ and $U^- \in \{U^*(z_n)\}_{n=1}^N$

Case 1. Consider the alternative candidate $(w, \epsilon, \tilde{U}^+, \tilde{U}^-)$ where w and ϵ are the same as in the optimal solution, but $\tilde{U}^+ = U^+ - \eta^+$ and $\tilde{U}^- = U^- + \eta^-$. The scalars (η^+, η^-) are positive and will be defined such that the alternative candidate gives the worker the same level of lifetime utility U as the optimal solution. Define z_i as the highest z_n satisfying $U^*(z_n) \leq U^-$, and z_j as the highest z_n satisfying $U^*(z_n) \leq U^+$. Then the following relationships hold:

- $z_i \leq z_j$
- $U^*(z_i) < U^- < U^*(z_{i+1})$ and $U^*(z_j) < U^+ < U^*(z_{j+1})$

For sufficiently low (η^+, η^-), it is the case that

$$U^*(z_i) < \tilde{U}^- < U^*(z_{i+1}) \quad \text{and} \quad U^*(z_j) < \tilde{U}^+ < U^*(z_{j+1}) \tag{40}$$

The requirement that the alternative candidate gives the same utility to the worker implies that

$$\eta^+ = \left(\frac{1 - \pi(\epsilon)}{\pi(\epsilon)} \frac{[1 - \lambda_e + \lambda_e F(z_i)]}{[1 - \lambda_e + \lambda_e F(z_j)]} \right) \eta^- \tag{41}$$

The values (η^+, η^-) satisfy conditions (40) and (41).

Let V denote the value of the contract to the firm under the optimal solution, and $\tilde{V}$ the value associated with the alternative candidate. Then

$$\tilde{V} - V = \beta\psi(1 - \delta) \Big\{ \big[aV(\tilde{U}^+) + bV(\tilde{U}^-) \big] - \big[aV(U^+) + bV(U^-) \big] \Big\} \tag{42}$$

where

$$a = \pi(\epsilon) \, [1 - \lambda_e + \lambda_e F(z_j)] > 0 \tag{43}$$
$$b = (1 - \pi(\epsilon)) \, [1 - \lambda_e + \lambda_e F(z_i)] > 0 \tag{44}$$

Let $\alpha = \frac{a}{a+b}$, then $\tilde{V} > V$ reduces to

$$\alpha V(\tilde{U}^+) + (1 - \alpha) V(\tilde{U}^-) > \alpha V(U^+) + (1 - \alpha) V(U^-) \tag{45}$$

By construction, due to requirement (41),

$$\alpha\tilde{U}^{+} + (1-\alpha)\tilde{U}^{-} = \alpha U^{+} + (1-\alpha)U^{-} \tag{46}$$

This can be seen from

$$\alpha\tilde{U}^{+} + (1-\alpha)\tilde{U}^{-} = \alpha U^{+} + (1-\alpha)U^{-} - \alpha\eta^{+} - (1-\alpha)\eta^{-} \tag{47}$$

and

$$-\alpha\eta^{+} - (1-\alpha)\eta^{-} = -\frac{a}{a+b}\eta^{+} + \frac{b}{a+b}\eta^{-} = \frac{a}{a+b}\underbrace{\left[-\eta^{+} + \frac{b}{a}\eta^{-}\right]}_{=0 \text{ by } (41)} = 0 \tag{48}$$

Therefore, if $V(U,z)$ is concave in U, since $U^{+} - U^{-} > \tilde{U}^{+} - \tilde{U}^{-}$ condition (45) holds. Thus, $\tilde{V} > V$ and $U^{+} \neq U^{-}$ cannot be optimal.

Case 2. Define $(z_i, z_j, \tilde{U}^{+}, \tilde{U}^{-})$ as in Case 1. The following relationships hold:

- $z_i \leq z_j$
- $U^*(z_i) < U^{-} < U^*(z_{i+1})$ and $U^*(z_j) = U^{+} < U^*(z_{j+1})$

For sufficiently low (η^{+}, η^{-}), it is the case that

$$U^*(z_i) < \tilde{U}^{-} < U^*(z_{i+1}) \quad \text{and} \quad U^*(z_{j-1}) < \tilde{U}^{+} < U^*(z_j) < U^*(z_{j+1}) \tag{49}$$

The requirement that the alternative candidate gives the same utility to the worker implies that

$$\eta^{+} = \left(\frac{1-\pi(\epsilon)}{\pi(\epsilon)} \frac{[1-\lambda_e+\lambda_e F(z_i)]}{[1-\lambda_e+\lambda_e F(z_{j-1})]}\right)\eta^{-} \tag{50}$$

The values (η^{+}, η^{-}) satisfy conditions (49) and (50).

Let V denote the value of the contract to the firm under the optimal solution, and $\tilde{V}$ the value associated with the alternative candidate. Recall that the incumbent keeps the worker if the outside competitor is not willing to offer a higher lifetime utility than he has promised to deliver. Then

$$\tilde{V} - V = \beta\psi(1-\delta)\Big\{\left[a'V(\tilde{U}^{+}) + bV(\tilde{U}^{-})\right] + \underbrace{\left[V(U^*(z_j)I(z > z_j) - V(U^{+})\right]}_{c \geq 0} - \left[a'V(U^{+}) + bV(U^{-})\right]\Big\} \tag{51}$$

where $a' = \pi(\epsilon)\left[1-\lambda_e+\lambda_e F(z_{j-1})\right] > 0$ and b is defined as in Case 1. Let $\alpha' = \frac{a'}{a'+b}$. Recall that in the present case $U^*(z_j) = U^{+}$, hence the term c is non-negative. Then $\tilde{V} > V$ reduces to

$$\alpha' V(\tilde{U}^{+}) + (1-\alpha')V(\tilde{U}^{-}) > \alpha' V(U^{+}) + (1-\alpha')V(U^{-}) \tag{52}$$

By construction

$$\alpha'\tilde{U}^{+} + (1-\alpha')\tilde{U}^{-} = \alpha' U^{+} + (1-\alpha')U^{-} \tag{53}$$

Hence, given that $U^{+} - U^{-} > \tilde{U}^{+} - \tilde{U}^{-}$, condition (52) holds and $U^{+} \neq U^{-}$ cannot be optimal.

Case 3. Here $(\eta^{+}, \eta^{-}, V, \tilde{V})$ are defined as in Case 1. Then $\tilde{V} > V$ and $U^{+} \neq U^{-}$ cannot be optimal.

Case 4. Here $(\eta^{+}, \eta^{-}, V, \tilde{V})$ are defined as in Case 2. Then $\tilde{V} > V$ and $U^{+} \neq U^{-}$ cannot be optimal.

□

A.2 Data

The Survey of Income and Program Participation (SIPP) is a longitudinal survey of representative households in the United States, administered by the U.S. Census Bureau. The survey focuses on collecting data at high frequencies on individuals' income sources and amounts, their labor market status as well as eligibility for and participation in government programs. The SIPP consists of a set of partially overlapping panels, each between two and four years in duration, starting from 1984. Public use data sets from the SIPP published by the U.S. Census Bureau are provided on the homepage of the NBER.[12]

Over the length of one panel, households are interviewed every four month. At each interview, a detailed monthly labor market history (employers, hours, earnings, job characteristics, employment turnover) for each member of the household over the preceding four months is collected, with some variables being recorded even at a weekly frequency. In particular, detailed information for up to two jobs the individual has held over those four months (referred to as the *wave*) are recorded.

The SIPP 2004 panel covers the period from October 2003 to December 2007. Data are released only in core wave files, and longitudinal sampling weights, constructed at the end of data collection, are provided in separate files. We restrict our attention to observations from January 2004 to December 2006 – January 2007 observations are included for transition rates and wage changes between months – for four reasons. First, this avoids the problem of the sample size for the first and the last three months of any SIPP panel being much smaller due to the rotating design of data collection. Second, the sample period stops early enough before the onset of the current financial and economic crisis. Third, there were no changes in the federal minimum wage during the sample period, as such changes occurred in September 1997 and in July 2007. Fourth, there are suitable longitudinal weights available for the waves corresponding to the sample period. We drop all observations for individuals whose entries on person characteristics (gender, age, and race) are inconsistent over time, or who were in the Armed Forces at some point during the panel span. Furthermore, we restrict the sample to male workers between the age of 20 and 65 years who were employed at least in one month during the panel span in a job that is neither self-employment nor family work without pay. Our basic sample contains information on 6,444 individuals for whom we can use the longitudinal weight variable *lgtpnwt3*. The corresponding weighted sample size in each month of our sample period is around 66 million people, and the average age is 40 years.

We use the weekly record of a person's labor market status in the second week of a month, *rwkesr2*, in order to categorize individuals as employed, unemployed or not in the labor force for a given month. Our assignment to labor market states is similar to Nagypal (2008). In particular, a person is taken to be employed if the status is *"with a job or business - working"* or *"with a job or business - not on layoff, absent without pay"*. He is recorded as unemployed if the status is *"with a job or business - on layoff, absent without pay"* or *"without a job or business - looking for work or on layoff"*, and as not in the labor force if it is *"without a job or business - not looking or on layoff"*. Based on this classification, the average participation rate in our sample is 93%, while the

[12] http://www.nber.org/data/sipp.html

average unemployment rate is 3.6%.[13] The average monthly rate of transitions from employment to unemployment is 23.3%, and the rate of transitions from employment to unemployment is 0.8%.

The SIPP survey collects information on up to two wage and salary jobs per person for a given wave. Job-specific information is recorded in the following way: A person's most important job is recorded as job 1 throughout the wave. If an individual held more than one job within the wave, the second most important is recorded as job 2, even if the two jobs were never held simultaneously. Whether a person held a particular job during a given month or week within the wave can be inferred from the recorded starting and ending dates of each job. Taking into account only jobs that are neither unpaid family work nor in the Armed Forces, we determine an individual's main job in a given month by the following sequence of priorities: (i) job 1 if it was held in week 2, (ii) job 2 if it was held in week 2, (iii) job 1 if it was held at some point during the given month, and (iv) job 2 if it was held at some point during the month.

Based on our definitions of a person's labor market status and main job, we construct a measure of the number of job-to-job transitions between two consecutive months which comprises all workers who were employed in the second week of both months, were not unemployed in any of the weeks between, held main jobs with different employer identification numbers in the two months, held each of the main jobs in the second week of the respective month, and did not return to a job previously recorded as their main job. The average fraction of employed workers in our sample who make a job-to-job transition between the current and the following month is 1.32%.[14]

Regarding the estimation of residual wages, we further restrict our sample to workers employed in full-time jobs, that is, jobs for which they report to be usually working 35 hours or more per week. We first impute a worker's real hourly wage at his main job in a given month, since the SIPP records some crucial variables only once per wave. For around one-half of the observations in our sample a regular hourly wage rate pertaining to the whole wave is reported for the main job. For the remaining observations, we calculate an average hourly wage over the wave from total earnings in this job over four months, the number of hours typically worked in the job, and the total number of weeks employed in the job throughout the wave. For both types of hourly wages, we use the annual CPI from the BLS to express real hourly wages in constant 2004 dollars. Moreover, in accordance with related empirical studies, we exclude wage observations that fall below one-half of the nominal minimum wage rate as well as observations above $211, a value which corresponds to the 99.9th-percentile of real wages in our sample.[15]

Furthermore, we construct variables reflecting a number of worker characteristics. Following Eckstein and Nagypal (2004), we assign individuals to one of five education groups, based on the highest grade or degree obtained. The five categories correspond to high school dropouts, high school graduates, workers with some college education, college graduates, and post-graduate degree

[13]These figures compare to an average participation rate of 86% and unemployment rate of 4.6% among all men aged 20 to 64 years reported by the Bureau of Labor Statistics (from CPS data) for the same time period. See Nagypal (2008) for a discussion on differences between the CPS and the SIPP regarding the categorization of a person's labor market status which lead to lower records of unemployment in the SIPP.

[14]Our definition of job-to-job transitions closely corresponds to the one used in Menzio et al. (2012), but their definitions of a person's main jobs and labor market status are different.

[15]See for example Katz and Autor (1999) for similar sample restrictions.

holders. For those individuals who report implausible changes in education levels over the panel span (a decrease in education, or a sharp increase that is not associated with temporary school enrolment or with gaps in observations), we impute the number of years of education completed, based on the person's most frequently reported level out of ten finer education categories. Finally, we construct dummy variables for being non-white as well as for all of the four large regions of the United States as classified by the U.S. Census Bureau.

We obtain our estimates of residual wages by running a pooled regression of log real hourly wages on the five broad education groups, the non-white dummy, the region dummies as well as year dummies. The estimated standard deviation of log residuals is 0.51, while the mean and standard deviations of the exponentiated residual wages are 1.15 and 0.76, respectively.

Using the residuals from the above regression, we calculate statistics of workers' log wage changes both within and between jobs. Our observations for wage changes within a job include all workers from the wage sample who held the same main job across two waves of the panel during the sample period. However, since the distribution of log residual wage changes within a job has very long, flat tails on both sides, we exclude estimates below and above three standard deviations from the mean. These cut-offs correspond to decreases by more than 50% and increases by more than 90% in a worker's wage between two waves, and imply dropping around 2% of the sample. The mean and standard deviation of the remaining observations are 0.007 and 0.14, respectively, and the fraction of negative changes is 33.5%.

The observations for wage changes between jobs includes all workers from the wage sample who experienced a job-to-job transition as defined above between two consecutive months during our sample period. Again, we eliminate extreme values by trimming the sample at three standard deviations below and above the mean. This step excludes observations with decreases by more than 75% and increases by more than 290%, and reduces the sample by around 2%. The mean of the remaining observations is 0.029, and the standard deviation is 0.36. Moreover, 38.4% of the log wage changes associated with job-to-job transitions are negative. For the subsample of positive wage changes, the mean and standard deviation are 0.28 and 0.26, respectively.