



Essays on Wage Inequality from a Macroeconomic Perspective

Susanne Forstner

Thesis submitted for assessment with a view to obtaining the degree
of Doctor of Economics of the European University Institute

Florence, February 2013

European University Institute
Department of Economics

Essays on Wage Inequality from a Macroeconomic Perspective

Susanne Forstner

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Examining Board

Prof. Árpád Ábrahám, European University Institute (Supervisor)
Prof. Gueorgui Kambourov, University of Toronto
Prof. Ramon Marimon, European University Institute
Prof. Josep Pijoan-Mas, CEMFI Madrid

© Susanne Forstner, 2013

No part of this thesis may be copied, reproduced or transmitted without prior permission of the author

Abstract

This thesis contains two chapters on the sources of residual wage inequality. The first chapter contributes to attempts at explaining the increase in wage inequality in the U.S. labor market over the past few decades. I address the question of how much of this increase can be attributed to factors associated with job-to-job mobility. For this purpose, I develop a search model with on-the-job search, anticipation of job destruction, and costs to workers when switching jobs. The quantitative analysis involves calibrating the model to match characteristics of the U.S. labor market in the mid-1980s and the mid-2000s. I find that changes in job-to-job mobility have a significant quantitative impact on residual wage inequality. In particular, up to one-half of the observed inequality increase is accounted for by the composite effect of three mobility determinants. Among them, the arrival probability of offers on the job plays the leading role, whereas the impact of job switching costs is negligible. In addition, changes in the conditions of job loss amplify the effect of offer arrivals.

In the second chapter, joint work with Árpád Ábrahám and Fernando Álvarez-Parra, we study the impact of moral hazard in labor contracts on residual wage inequality. The tool of our analysis is a search model with job-to-job mobility and firm competition for workers, where firms offer long-term contracts to risk-averse workers in the presence of repeated moral hazard. For a quantitative analysis, we calibrate the model to match characteristics of the U.S. labor market derived from micro data from the mid-2000s. We find that, on balance, moral hazard increases residual wage inequality by around six percent. The direct effect of providing incentives through wage variation accounts for a moderate contribution to inequality increase. In addition, moral hazard affects the wage distribution through several indirect effects, as firms adjust the levels of effort implemented and the wage offers made to workers in response to increased effort costs. Through their particularly strong impact on the lower parts of the wage distribution, such effects contribute substantially to the overall rise in inequality. The main reason is that, under moral hazard, low wage workers spend significantly less effort.

Acknowledgements

First and foremost, I want to thank Árpád Ábrahám, my main advisor, for his continuous guidance and support. The present thesis has benefited enormously from our frequent discussions. His encouragement as well as his advice in many dimensions of economic research have been invaluable throughout the process.

I am also indebted to Ramón Marimón, my second advisor, for most valuable comments and suggestions, and, in particular, for remarks that induced me to think about the broader context of my work.

In writing this thesis, I have benefited from comments by many more researchers. Among those, I am particularly grateful to Jérôme Adda, Russell Cooper, Giancarlo Corsetti, Georg Dürnecker, Nicola Pavoni, and participants at the EUI Macro Student Lunch and the Vigo Workshop on Dynamic Macroeconomics 2011 for useful discussions and comments.

Last, but not least, I want to thank my friends and family, Alexis, Harald, Kathi, Vera, Anneliese, Helmut, and Martina. Their support, their criticism, and their company despite our geographical distance, have made the last few years an especially rewarding experience.

Contents

Preface	1
1 Job-to-job mobility and wage inequality: a quantitative assessment	5
1.1 Introduction	5
1.2 Empirical background	9
1.2.1 Wage inequality	9
1.2.2 Worker turnover	12
1.2.3 Factors affecting job-to-job mobility	15
1.3 The model	16
1.3.1 The environment	16
1.3.2 The workers' optimization problem	18
1.3.3 Equilibrium	20
1.4 Quantitative Analysis	20
1.4.1 The experiments	20
1.4.2 Functional forms and distributions	21
1.4.3 A baseline calibration	21
1.5 Results	25
1.5.1 The main experiment	25
1.5.2 A supplementary experiment	28
1.5.3 Discussion	30
1.6 Concluding remarks	34
A Data Appendix	35

2	The effect of moral hazard on wage inequality with on-the-job search and employer competition	39
2.1	Introduction	39
2.2	The Model	44
2.2.1	The setup	45
2.2.2	The optimal contract	48
2.2.3	On continuation values in the absence of outside offers	57
2.2.4	Stationary equilibrium	58
2.3	Analysis of a Simplified Model	59
2.4	Quantitative Analysis	63
2.4.1	Calibration	63
2.4.2	Computational experiments	69
2.5	Concluding Remarks	81
B	Appendix	83
B.1	Proof of Lemma 2	83
B.2	Data	86

Preface

The papers of this thesis contribute to the research on wage inequality. As is well known, the analysis of economic inequality, of which unequal labor incomes represent one particular aspect, has a broad and deep background extending over several fields. To lay it out in detail would require an outline of some conceptions from modern political philosophy and of different perspectives on the assessment of inequalities. As regards the income dimension of economic inequality, unequal wages are known to determine to a large extent the inequality in wealth and consumption across households because labor income is the main source of income for most of them. Finally, the empirical evidence of substantial wage variation across workers who do not differ in terms of observable characteristics, points out clearly – in addition to the standard equity concerns – economic efficiency issues too.¹

The two chapters outlined below have, on the one hand, in common the focus on sources of wage inequality across observedly identical workers and the emphasis on a quantitative approach. On the other hand, they offer two different perspectives on their main subject. One study is about observed change in frictional wage inequality over time and concentrates on a group of three determinants simultaneously. By contrast, the other study focuses on the extent of wage inequality within a given period and singles out one specific informational friction in worker-firm relations for analytical scrutiny.

The first chapter makes a contribution to the various attempts to explain the increase in wage inequality in the U.S. labor market over the past few decades. More specifically, I address the question of how much of the observed rise in residual wage inequality can be associated with the parallel increase in job-to-job mobility. The motivation to examine this relationship derives from the close correspondence in search models of the labor market between the factors determining worker mobility and those shaping the wage distribution.

¹A concise presentation of this view is found in Mortensen (2003).

For the quantitative analysis required, I develop a search model with on-the-job search, anticipation of job destruction, and costs to a worker when leaving his employer. The model, besides incorporating several determinants of job-to-job mobility, allows for wage loss upon job change, a feature which is commonly observed in data on job-to-job transitions. Starting from a calibration of the model to match certain characteristics of the U.S. labor market in the mid-1980s, I re-calibrate, in a second step, selected parameters in such a way that the model matches changes in worker turnover between the mid-1980s and the mid-2000s. As a result of this computational experiment, I find that changes in job-to-job mobility have a significant quantitative impact on wage inequality across ex-ante identical workers. In particular, up to one-half of the empirically observed increase in residual wage inequality is accounted for by the composite effect of three determinants of job-to-job mobility. Among them, the arrival probability of offers on the job adopts the leading role, whereas the impact of job switching costs turns out to be negligible. In addition, changes in the conditions of job loss – represented by the probability of receiving notice – amplify the impact of the arrival probability. As the main feature of the mechanism behind the rise in inequality, I identify a shift from voluntary to involuntary job-to-job transitions, which originates from an increase of offer arrivals on the job and a parallel decrease of the probability of job loss.

The second chapter, which is joint work with Árpád Ábrahám and Fernando Álvarez-Parra, focuses on the impact the presence of moral hazard in labor contracts exerts on the cross-sectional wage distribution. More specifically, we attempt to quantify the effect of moral hazard on residual wage inequality, break the total effect down into several components, and explore the different mechanisms which together produce the overall impact. The tool of our analysis is a search model enriched by features of job-to-job mobility and firm competition for workers à la Postel-Vinay and Robin (2002). In this framework, firms offer to risk-averse workers long-term contracts, taking into account the presence of repeated moral hazard, which arises from the informational friction of non-observability of worker effort. In this situation, an optimal labor contract prescribes the provision of incentives to encourage effort through variation of wages across output realizations. In view of this mechanism of incentive provision through wage dispersion, intuition suggests that moral hazard should increase residual wage inequality. In our

quantitative analysis we perform a series of computational experiments which are based on a benchmark calibration of the model to U.S. labor market data of the mid-2000s. Our main result is that, on balance, moral hazard increases residual wage inequality by around six percent. Quantitative experiments intended to disentangle partial effects and to explore underlying mechanisms suggest that the direct effect of providing incentives to encourage effort through wage variation accounts for a moderate contribution to inequality increase. In addition, non-observability of effort affects the wage distribution through several indirect effects which originate from changes in effort costs to firms. In the presence of moral hazard, effort is more expensive, so that firms adjust both the levels of effort implemented and the wage offers made to workers coming from unemployment or from other firms. Among these indirect effects, the one most closely linked to outside wage offers and firm competition weakly counteracts the increase of inequality. By contrast, through their particularly strong impact on the lower parts of the wage distribution, other indirect effects contribute substantially to the overall rise in inequality, due to the fact that, under moral hazard, firms demand significantly lower levels of effort from low wage workers.

Chapter 1

Job-to-job mobility and wage inequality: a quantitative assessment

1.1 Introduction

Since the late 1970s the labor market in the United States has exhibited a persistent widening of the overall wage distribution, a trend which was to a large extent due to an equally persistent increase in the dispersion of wages among observationally equivalent workers. For this rise in residual wage inequality a number of potential explanations have been put forward. Most prominent among them are those based on the arguments that the returns to unobservable worker characteristics have grown, that the economy has become more turbulent in the sense of workers' job-specific skills depreciating more strongly upon job loss, or that the overall labor market risk for workers has risen substantially.

In the present paper I investigate another potential source of increase in residual wage inequality, namely, a rise in job-to-job mobility. A straightforward motive behind examining worker transitions from this angle is that in search models of the labor market the factors determining labor mobility correspond closely to those shaping the wage distribution. Another motive derives from the fact that the rise in wage inequality in the United States between the late 1970s and the mid-2000s occurred simultaneously with an increase in the rate of job-to-job transitions.

Among the determinants of job-to-job mobility are the level of search technology, matching efficiency and costs that workers incur when changing employers. At the same

time, mobility determinants such as these have the potential to affect the wage distribution through changes in reservation wages and in the composition of worker transitions. For instance, in a basic search model with on-the-job search both the arrival probability of offers on the job and job switching costs have an effect on the reservation wage of unemployed workers. While, on the one hand, a rise in the above probability lowers this reservation wage, an increase in job-switching costs, on the other hand, raises it.

In the present analysis, I examine in quantitative terms the relationship between the increase in job-to-job mobility and the rise in wage inequality in the United States between the mid-1980s and the mid-2000s. In particular, I address the following question: "How much of the observed increase in residual wage inequality can be associated with increased job-to-job mobility?"

In order to propose an answer, I develop a job search model with on-the-job search, anticipated job destruction, and job switching costs. The model not only features various determinants of worker turnover, but also allows for job-to-job transitions involving wage loss, where the latter type of transition arises from workers' anticipation of job destruction and their pre-emptive job search during a notice period. Within this framework, I quantify the impact of determinants of job-to-job mobility on wage inequality by the following counterfactual experiment. First, I calibrate the model with a view to matching characteristics of the U.S. labor market in the mid-1980s. In a second step, I recalibrate only parameters closely linked to the turnover of employed workers in order to match observations on worker mobility in the mid-2000s. A comparison of cross-sectional wage inequality between the two scenarios provides a quantification of the impact of job-to-job mobility on inequality. For the major part of this analysis, I use micro data from the Survey of Income and Program Participation.

I find that, within this framework, changes in job-to-job mobility have a significant quantitative impact on residual wage inequality. More specifically, up to one-half of the observed inequality increase between the first and the second time period are accounted for by the response of workers to simultaneous change in three determinants of job-to-job mobility. Among them, the arrival probability of offers on the job adopts the leading role, whereas the impact of job switching costs turns out to be negligible. Changes in the conditions of job loss – represented by the probability of receiving notice – amplify the

impact of the arrival probability.

In the remainder of this section, I provide a selective overview of the literature related to the present work and outline my contribution. The foundation for the framework of analysis is the large literature on search-theoretic models of the labor market, in which wage differences across ex-ante identical workers emerge as a consequence of market frictions. In a recent paper, Hornstein et al. (2011) assess the degrees to which the basic job search model and various extensions to it, when plausibly parameterized, are compatible with empirical evidence on the extent of wage dispersion. One of their findings is that the inclusion of on-the-job search in the model is crucial to obtaining plausible levels of wage dispersion.

As regards studies of the relationship between job-to-job mobility and the wage distribution, the present analysis is related to Jolivet et al. (2006) and Postel-Vinay and Robin (2002). The first paper attempts, in the context of a cross-country study, a broad empirical validation of the correspondence between determinants of labor turnover and determinants of overall wage distributions. The second paper evaluates the contribution of labor market frictions to wage inequality for a French data set. The model framework used by the authors features on-the-job search in the presence of inter-firm competition for employed workers.

A common characteristic of labor market transition data is a significant fraction of job-to-job transitions associated with a wage loss. At least three different approaches to modeling this feature are found in the literature. While one of these (Dey and Flinn (2005)) builds on compensating differentials, another one (Postel-Vinay and Robin (2002)) invokes expectations of future wage growth as a reason to accept a wage cut upon job change. A third modeling approach (Jolivet et al. (2006)) introduces exogenous reallocation shocks with a view to capturing immediate re-employment after layoff. By contrast, in the present model setup, job loss is in all cases preceded by pre-announcement, which leads to pre-emptive job search by the worker concerned during a notice period. This approach, amounting to a more elaborate micro-modeling of the concept of reallocation shocks, follows Bowlus and Vilhuber (2002) and Nagypal (2008).

An early theoretical analysis of job search with job changing costs is that of Hey and McKenna (1979), who focus on constant costs to workers of switching jobs. Burgess

(1992) presents a theoretical analysis of the implications of job-changing costs for aggregate unemployment. More recently, Cabrales et al. (2008), in the context of studying the implications of equity concerns for labor market outcomes in general, examine the impact of changes in mobility costs on wage dispersion both within and across firms.

Finally, with regard to explaining the increase in residual wage inequality in the United States since the late 1970s, the papers by Violante (2002) and Kambourov and Manovskii (2009) are among the most prominent contributions to the debate on this topic. Both studies emphasize the relevance of job-specific skills and rely in their respective mechanisms on a notion of increased labor market turbulence which leads to a rise in wage inequality.

The present paper makes a quantitative contribution to the discussion on frictional wage dispersion on several points. First, I calibrate a parsimonious search model which takes into account explicitly both mobility costs and the anticipation of job loss. In this connection, pre-emptive job search as a worker's response to the threat of imminent layoff adds another dimension to the nature of job-to-job transitions, while the inclusion of mobility costs broadens the basis for quantifying the impact of job-to-job transitions. Second, I show that change in determinants of job-to-job mobility – defined in a broad sense – did have a significant quantitative impact on residual wage inequality in the U.S. labor market. Third, in the present framework, inequality increase arises from a mechanism which is different from the mechanisms proposed in Violante (2002) and Kambourov and Manovskii (2009). While those papers rely on increased turbulence as a source of the observed rise in inequality, my analysis centers on better chances for workers to move up the wage ladder. Finally, the computational experiments of the exercise provide a quantification of the contribution of several mobility determinants to the observed increase in U.S. wage inequality between the mid-1980s and the mid-2000s.

The paper is organized as follows: Section 1.2 outlines the empirical background to the present work. Section 1.3 presents the model used in the analysis, states the optimization problem, and defines stationary equilibrium. In Section 1.4 I discuss in some detail how the model is put to use in a quantitative analysis. The results of the analysis are presented and discussed in Section 1.5, and Section 1.6 concludes.

1.2 Empirical background

In this section, I provide an overview of empirical evidence on developments in the U.S. labor market which motivate the questions addressed in the present paper, inform the design of the framework for the analysis, and outline its empirical basis. First, I present facts about change in wage inequality in the United States over the past few decades. Then, I move to empirical information on worker turnover, in particular, the evolution of job-to-job mobility since the late 1970s. This information is supplemented by data on transitions from employment to unemployment. Finally, I discuss some evidence on determinants of worker mobility and highlight developments that are likely to have affected these underlying factors.

1.2.1 Wage inequality

A large number of empirical studies have documented the substantial and persistent widening of the cross-sectional distribution of wages in the United States from the 1970s onwards. Katz and Autor (1999) provide a comprehensive survey of the pertinent literature up to the late 1990s. In addition, they present their own estimates from March Current Population Survey (CPS) data, which produce a picture similar to the findings of other authors. Between the late 1970s and the mid-1990s, weekly labor earnings of the 90th-percentile relative to the 10th-percentile worker increased by over 25 percent for both men and women. Eckstein and Nagypal (2004) re-examine the main trends in the U.S. wage distribution, using March CPS data from 1962 up to 2003. They find a large and persistent increase in overall wage inequality of men as well as women, starting in the late 1970s and continuing throughout the period up to the early 2000s. Figure 1.1 shows the evolution of the overall standard deviation of log wages of men in their sample.

Part of the observed increase in wage inequality in the United States over the last 35 years can be explained by changes in relative wages of different groups of workers, distinguished by observable characteristics such as sex, education, and experience.¹ For example, the rise in returns to skills since the late 1970s, measured by the increase in the college wage premium, has received broad attention as a source of the widening of the

¹Katz and Autor (1999) and Eckstein and Nagypal (2004) also present detailed analyses of the changes in wage differentials between groups.

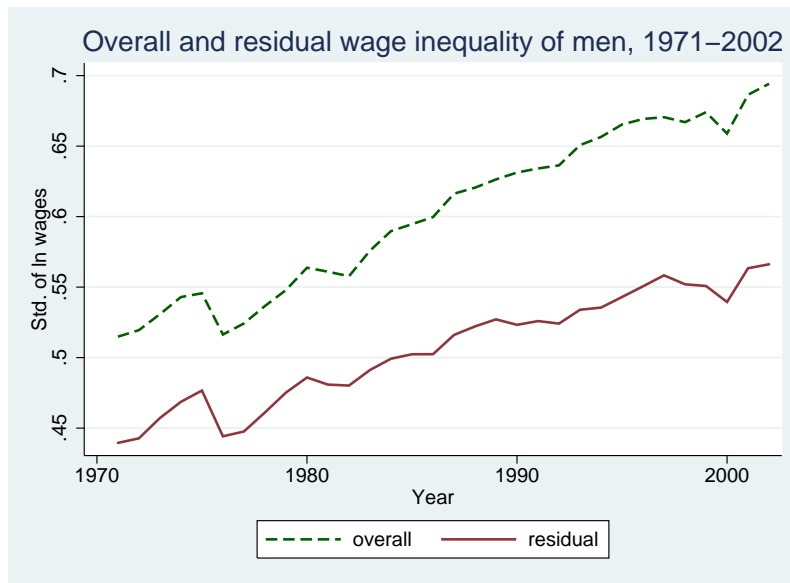


Figure 1.1: Standard deviations of log weekly wages, and of residuals of wage regressions, March CPS 1971 to 2002. Data and estimations from Eckstein and Nagypal (2004).

wage distribution. In addition, a substantial increase in the dispersion of wages within narrowly defined demographic and skill groups of workers was recorded for the same time period.² For instance, Katz and Autor (1999) analyze the evolution of the residuals from Mincerian wage regressions using March CPS data from 1964 to 1996. They report that the ratio of the 90th-to-10th percentile of residual wages has increased significantly for men since the mid-1970s, and for women since the 1980s.³ Eckstein and Nagypal (2004) estimate similar wage regressions using data up to 2003 and controlling in addition for occupational categories.⁴ The results confirm the findings of Katz and Autor (1999): The within-group standard deviation of weekly wages for men increased by around 23%, from an average of 0.46 in the late 1970s to 0.57 in the early 2000s. The same measure of residual inequality for women started to increase only in the 1980s, rising by about 20%,

²See Juhn et al. (1993) for an assessment of the contribution of residual wage inequality to the overall increase in wage inequality.

³Katz and Autor (1999) estimate separate regressions for men and women, for each year, of log weekly wages on eight education dummies, a quartic in labor market experience, interaction terms between three broad education groups and the quartic in experience, three regional dummies, and two race dummies. Their sample includes only full-time full-year wage and salary workers.

⁴Eckstein and Nagypal (2004) estimate regressions for men and women, for each year, of log weekly wages on five education dummies, a quadratic function in experience, four regional dummies, two race dummies, and three occupational dummies. Their sample includes only full-time full-year wage and salary workers between 22 and 65 years of age.

from an average of 0.42 in the early 1980s to 0.50 in the early 2000s.⁵ Estimation results for residual wage inequality of men from Eckstein and Nagypal (2004) are depicted in Figure 1.1. They demonstrate how closely the increase in residual wage inequality paralleled the rise of overall inequality over the period from the late 1970s to the early 2000s.

Since the focus of the present analysis is on change in residual wage inequality, I summarize in the following two tables empirical information on the rise in inequality between two benchmark periods. Both tables report the same three inequality measures, but differ with respect to data sources. While Table 1.1 shows levels of residual wage inequality of men as estimated from CPS data by Eckstein and Nagypal (2004), Table 1.2 provides comparable figures which I estimate from SIPP data sets.⁶ The inequality measure reported in the last row of each table, the mean-to-fifth-percentile ratio, is used throughout the paper as a proxy for the mean-to-min ratio.

Table 1.1: Measures of residual wage inequality of men (CPS)

	1985	2002	Change (%)
Std($\ln w$)	0.502	0.566	12.70
q90/q10(w)	3.142	3.644	15.98
Mean/q05(w)	2.576	2.876	11.65

Table 1.2: Measures of residual wage inequality of men (SIPP)

	1985	2005	Change (%)
Std($\ln w$)	0.437	0.49	11.91
q90/q10(w)	2.927	3.15	7.49
Mean/q05(w)	2.332	2.45	4.96

The percentage increase in inequality measures between the two periods is substantial,

⁵Kambourov and Manovskii (2009) and Heathcote et al. (2010) find very similar time patterns for the evolution of the within-group variance of male hourly wages since the 1970s. Their data are taken from the Panel Study of Income Dynamics, 1969 to 1996, and 1967 to 2002, respectively.

⁶Details concerning the SIPP data sets used in the present paper, sample selection and construction of variables, as well as estimation specifications are documented in the Appendix A.

ranging from around 5% for the SIPP-based mean-to-min ratio to around 16% for the CPS-based ninety-to-ten ratio. Increases are seen to differ across measures and between data sources with estimates from CPS data being consistently higher than their SIPP counterparts. In Section 1.5 below the figures of Tables 1.1 and 1.2 provide the empirical benchmark against which the results of computational experiments are evaluated.

1.2.2 Worker turnover

The major part of this subsection on labor market transitions of employed workers is about moves between employers. The importance of these job-to-job transitions in the process of reallocating labor in the U.S. economy has recently been documented by a number of empirical studies. For instance, Fallick and Fleischman (2004) and Nagypal (2008), analyzing monthly data from the CPS since 1994, find that the monthly rate of job-to-job transitions is on average two to three times larger in magnitude than that of monthly flows of workers from employment to unemployment. For the time period from January 1994 to December 2003, Fallick and Fleischman (2004) report that each month on average 2.6 % of employed workers left one employer for another, whereas only 1.3 % of employed workers moved into unemployment. Furthermore, they find that an average of two-fifths of new jobs started over the above period were associated with direct employer changes. Nagypal (2008), using data up to August 2007, estimates that within a month on average 2.9 % of employed workers change their employer, and 0.9 % move to unemployment.⁷

Moreover, a number of empirical analyses document that significant changes in the size and composition of worker flows in the United States have taken place during the past few decades. Stewart (2002) is among the first authors who measure, in a consistent manner, the incidence of job separation as well as its decomposition into different flows over a long time period. His measures are based on March CPS data from 1977 to 2001. In his analysis of the evolution of the rate of job-to-job transitions since the mid-1970s, this author finds that the incidence of such transitions increased steadily. Thus, the average fraction of employed workers experiencing an employer-to-employer transition within a year rose from 8.6 % in 1975 to 13.7 % in the year 2000. Sherk (2008), who applies the

⁷Nagypal (2008) obtains similar results from data of the Survey of Income and Program Participation for the same time period.

same method of analysis to CPS data up to the year 2007, reports that the fraction of job-to-job changers increased to roughly 12.5 % in 2007.

In view of the emphasis of the present analysis on job-to-job mobility as a source of rising wage inequality, I present in Figure 1.2 some basic information in a compact form. The figure shows the time series of the annual incidence of job-to-job transitions of men from 1976 to 2007 as estimated by Sherk (2008) together with its linear and HP-filtered trends.⁸ Using linear-trend values estimated on the basis of this time series, I find that the increase in the transition rate between the benchmark years 1985 and 2005 amounts to 14%. This result provides the empirical part of the motivation for examining the relationship between parallel increases in job-to-job mobility and wage inequality. In addition, in the quantitative analysis laid out in Section 1.4 empirical observations on the rate of job-to-job transitions play a central role.

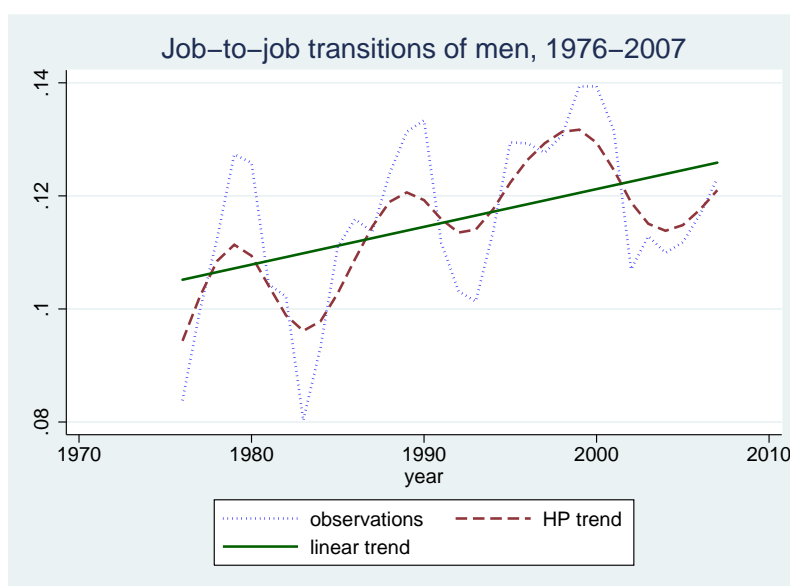


Figure 1.2: Evolution of the annual incidence of job-to-job transitions of men. The HP trend reflects a smoothing parameter of $\lambda = 6.25$. Based on estimations by Sherk (2008), March CPS 1976 to 2007.

The analysis presented in this paper distinguishes between two different types of job-to-

⁸Measurement of the job-to-job transition rate in Sherk (2008) is based on the methodology of Stewart (2002). The variable reported is the estimated number of workers who experienced at least one job-to-job transition over the time period of roughly one year, divided by the number of workers who were employed for at least one week during that year. James Sherk kindly provided his estimates, as well as documentation on the estimation procedure.

job transition, according to whether a move is associated with wage gain or wage loss. In connection with modeling this distinction, transitions from employment to unemployment need to be taken into account as well. Since change in these transitions has an impact on the main result of Section 1.5, the evolution of the corresponding transition rate over the recent past is briefly outlined here. Figure 1.3 shows the time series of the annual incidence of job-to-unemployment transitions of men from 1976 to 2007 - as estimated by Sherk (2008) - together with its linear and its HP-filtered trends. The figure displays a strong downward trend over the three decades surveyed here.

As in the case of job-to-job transitions, linear-trend values, estimated on the basis of this time series, can be used to assess change in the rate of job-to-unemployment transitions between the benchmark years 1985 and 2005. The decline in the transition rate over this twenty-year time span amounts to roughly -40% .

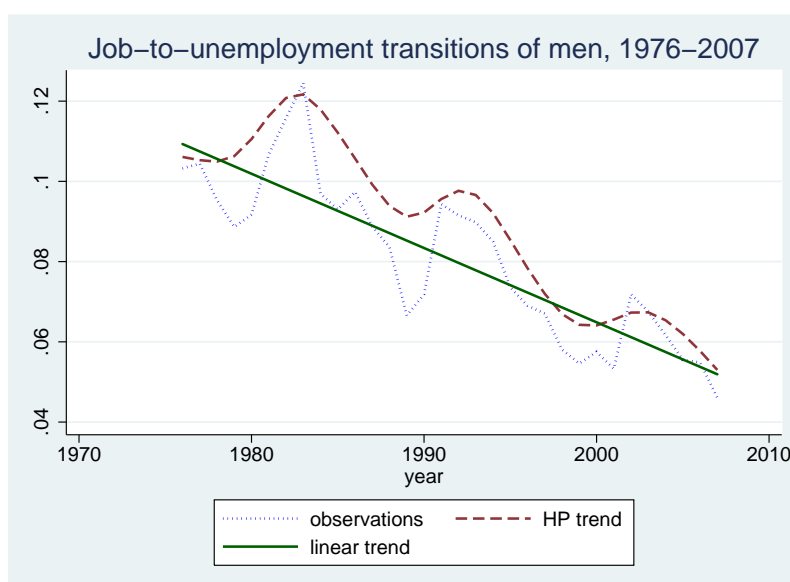


Figure 1.3: Evolution of the annual incidence of job-to-unemployment transitions of men. The HP trend reflects a smoothing parameter of $\lambda = 6.25$. Based on estimations by Sherk (2008), March CPS 1976 to 2007.

The persistent decline in the rate of employment-to-unemployment transitions in the United States between the late 1970s and the mid-2000s has been documented in many other empirical studies.⁹ Moreover, in a recent paper, Davis et al. (2010) examine the relationship between the decline in unemployment inflows and the parallel decline in vari-

⁹See Davis (2008) for a survey of empirical analyses.

ability of employment growth, job destruction, and job reallocation at the establishment level. One of their findings is that declining job destruction at the establishment level accounts for around 30% of the decline in employment-to-unemployment flows between the early 1980s and the mid-2000s.

1.2.3 Factors affecting job-to-job mobility

Among potential explanatory factors behind the observed increase in the rate of job-to-job transitions are improvements in search technology, increase in the efficiency of job-worker matching and reduction of mobility costs to workers. In the following paragraphs I sketch some empirical facts that can be linked to these three factors.

The level of search technology applied in the labor market must have risen significantly over the recent past, mainly due to the rapid spreading of Internet job search. Among the channels used to locate a job are Internet job boards and corporate websites which permit on-line job applications. In particular, growth in the use of job boards has been high, and projections are for this channel to replace newspaper advertisements in the not too distant future.¹⁰ The shift from newspaper to Internet implies a significant improvement of search technology in several dimensions. One is the increase in scope as well as speed of job search, since on the Internet a worker can screen a larger number of potentially interesting jobs much more rapidly than through other media. Another consequence of the growing presence of the Internet in job search is a higher efficiency of matching workers to jobs. Among the reasons for more efficient matching through Internet-based job search are the possibilities of having more initial meetings between worker and potential employer and of on-line screening of candidates.

The last factor in the previous list hints at the fact that changing employers is associated with transaction costs to the worker. These costs, which can be of a pecuniary or a psychological nature, arise from leaving the old job or from accepting a new one. A prime example for mobility costs to workers switching from one employer to another is the lack of full transferability of employer-based pension plans. In the United States, benefits from such plans make up a significant part of workers' retirement income.¹¹ Employer-based

¹⁰See Autor (2001).

¹¹Payments from these pension plans are additional to retirement income provided through the publicly administered Social Security. Munnell and Perun (2006) report that in 2004, employer-sponsored pension

pension plans can be either defined benefit (DB) or defined contribution (DC) systems. Upon changing employers, workers lose most of their pension claims accumulated within a DB pension plan, which is not the case for defined contribution plans. Starting in the late 1970s, a number of changes to U.S. legislation regulating the tax status of employer-based pension plans have induced a general shift from defined benefit to defined contribution plans.¹² This implies that today the average U.S. worker faces much lower job changing costs associated with the loss of nontransferable pension claims than thirty years ago.

1.3 The model

1.3.1 The environment

Time is discrete and indexed by $t = 0, 1, 2, \dots$. The economy is populated by a large number of ex-ante identical, risk-averse individuals who derive utility from consumption and suffer disutility when leaving a particular job. Individuals can be either employed or unemployed, but have no labor force participation choice. When employed, they receive a period wage w_t , whereas when unemployed, they receive a constant level b of unemployment benefits. Individuals cannot save, thus, they always consume their period income. Workers search for jobs both while unemployed and while employed, with job search being exogenous and costless to workers. Furthermore, at the end of each period, workers face a constant probability σ of dying, in which case they exit the economy and are replaced by unemployed, new agents. The size of the worker population is therefore constant over time and can be normalized to one.

An unemployed worker searches for a job in such a way as to receive in each period one job offer with exogenous probability λ_u , and no offer with probability $(1 - \lambda_u)$. A job offer is associated with a productivity draw z^o from a known constant distribution with cumulative density function $F(z^o)$. If the unemployed worker accepts the offer and starts

income accounted for 19% of total income for people above 65 years, and that private pension plans made up 24% of the wealth holdings of a "typical" household shortly before retirement.

¹²Most importantly, the Revenue Act of 1978 introduced the so-called "401(k)" type of employer-based defined-contribution pension plans. Since the early 1980s, the number of 401(k) plans has grown rapidly and continually, so that by now they account for the majority of employer-based DC plans in use (see Poterba et al. (2007) for detailed empirical evidence). Moreover, the relative importance of DB and DC types among private pension plans has been reverted since the early 1980s (see Buessing and Soto (2006) and Munnell et al. (2006) for detailed empirical evidence).

working in this job at the beginning of the next period, he will produce output z^o per period in this job.

Employed workers are paid their marginal product, therefore wages are equal to the worker's productivity in the current job, z . Furthermore, they search for alternative job opportunities while being employed. In analogy to an unemployed worker, an employed worker receives exactly one job offer per period with exogenous probability λ_e . This offer is associated with a productivity draw z^o from the same offer distribution as faced by unemployed workers, $F(z^o)$. If a worker accepts an outside job offer z^o , he starts working at the new job with productivity z^o at the beginning of the next period.

Existing jobs may be destroyed exogenously, but only with advance notice to the worker. This feature is modelled in the following way: Each job-worker pair is characterized, in addition to the worker's productivity, by a state indicator variable s_t . The indicator is set to $s_t = 1$ for a normal employment position for which no notice of layoff has been given in the current period. It switches to $s_t = 2$ as soon as a worker receives notice of imminent layoff. Within a given period, workers in normal jobs receive such a notice with probability η . Once notice of layoff has been given, the job concerned is destroyed at the end of a period with probability δ . New employment relationships always start as normal jobs. Moreover, once notice of layoff has been given, the job cannot revert to the status of a normal job.

The feature that changing employers is associated with transaction costs to workers is captured by introducing a one-time utility loss workers suffer whenever they leave a job. This setup reflects the idea of workers losing part of their employer-based pension claims. For the sake of parsimony, job leaving costs are modelled as being constant across job-worker pairs and are denoted by χ . Workers incur this utility loss at the time they leave a job, be it because they are laid off, because they voluntarily quit to unemployment, or because they leave their employer for another job.

The timing within a period is the following: At the beginning of the period, each employed worker observes the current status of his job s_t . Based on this, he decides whether to continue in the present job or leave to unemployment and incur the associated job leaving costs. If he decides to stay, production takes place, and the worker receives the period wage. Then, both unemployed and employed workers receive a new job offer z^o

with the respective probabilities, and decide whether to accept the offer or stay in their current position. If the offer is accepted, the worker starts the new job at the beginning of the next period at the already observed productivity z^o . Employed workers who choose to accept a new job offer incur the job leaving costs at the beginning of the next period. At the end of the current period, workers having received notice of layoff are dismissed with probability δ , and workers currently employed in normal jobs receive notice with probability η . Finally, workers die with probability σ , and are replaced by unemployed individuals.

1.3.2 The workers' optimization problem

A worker's period utility from consumption is denoted by

$$u(c_t) \tag{1.1}$$

where $u(\cdot)$ is increasing and concave. Consumption c_t is given by z_t if the worker is employed, and by b if he is unemployed. In addition, a worker incurs utility cost χ every time he leaves a job. Workers discount future periods with the common discount factor $\beta \in (0, 1)$.

A job is characterized by the combination of productivity z and status s , where $s = 1$ indicates a normal job, and $s = 2$ signifies that the worker has received notice of layoff. Let $\mathcal{W}(z, s)$ denote the beginning-of-period value to a worker of continuing in his current job (z, s) , and let \mathcal{U} denote the beginning-of-period value to a worker of being unemployed.

The value function associated with the problem of a worker holding a job (z, s) at the beginning of the period is given by

$$\bar{\mathcal{W}}(z, s) = \max \left[\underbrace{\mathcal{W}(z, s)}_{\text{stay}}, \underbrace{\mathcal{U} - \chi}_{\text{quit}} \right] \tag{1.2}$$

The value of being unemployed is

$$\mathcal{U} = u(b) + \beta(1 - \sigma) \left\{ (1 - \lambda_u)\mathcal{U} + \lambda_u \int \max \left[\underbrace{\bar{\mathcal{W}}(z^o, 1)}_{\text{accept}}, \underbrace{\mathcal{U}}_{\text{reject}} \right] dF(z^o) \right\} \tag{1.3}$$

The value of being employed in a job (z, s) is

$$\mathcal{W}(z, s) = u(z) + \beta(1 - \sigma) \left\{ (1 - \lambda_e)\Phi(z, s) + \lambda_e \int \max \underbrace{[\overline{\mathcal{W}}(z^o, 1) - \chi]}_{\text{accept}}, \underbrace{\Phi(z, s)}_{\text{reject}} dF(z^o) \right\} \quad (1.4)$$

where $\Phi(z, s)$ is the expected continuation value of a job of type (z, s) . It is given by the following expressions for the two possible values of job status:

$$\Phi(z, 1) = (1 - \eta)\overline{\mathcal{W}}(z, 1) + \eta\overline{\mathcal{W}}(z, 2) \quad (1.5)$$

$$\Phi(z, 2) = \delta[\mathcal{U} - \chi] + (1 - \delta)\overline{\mathcal{W}}(z, 2) \quad (1.6)$$

The optimal decisions of workers, both employed and unemployed, can be characterized by reservation productivities. For unemployed workers, the reservation productivity z^u above which he accepts a job offer z^o satisfies

$$z^o > z^u \Leftrightarrow \overline{\mathcal{W}}(z^o, 1) - \mathcal{U} > 0$$

Similarly, the reservation productivity $z^e(s)$ below which an employed worker decides to quit his current job, satisfies

$$z > z^e(s) \Leftrightarrow \mathcal{W}(z, s) + \chi - \mathcal{U} > 0$$

Finally, the reservation productivity $z^s(z, s)$ above which a worker currently employed at (z, s) accepts an outside job offer z^o satisfies

$$z^o > z^s(z, s) \Leftrightarrow \overline{\mathcal{W}}(z^o, 1) - \chi - \Phi(z, s) > 0$$

1.3.3 Equilibrium

I analyze stationary equilibria of the present model, which consist of the following components:

1. the value function $\bar{W}(z, s)$ with associated functions $\mathcal{W}(z, s)$ and \mathcal{U} , corresponding to the solution of the workers' problem,
2. policy functions for workers' decisions on which jobs to accept when unemployed, in which job states to voluntarily quit to unemployment, and which outside job offers to accept when employed, summarized by the respective reservation productivities z^u , $z^e(s)$, and $z^s(z, s)$,
3. and a stationary distribution $\mu(z, s, e)$ of workers over productivity levels z , job statuses s , and the states of being employed ($e = 1$) and unemployed ($e = 0$).

1.4 Quantitative Analysis

The main goal of this analysis is to examine the relationship between the observed increase in workers' job-to-job mobility in the United States, and the observed increase in residual wage inequality over the period from the mid-1980s to the mid-2000s. The subsections below describe in some detail how the model of Section 1.3 is put to use in an attempt to provide an answer to the above question. At the core of the present analysis are computational experiments which demand calibration of the underlying model. In this context, it is required to make assumptions about functional forms and distributions, to find a solution for the endogenous variables in terms of exogenous variables and parameters, and to select values for the parameters of the model.¹³

1.4.1 The experiments

In the context of the present analysis three different computational experiments are carried out. All of them start from a baseline calibration of the model to match a number

¹³This outline of the quantitative analysis adopts the conceptual framework and terminology of Canova (2007), Chapter 7.

of features of the U.S. labor market in the mid-1980s. Each experiment involves a recalibration to empirical statistics from the mid-2000s. According to their different analytical goals, experiments vary with respect to the groups of free parameters. The first experiment explores the combined effect of three mobility parameters on wage inequality. The second exercise focuses on the role of only one of these parameters – the arrival rate of job offers to employed workers – only. Finally, the model is recalibrated in full analogy to the baseline calibration in order to obtain a comprehensive assessment of parameter change between the two time periods.

1.4.2 Functional forms and distributions

In order to analyze the model of Section 1.3 numerically, I make the following assumptions on functional forms and distributions:

The period utility of workers from consumption is of the CRRA class,

$$u(c) = \begin{cases} \frac{c^{1-\gamma}-1}{1-\gamma} & \text{for } \gamma \neq 1 \\ \ln(c) & \text{for } \gamma = 1 \end{cases} \quad (1.7)$$

where γ is the coefficient of relative risk aversion.

Wage offers z^o are drawn from a log-normal distribution with parameters μ_z and σ_z^2 , so that

$$\ln(z^o) \sim \mathcal{N}(\mu_z, \sigma_z^2) \quad (1.8)$$

1.4.3 A baseline calibration

For a given choice of parameter values, a discretized version of the model is solved by standard value function iteration. The distribution of workers over wages and states of employment in stationary equilibrium is obtained through iteration on transition probabilities. Given the assumptions of the previous subsection, values for the following eleven parameters need to be selected:

1. γ the coefficient of relative risk aversion,
2. β the time discount factor,
3. σ the probability of dying,
4. b the level of unemployment benefits,
5. λ_u the probability of receiving a job offer while unemployed,
6. λ_e the probability of receiving a job offer while employed,
7. μ_z the mean of the log wage offer distribution,
8. σ_z^2 the variance of the log wage offer distribution,
9. η the probability of an employed worker receiving notice of imminent job destruction,
10. δ the probability of a job being destroyed at the end of the current period, and
11. χ the utility cost of leaving a job.

Since the mechanism of the present model focuses on individuals' labor market transitions, which are observed to occur with high frequency, I choose the length of the model period to be one month. Values for five parameters (γ , β , σ , λ_u , and δ) are set exogenously, while for the other six (b , λ_e , μ_z , σ_z^2 , η , and χ) they are determined endogenously.

Individually selected parameters

In accordance with related studies, I set the coefficient of relative risk aversion (γ) at one, reflecting logarithmic period utility from consumption, and the discount factor (β) at 0.9963. The probability of dying (σ) is set at 0.0021, a value which corresponds to an expected length of an individual's working life of forty years.¹⁴ The probability of receiving a job offer while unemployed (λ_u) is normalized to one.¹⁵ For the probability of job destruction after having received notice (δ), I choose the value 1/3. The resulting

¹⁴See for example Bils et al. (2008), Heathcote et al. (2010), and Kambourov and Manovskii (2009) for comparable parameter choices in related analyses.

¹⁵In the present model, this parameter and the mean of the wage offer distribution are jointly not well identified. For this reason I choose to normalize λ_u and to determine μ_z endogenously.

average duration of the notice period of three months is meant to capture a combination of actual notices with informal information on future job loss that workers receive through other channels. The values for parameters set exogenously are summarized in Table 1.3.

Table 1.3: Values of individually selected parameters (1980s)

γ	1
β	0.9963
σ	0.0021
λ_u	1
δ	0.3333

Jointly selected parameters

The remaining six parameters (b , λ_e , μ_z , σ_z^2 , η , and χ) are determined jointly in such a way that the values of eight selected statistics, when computed for the corresponding stationary equilibrium of the model, closely match their empirical counterparts observed for the U.S. in the mid-1980s. The choice of target statistics is informed by the focus of the analysis on worker turnover, job-to-job mobility and wage distributions. For the calculations of empirical target values I use data from the SIPP 1985 whenever possible and indicate exceptions to this rule in the description below. In the following paragraph I introduce briefly those statistics that serve to match observable features of the labor market relevant to the goal of this exercise.

The statistic used with a view to determining the level of unemployment benefits (b) is the unemployment rate (u). The empirical target for this statistic is 6.4%, the average of monthly unemployment rates of men in 1985 as reported by the Bureau of Labor Statistics. In view of the parameters of the wage offer distribution (μ_z and σ_z^2), I use the mean and the variance of initial wages (μ_{zi} and σ_{zi}^2), that is, of wages of workers who move from unemployment to employment. In order to determine the probability of receiving a job offer while being employed (λ_e), I use the rate of job-to-job transitions (τ^{ee}). The empirical target employed is the linear-trend value of job-to-job transitions for 1985, based on a time series of annual job-to-job transition rates for men estimated by

Sherk (2008). The annual value of the target statistic is 11%, corresponding to 0.92% at a monthly level. With a view to determining the probability of receiving notice of imminent job destruction (η), I include the rate of employment-to-unemployment transitions (τ^{eu}) among the targets. Furthermore, I use the mean of positive wage change upon job-to-job transitions ($\Delta \ln w_+^{ee}$) as a target that can be associated with the costs to workers when leaving a job (χ). Finally, two more targets are included in the calibration exercise, namely the rate of unemployment-to-employment transitions (τ^{ue}) and the variance of log wages (σ_w^2). The values for the six parameters determined endogenously are given in Table 1.4, while empirical as well as simulated values of the corresponding target statistics are presented in Table 1.5.

The estimated low value of b is consistent with other empirical studies on wage dispersion.¹⁶ The value of the arrival rate of offers on the job λ_e implies that on average employed workers receive an outside offer every ten months. Moreover, the level of this parameter relative to λ_u is consistent with findings in other empirical studies. The magnitude of job switching costs as implied by the estimated value of χ can be assessed via the reservation wage of workers employed in normal jobs. Given the present estimate of χ , these workers are willing to change jobs for a wage gain of less than one percent, meaning that job switching costs are very low. Finally, the estimate of η is higher than the rate of job-to-unemployment transitions in the data, indicating that a significant fraction of workers find outside job offers during the notice period.

Table 1.4: Values of jointly selected parameters (1980s)

b	0.0306
λ_e	0.0974
χ	0.4553
μ_z	-1.0835
σ_z^2	0.5139
η	0.0168

¹⁶See Hornstein et al. (2011) for a survey of empirical search models and the observation that the value of leisure in search models needs to be very low in order to obtain a plausible level of wage dispersion.

Table 1.5: Simulated vs. empirical statistics (1980s)

	Model	Data
u	0.0636	0.0640
τ^{ue}	0.2481	0.2600
τ^{ee}	0.0092	0.0092
τ^{eu}	0.0148	0.0150
$\Delta \ln w_+^{ee}$	0.3706	0.2700
μ_{zi}	-0.1699	-0.1700
σ_{zi}^2	0.1238	0.2321
σ_w^2	0.1571	0.1907

1.5 Results

1.5.1 The main experiment

The counterfactual experiment discussed in this subsection can be characterized by two features. First, parameter values underlying the experiment are the result of a two-step calibration exercise. Second, the exercise is designed so as to quantify the relationship between job-to-job mobility and wage inequality.

The point of departure for the present experiment is the baseline calibration of the model to observations of the early 1980s, as reported in Section 1.4.3, which is also the first step of the procedure outlined here. In the second step, I recalibrate the model so as to match increased job-to-job mobility in the early 2000s. For this recalibration I keep most of the model parameters fixed at their 1980s values, allowing only three determinants of worker turnover, namely, λ_e , χ , and η , to vary. The early-2000s-values of these three mobility parameters are chosen to match three empirical observations on worker transitions in that period: the rate of job-to-job transitions, the mean wage increase upon job-to-job transition, and the rate of employment-to-unemployment transitions. Finally, I obtain a quantification of the impact of job-to-job mobility on wage inequality by comparing wage distributions in the stationary equilibria corresponding to the two steps of the calibration.

Table 1.6 presents the basic results of the experiment outlined above, that is, the

values of the three mobility parameters together with the associated target statistics. The table documents, first of all, the fact that the model, through changes in determinants of job-to-job mobility, reproduces perfectly the empirically observed increase of 13% in the rate of job-to-job transitions between the two periods. Second, it shows that the other two targets are less well matched, with notable differences in goodness of fit between levels and change on the one hand and between the two statistics on the other. As regards parameter values reported in the table, changes from the first to the second period are seen to differ significantly between the three mobility determinants. The arrival rate of offers on the job increases by 32%, whereas the probability of receiving a notice falls by 26%, and job leaving costs remain virtually unchanged.¹⁷

Table 1.6: Parameter values and statistics

		1980s	2000s	Change (%)
Parameter values:				
	λ_e	0.0974	0.1289	32.34
	χ	0.4553	0.4222	-7.27
	η	0.0168	0.0124	-26.19
Statistics:				
Model	τ^{ee}	0.0092	0.0104	13.04
	$\Delta \ln w_+^{ee}$	0.3706	0.364	-1.78
	τ^{eu}	0.0148	0.0108	-27.03
Data	τ^{ee}	0.0092	0.0104	13.04
	$\Delta \ln w_+^{ee}$	0.27	0.29	7.41
	τ^{eu}	0.015	0.012	-20.00

The results of Table 1.7 put the focus on the main question addressed by the experiment, namely, that about the quantitative effect that the increase in job-to-job mobility in the United States between the two time periods could have had on the residual wage distribution, and, in particular, on residual wage inequality. The first two rows in the

¹⁷The change in χ reported in Table 1.6 is negligible relative to the precision of the numerical algorithm.

table present the mean and the standard deviation of wages, showing that the counterfactual experiment produces a substantial increase of both statistics. Change within the experiment in wage inequality is assessed by three different measures, all of which show a sizeable increase between the first and the second time period. It amounts to around six percent for the mean-min and the ninety-ten ratios and is about half this size for the standard-deviation measure.

Table 1.7: Cross-sectional wages and inequality

	1980s	2000s	Change (%)
Mean(w)	1.092	1.186	8.55
Std(w)	0.529	0.578	9.13
Std($\ln(w)$)	0.396	0.410	3.50
q90/q10(w)	2.764	2.927	5.92
Mean/q05(w)	1.855	1.976	6.50

A comparison of the results of Table 1.7 with empirically measured change in wage inequality indicates how much of that change can be associated with the empirical fact of rising job-to-job mobility. Table 1.8 shows in its first two columns the actual change, calculated from two different data sources, of residual wage inequality over the period studied here for the three measures of the previous table. Inequality change resulting from the counterfactual experiment is reported along with the empirical figures. The fraction of observed inequality increase, as calculated, for example, from CPS data, that is accounted for by the model experiment ranges from one-quarter to slightly over one-half. It is highest for the mean-min ratio (56%), at the middle level for the ninety-ten ratio (37%) and lowest for the standard-deviation measure (28%).¹⁸

¹⁸The quantitative results in this section are based on comparisons of statistics on the cross-sectional distribution of wages in stationary equilibria of the model for different parametrizations. The dynamics of workers moving up the wage ladder through job-to-job transitions does, however, introduce a positive relation between age and earnings in the model. Since the empirical wage data are cleaned of experience (or age) effects in order to obtain residual wages, I perform a robustness check of the results in which I clean also the model data of age effects by regressing them on the same age terms that are used for the empirical data. It turns out that the age-earnings profile in the calibrated model is very flat and that the correction of the simulated data does not significantly affect the quantitative results. For instance, the change in wage inequality explained by the main experiment using model data that are cleaned for age

Table 1.8: Inequality change in the data and the experiment

	Data	Model	
	(CPS)	Experiment	Fraction(CPS)
Std($\ln w$)	12.70	3.50	0.28
q90/q10(w)	15.98	5.92	0.37
Mean/q05(w)	11.65	6.50	0.56

1.5.2 A supplementary experiment

A variation on the design of the main experiment of Section 1.5.2 can sharpen the focus on the role of job-to-job mobility for wage inequality. Instead of assessing the impact of change in three mobility parameters on inequality change, one of these parameters can be singled out for such an assessment. Motivated by the results of the previous section, I choose λ_e for playing the central role in an experiment of this special type. An additional argument for this choice is that informational frictions affecting on-the-job-search are captured by the frequency of job offer arrivals, which has a direct impact on the frequency of job-to-job transitions and affects the wage distribution. With this supplementary experiment I examine the case where all change in job-to-job mobility would arise from change in the rate of on-the-job-offer arrivals.

Table 1.9: Parameter value and statistic

		1980s	2000s	Change (%)
Parameter value:				
	λ_e	0.0974	0.1104	13.35
Statistic:				
Model	τ^{ee}	0.0092	0.0104	13.04
Data	τ^{ee}	0.0092	0.0104	13.04

effects ranges from 26% to 48% instead of from 28% to 56%.

The experimental setup is analogous to that of the previous subsection, now with all parameters except λ_e fixed at their 1980s-values, and only λ_e varying so as to reproduce the job-to-job transition rate in the mid-2000s. Table 1.9 shows that in the present framework the observed increase in job-to-job transitions could have been brought about by a 13% increase in the rate of offer arrivals only. The changes in wage inequality resulting from this experiment are presented in Table 1.10. Inequality, as reflected by the three measures employed here, increases significantly, but consistently less than in the previous experiment. A comparison of these figures with their empirical counterparts in Table 1.11 shows that the potential contribution of increasingly frequent offer arrivals to the rise in observed wage inequality ranges between one-eighth and one-quarter.

Table 1.10: Cross-sectional wages and inequality

	1980s	2000s	Change (%)
Mean(w)	1.0922	1.1036	1.04
Std(w)	0.5293	0.5398	1.98
Std($\ln w$)	0.3964	0.4034	1.77
q90/q10(w)	2.7636	2.8171	1.94
Mean/q05(w)	1.8550	1.9107	3.00

Table 1.11: Inequality change in the data and the experiment

	Data	Model	
	(CPS)	Experiment	Fraction(CPS)
Std($\ln w$)	12.70	1.77	0.14
q90/q10(w)	15.98	1.94	0.12
Mean/q05(w)	11.65	3.00	0.26

1.5.3 Discussion

In order to outline the background to the results reported in Section 1.5.1, Table 1.12 presents detailed information on the endogenous response of workers to parameter change in the main experiment. This information is shown for the two broad groups of unemployed and employed workers, where the latter group is in turn divided into the two subgroups of workers holding normal jobs (as defined in Section 1.3.1) on the one side and of workers who have received notice of job loss on the other. For each group levels of reservation wages, as computed for the two stationary equilibria, are given in the table.¹⁹ In addition, two variables describing job-to-job transitions are presented: the number of transitions relative to the size of a given group and the number of transitions originating in the group as a fraction of the total number of job-to-job transitions.

A first observation suggested by the figures in Table 1.12 is that the effects of simultaneous change in parameters λ_e and η on reservation wages virtually cancel each other out so that, on balance, parameter changes leave the levels of these wages nearly unchanged between the equilibrium of the mid-1980s and that of the mid-2000s. A second observation concerns the rates of job-to-job transitions within each group of employed workers. For both of them a sizeable increase in percentage terms between the two periods can be noted. It is large (30.7%) for the group of those employed workers who anticipate job loss. For the other subgroup the increase is considerably lower (14.0%) but still remarkable. Finally, the weights of the two groups within the total of employed workers differ significantly between the two equilibria. Put in a time framework, the difference represents a weight shift of 1.2 percentage points in favor of the group of workers holding normal jobs.

Based on these observations, the following link between the increase in job-to-job mobility and the rise in wage inequality examined through the main experiment can be established. First, the very small difference in the levels of the reservation wage of unemployed workers implies that the lower end of the support of the wage distribution does not shift noticeably between the two periods. Second, there are potential impacts from at least two different sources related to job-to-job transitions. One is the effect of reservation wages of employed workers which is insignificant in the present exercise due to the very

¹⁹The reservation wages of employed workers with respect to outside offers are exemplified by reporting the values at the mean wage of the 1980s scenario.

Table 1.12: The response of workers to parameter change

	1980s	2000s	Change (%)
Unemployed workers:			
reservation wage z^u	0.5558	0.5533	-0.45
Employed workers:			
all			
employment rate	0.9364	0.9506	1.52
rate of job-to-job transitions	0.0092	0.0104	13.04
in normal jobs			
share in total employment	0.9546	0.9665	1.25
reservation wage for job change $z^s(1.09, 1)$	1.0959	1.0968	0.08
rate of job-to-job transitions	0.0086	0.0098	13.95
share of job-to-job transitions in total	0.8923	0.9107	2.06
having received notice			
share in total employment	0.0454	0.0335	-26.21
reservation wage for job change $z^s(1.09, 2)$	0.5873	0.5866	-0.12
rate of job-to-job transitions	0.0218	0.0285	30.73
share of job-to-job transitions in total	0.1076	0.09180	-14.66

small change recorded for both worker groups. The other impact arises from the frequency of job-to-job transitions, which is significantly higher for both groups of employed workers in the second period. While both groups exhibit a mobility increase, there is a difference between them regarding the nature of job-to-job transitions. Transitions originating in the first group are invariably associated with wage gain whereas those from the second group may involve wage loss. In order to complete the argument underpinning a positive impact of mobility increase on wage inequality, the shift in favor of the first group of workers, documented in Table 1.12, can be invoked. This shift leads to a substantial increase in the excess of the number of transitions from the first group over that of the second group.

The implication is that a higher share of transitions involves moving up the wage ladder. Consequently, the mass of the stationary wage distribution shifts upwards between the first and the second period. Since, at the same time, the lower end of the distribution does not change its position significantly, the overall spread of wages is likely to be larger in the second period.

Table 1.13: Impact of individual parameter change on wage inequality

	Parameter	Std($\ln w$)
λ_e	32.3%	4.57%
χ	(-7.3%)	(0%)
η	-26.2%	-1.97%

In the remainder of this subsection, I discuss individual parameter change and its effect on wage inequality. To begin with, mobility costs do not change between periods in the main experiment, and therefore have no impact on inequality change. Regarding the other two free parameters, Table 1.13 provides evidence on the differential effect of change in the rate of offer arrivals on the one hand and in the probability of receiving notice of imminent job loss on the other. In the experiment, both parameters change substantially between the two periods, however, in opposite directions. Each individual change has clearly visible effects on levels of reservation wages and on residual wage inequality. The large increase in the arrival rate is associated with a significant decline in the reservation wage of unemployed workers on the one hand and in the reservation wages for job change of employed workers anticipating job loss on the other. This in turn leads to a large increase in all three inequality measures examined here. By contrast, the decline in the probability of receiving notice leads to an increase in the above reservation wages, and to a decrease in inequality measures. From a comparison of Table 1.13 with Table 1.7, the two parameters appear as counteracting forces whose combined impact produces the amount of inequality change shown in the latter table.

A synoptic discussion of the results reported in previous tables can shed some additional light on the role of individual parameters as determinants of wage inequality. Parameter λ_e , the arrival rate of offers on the job, is the protagonist in the relation be-

tween job-to-job mobility and wage inequality. When this parameter acts alone to increase job-to-job mobility, as is the case in the experiment of subsection 1.5.2, its increase is relatively modest (14%), and so is the consequent inequality increase as measured by the standard deviation of log wages (1.8%). The situation is different, however, in the main experiment where all three mobility parameters are at work simultaneously. Now, the change of λ_e is much larger (32%) and would, on its own, have increased wage inequality by a considerably higher amount (4.6%). The amplified role of the arrival rate in the main experiment arises from simultaneous change in the conditions of job destruction. The key parameter in this context, the probability of receiving notice (η), decreases by 26%. In isolation, this would have reduced wage inequality by a considerable amount (-2%).

Finally, first steps have been carried out towards a comprehensive assessment of parameter change between the two periods by recalibrating the model to data of the mid-2000s in analogy to the benchmark calibration. Results from the current stage of the exercise produce qualitatively identical changes in the three mobility parameters: a large increase in λ_e , virtually no change in χ , and a sizeable decline of η . The parameters of the distribution of log wage offers change significantly too: while the mean μ_z decreases, the variance σ_z^2 increases. In this connection, it is important to note that changes in the distribution of wage offers have an impact on the rate of job-to-job transitions observed in the economy, as they affect the probability for an employed worker, given his current wage, of receiving an attractive outside offer. Along these lines, Michelacci and Pijoan-Mas (2012), in their analysis of the impact of wage inequality on the number of hours worked, propose the following extended mechanism: An increase in the dispersion of wage offers, by raising the value of climbing up the job ladder, induces a worker to supply more hours in order to accumulate human capital faster. This, in turn, will raise the probability of his obtaining a job when competing against other applicants, and may increase job-to-job mobility in the aggregate. This outline of a mechanism and the comprehensive assessment in progress indicate the need to improve on determining the relative impact of the offer distribution on both job-to-job mobility and wage inequality in the present analysis.²⁰

²⁰Preliminary analyses of an extension of the present framework to include wage shocks within a job suggest that the results presented here are robust to such an extension. For instance, when incorporating a random walk process for log wages within a job with an exogenously set error variance of the same order of magnitude as estimated by Low et al. (2010) for residual wages from the 1993 panel of the SIPP, and repeating the baseline calibration and main experiment for the modified model, parameter changes in the

1.6 Concluding remarks

In this paper I examine the quantitative relationship between wage inequality and job-to-job mobility. A link between the two phenomena is suggested by basic features of job search models. Moreover, a quantitative analysis of this link is motivated by the empirical observation that in the U.S. labor market both residual wage inequality and the rate of job-to-job transitions increased between the late 1970s and the mid-2000s.

I address the question of how much of the observed inequality increase can be associated with increased job-to-job mobility. For this purpose, I develop a search model with on-the-job search, anticipated job destruction, and costs to workers when leaving a job. This framework not only features various determinants of job-to-job mobility, but also allows for job-to-job transitions involving wage loss when workers search for new jobs during a notice period. The main result of the paper is obtained through a computational experiment which involves calibration of the model to match labor market characteristics of the mid-1980s, and recalibration to match changes in labor turnover between this period and the mid-2000s. I find that within this framework, changes in job-to-job mobility do have a significant impact on wage inequality. In particular, up to one-half of the observed increase in residual wage inequality between the first and the second time period is accounted for by the composite effect of changes in determinants of job-to-job mobility.

There are at least three steps which future research along the lines of this analysis could take. The first step concerns a more refined treatment of the process of anticipated job destruction. It amounts to modifying the calibration so as to determine also the length of the notice period (δ) endogenously. The second step is to enhance the role of labor market risk in the model. This can be achieved, for example, by including, in addition to employment risk as incorporated in the feature of anticipated job destruction, wage risk in the form of shocks to wages within a job. Finally, it would be of interest to extend the present model to an equilibrium search framework.

experiment are quantitatively similar to those presented here. The increase in wage inequality produced by the experiment is even stronger than in the model without shocks, ranging from 47% to 72% of the empirically observed increase in different inequality measures.

A Data Appendix

The Survey of Income and Program Participation

Survey design. 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.²¹

Over the length of one panel households are interviewed every four months. 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 that an individual has held over those four months (referred to as the "wave") are recorded.

The 1985 panel. The SIPP 1985 panel data set covers the period from October 1984 to July 1987. It contains information on 34,346 individuals aged 15 years and over from the civilian non-institutional population. After the completion of data collection sampling weights for individual observations are constructed in order to correct for sample attrition, leaving a full panel sample of 18,475 individuals.

In addition to the full panel file, which contains most variables and the longitudinal sampling weights, I extract information on the weekly labor market status, the starting and ending dates for particular jobs, and the monthly cross-sectional sampling weights from the core wave files. The sample is restricted to male workers between the age of 20 and 65 years who are employed at least in one month during the panel span in a job that is neither self-employment nor family work without pay. Moreover, I drop individuals who are in the Armed Forces at some point within the panel span.

²¹<http://www.nber.org/data/sipp.html>

For categorizing individuals as employed, unemployed, or not-in-the-labor-force in a particular month, I use the labor force status recorded for the second week of the month in the core wave files. In particular, a person is classified as employed if he had a job or business in that week and was not looking for work or on layoff. He is classified as unemployed if he had no job or business in that week and was looking for work or on layoff, or if he had a job or business, but was absent without pay and looking for work or on layoff.

The SIPP survey collects information on up to two wage and salary jobs per person for a given wave. In the core wave files job-specific information is recorded in the following way: A person's most important job is recorded as job one throughout the wave. If an individual held more than one job within the wave, the second most important is recorded as job two, even if the two jobs were never held simultaneously. The full panel file contains an edited version of this job-specific information, where two jobs only appear in the same month if they were both held during that month. In my analysis, I use job one as reported in the full panel file as the main wage job held by an individual in a given month if it is neither unpaid family work nor a job in the Armed Forces. Only if job one does not fulfil these criteria, job two is used instead. In order to extract starting and ending dates, jobs are then matched with records from the core wave files. Based on these dates, I record a job-to-job transition to have taken place every time a worker changes his main job from one month to the next under the condition that this job change was associated with a gap of less than 15 days between ending the old and starting the new job, and with a maximum of one month of overlapping job dates.

Concerning the calculation of hourly wages, the SIPP reports crucial variables only once per wave. For around one-half of the observations in my sample a regular hourly wage rate over the whole wave is reported. For the rest, I calculate the 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, I use the annual CPI from the Bureau of Labor Statistics to express real hourly wages in constant 1980 dollars. In accordance with related empirical studies, I exclude wage observations that fall below one-half of the nominal

minimum wage rate.²²

The sample is further restricted to full-time jobs, that is, to jobs for which individuals report to be usually working 35 hours or more per week. Moreover, observations from the first and last three panel months are dropped, because the sample size for these months is much smaller due to the rotating nature of interviewing in the SIPP. This leaves a sample of 4,986 individuals and 97,121 observations for the analysis.

For the estimation of residual wages, a number of variables reflecting worker characteristics need to be constructed from the data set. Following Eckstein and Nagypal (2004), I assign individuals to one of five education groups based on the number of completed years of schooling. The five categories correspond to high school dropouts, high school graduates, workers with some college education, college graduates, and post-graduate degree 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), I impute the number of years of education completed on the basis of the person's most frequently reported level out of ten finer education categories. Based on these finer education categories, I calculate a person's potential labor market experience as age minus years of education minus five. Furthermore, I 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.

A pooled regression of log real hourly wages on the five broad education groups, a quadratic in potential experience, interaction terms between education groups and the quadratic in experience, the non-white dummy, the region dummies, as well as year dummies is run in order to obtain residual wages. Starting wages of workers are estimated from the residual wages of those individuals who report to have been unemployed as defined above in the month prior to the wage observation.

The 2004 panel. 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. After applying sample selection criteria analogous to those for the 1985 panel, I obtain a data set with 5,938 individuals and 159,059 observations for use in the analysis. The construction of

²²See for example Katz and Autor (1999) for similar sample restrictions.

variables is in large part similar to that for the 1985 data set. Accordingly, the following paragraphs outline only differences between the two periods arising either from changes in the questionnaire or from changes in the format of the public use files.

Similar to the previous analysis, I use the labor force status reported for the second week of the month in order to classify individuals as employed or unemployed. Due to a slight change in the questionnaire, a person is now taken to be employed if he had a job or business in that week and was either working or absent without pay but not on layoff. He is classified as unemployed if he either had a job or business, but was absent without pay on layoff, or had no job or business and was looking for work or on layoff.

Since for the 2004 panel only core wave files are available, starting and ending dates for jobs need to be used to determine whether a particular job recorded for a wave was actually held during a particular month. In determining the main job held by an individual in a given month, priority is given to the job recorded as number one in the core file, as well as to jobs held during the second week of the month.

Regarding the education of individuals, the 2004 panel provides information on the highest grade or degree obtained instead of the number of years completed. In order to construct education groups corresponding to the ones for the 1985 data, I use the same classification conversion as Eckstein and Nagypal (2004).

Chapter 2

The effect of moral hazard on wage inequality with on-the-job search and employer competition

(with Árpád Ábrahám and Fernando Álvarez-Parra)

2.1 Introduction

To understand the fundamental determinants of earnings inequality is a classic problem in economics. It is well known that a large part of earnings differences cannot be explained by individuals' observable characteristics. Recent contributions to the literature have evaluated carefully how much of residual wage inequality can be attributed to search frictions in the labor market.¹ In this paper, we add to the analysis another reason for why people with the same characteristics receive different wages, namely, incentive pay. If a worker's effort on the job is private information, wage payments need to vary across different levels of match output in order to provide incentives. Combining on-the-job search and employer competition, on the one hand, with a moral hazard problem in the worker-firm relation, on the other, is of interest for two reasons. First, including both mechanisms in a search model allows for a rich structure of wage inequality within and between firms. Second, and more importantly, it can be shown that the two mechanisms are closely linked. Consequently, it is not obvious that the introduction of moral hazard does indeed increase wage dispersion compared to standard models of on-the-job search and employer competition.² In addition, the interaction between the two mechanisms may

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

²See for example Postel-Vinay and Robin (2002).

affect the estimation of such models as well as the policy conclusions drawn from analyses based on such frameworks.³

As regards the empirical background, the significance of moral hazard for employment relations is indicated by evidence on performance-dependent pay. For example, MacLeod and Malcomson (1998) report for the year 1990 that 24% of young workers in the United States and 34% of workers in the United Kingdom received some form of performance-related compensation. Likewise, job-to-job mobility is documented to be a quantitatively important component of worker turnover. For instance, Fallick and Fleischman (2004) estimate that, for the period from 1994 to 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. Finally, with regard to firm competition, it seems plausible to assume that in case a worker announces his move to another job, his current employer will try to retain him by countering the outside offer.

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 with employer competition. In our model, risk-neutral firms offer long-term contracts to risk-averse, ex-ante identical workers in a labor market with search frictions. Output of a worker-firm match is a function of firm-specific productivity and a stochastic productivity component that depends on the worker's unobservable effort. The informational friction, internal to the firm, introduces into the worker-firm relationship the well-known trade-off between providing insurance and instilling incentives. The optimal balance between these two objectives implies an efficient level of wage variation across output realizations. In addition, our model features one-sided limited commitment, as firms are assumed to commit to the wage contract, whereas workers may quit to unemployment or leave for another job. With respect to on-the-job search and firm competition for workers, we assume that workers randomly receive alternative job offers, and that it is private information whether or not a worker has received an offer. As soon as an offer is disclosed, the current and the potential future employer enter a Bertrand competition for the worker where offers are made in terms of wage contracts.

In our framework, wage differences between ex-ante identical workers originate from

³In particular, the analysis of minimum wage policies as in Lise et al. (2012) may be affected significantly.

search frictions and the informational friction specific to the moral hazard problem. On the one hand, search frictions lead to differences between workers with different histories of unemployment spells and job offers. For instance, wages differ between employed workers who have received outside job offers and those who have not. Furthermore, variation in productivity levels between firms leads to wage differences between offers. On the other hand, labor contracts prescribe, as an optimal response to the informational friction of unobservable effort, wages that are dispersed across output realizations. This mechanism of providing incentives reduces risk sharing, making a worker's history of productivity shocks relevant for his wage sequence. Accordingly, a direct effect of moral hazard is that of inducing wage differences among workers with different histories of productivity realizations.

The latter argument suggests that the presence of moral hazard in labor contracts unambiguously increases residual wage inequality. This would indeed be the case if the only effect of moral hazard were a mean preserving spread of wages around the levels determined by a model with observable effort. However, in the presence of on-the-job search and employer competition, moral hazard affects the wage distribution also indirectly through a number of channels. Such effects can be traced back to a common source, namely, the fact that incentive provision through utility variation makes worker effort more costly to firms. In other words, providing the same lifetime utility and implementing the same effort level as under effort-observability reduces the profits of a firm. The additional effort costs affect the levels of wage offers to unemployed workers, and to employed workers in the course of firm competition. Moreover, they impact on the dynamics of wages within a job.

The main goal of the present paper is to disentangle and characterize the direct and indirect effects of moral hazard on the wage distribution, and to assess them quantitatively. For the latter purpose, we calibrate our model to key moments of cross-sectional wage dispersion and of individual wage dynamics in the U.S. labor market. In the following paragraphs, we describe the indirect effects of moral hazard and outline mechanisms underlying their impact.

One of the channels through which increased effort costs affect wage dispersion is linked to the maximum levels of lifetime utility firms are willing to offer. The reduction in profits under moral hazard implies that these critical utility levels decrease. At the same

time, the lower bound on lifetime utility, which is given by the value of unemployment, remains unchanged. In consequence, the distribution of lifetime utilities in the economy is compressed downwards from the top, leading to a compression of the wage distribution too. One should expect this effect to be quantitatively moderate for the following reason: Using a simplified model, it can be shown that, if effort levels do not change between the two scenarios, average costs to the firm and, hence, average wages have to be equal at the firm's respective break-even points. Therefore, in such a model, the average wage associated with critical utilities remains the same.

The introduction of moral hazard has comparatively much stronger indirect effects on the lower parts than on the rest of the wage distribution. In this connection, a key observation is that in both scenarios firms have to deliver the same level of lifetime utility to workers at the lower bound of the wage distribution, namely, the value of the outside option of unemployment. Under moral hazard, this utility level has to be provided through an output-dependent wage path. *Ceteris paribus*, average wages of recently unemployed workers therefore need to be higher in order to compensate risk-averse workers for earnings variation. Moreover, due to the mechanism described above, workers expect lower levels of lifetime utility from outside offers. Consequently, their continuation utility decreases, leading to an additional increase in average wages with the first employer after an unemployment spell. The resulting compression effect of moral hazard on the wage distribution emerges clearly in the aforementioned simplified model when effort levels are held constant.

Firms react, however, to the increase in effort costs by lowering the levels of effort prescribed to their workers. If these changes were near proportional at all levels of lifetime utility, they would not impact significantly on wage dispersion because the cost of effort compensation would change uniformly. But the change in effort turns out to be much larger at low levels of lifetime utility, due to the fact that incentives need to be provided through a spread in future lifetime utilities. At low levels of utility (and hence wage), this spread is largely constrained by the requirement that, for any history, workers need to be at least weakly better off working than being unemployed. The large decrease in effort levels in this range of utility implies a parallel decrease in effort compensation to the worker and, *ceteris paribus*, a significant decrease in the lowest percentile wages. As a

result, the indirect effect mediated through changes in effort levels counteracts the above described upward compression of the wage distribution.

Since, through the various channels, moral hazard affects the wage distribution in ways that imply inequality changes in both directions, its overall impact on wage dispersion can be determined only by a quantitative analysis. Using our calibrated model for this assessment, we find that, on balance, the presence of moral hazard in labor contracts increases residual wage inequality by around six percent. Computational experiments suggest that the direct effect of incentive provision leads to a moderate inequality increase due to wage dispersion within groups of workers with the same job offer history. Among the indirect effects of moral hazard, the one most closely related to outside wage offers and firm competition counteracts the direct effect, exerting a modest influence on inequality. In stark contrast to this, non-observability of effort, through other indirect effects, has a particularly strong impact on the lower parts of the wage distribution. The main reason is that, in response to higher effort costs, firms demand significantly lower levels of effort from low wage workers. As a result, within the lower half of the wage distribution, inequality increases more than proportionally, thus contributing substantially to the overall impact of moral hazard.

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 (2011a) and Tsuyuhara (2011b) develop a model of directed

on-the-job search with dynamic moral hazard. Within that framework, and similar to our model, both work incentives and job-to-job mobility induce dispersion in labor productivities among ex-ante identical workers. The main focus of the papers is on the longitudinal characteristics of the wage contract and on the response of labor market variables to aggregate shocks. The contribution of moral hazard to observed wage inequality is, however, neither qualitatively nor quantitatively evaluated. Moreover, as opposed to our model, employers are assumed not to react to outside job offers received by their workers.

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. We find that the overall increase in wage inequality attributable to moral hazard amounts to around six percent. Moreover, inequality increase is comparatively more pronounced within the lower parts of the wage distribution.

The paper is organized as follows: Section 2.2 presents the model used in our analysis, states the optimal contract problem, and provides two theoretical results. In Section 2.3 we consider a simplified version of the model and illustrate analytically key parts of the channels through which the presence of moral hazard impacts on the wage distribution. Section 2.4 returns to the quantitative analysis of the full model. The first part describes in some detail the calibration of our model, while the second part presents the quantitative results together with a discussion of the underlying mechanisms. Section 2.5 concludes the paper. Some analytical results and a more extensive data description are presented in the appendices.

2.2 The Model

This section presents the model on which our study is based and states the central contracting problem.⁴ It starts with an exposition of the basic environment in which the

⁴The framework of analysis builds on Alvarez-Parra (2008).

analysis is carried out. Within this framework, we formulate the optimal contract problem and define as well as characterize a number of key variables pertinent to the problem. Finally, and with a view to the quantitative analysis presented in Section 2.4, we state two theoretical results and provide a concise definition of equilibrium in the present context.

2.2.1 The setup

In the present framework, time $t = 0, 1, 2, \dots$ is discrete with an infinite horizon. The economy is populated by a large number of workers I_w and of firms I_f . Workers are risk-averse and ex-ante identical, whereas firms are risk-neutral and differ with respect to their productivity. Each firm can employ only one worker. Both I_w and I_f are 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 at a common discount factor $\beta \in (0, 1)$.

Firms

Each firm is characterized by a productivity level $z \in \mathcal{Z} = \{z_1, z_2, \dots, z_N\}$ where $z_n < z_{n+1}$. The value is drawn from a distribution with cumulative distribution function $F(z)$ when the firm is established and remains constant over time. Firm output y is a function of both the non-stochastic firm-specific productivity level z and a stochastic worker-specific productivity factor A which depends on the effort ϵ spent by the employed worker over a given period.

Firms use a production technology according to which output y is given by

$$y = y(z, A) = zA \tag{2.1}$$

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

$$A = \begin{cases} A^+ & \text{with probability } \pi(\epsilon) \\ A^- & \text{with probability } 1 - \pi(\epsilon) \end{cases} \tag{2.2}$$

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.

Workers

Each worker can be either unemployed or employed by one of the firms active in a given period. For each period, the probability for a worker to die within this period is $(1 - \psi)$. A worker derives utility from consumption and suffers disutility from spending effort while working on his job. The period utility $u(c)$ from consumption c is assumed to be a continuous, strictly increasing and strictly concave function ($u'(c) > 0, u''(c) < 0$) 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 ($g'(\epsilon) > 0, g''(\epsilon) \geq 0$) with $g(\epsilon) \geq 0$ for $\epsilon \in [0, \bar{\epsilon}]$. 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, the unemployed worker's period utility amounts to $u(b)$. In unemployment, the probability of receiving exactly one job offer within a given period is λ_u . The productivity type z of the firm making such an offer is a random draw from distribution $F(z)$. While employed, a worker spends effort ϵ and receives wage w within a time period, implying a period utility of $u(w) - g(\epsilon)$. For the employed worker, the probability of receiving exactly one outside job offer within a given period is λ_e , where, in this case too, the productivity type \tilde{z} of the firm making the offer is drawn from distribution $F(\cdot)$.

Interaction between workers and firms

When a worker and a firm meet, the latter offers the former a long-term contract which is valid as long as the worker does not bring in an outside job offer. The contractual relationship is asymmetric in that the firm commits to the contract, whereas the worker may walk away at any time. Independent of commitment, there is a positive probability δ that within a given period an employment relationship gets destroyed exogenously. The two types of endogenous dissolution of a worker-firm relationship are quitting into unem-

ployment or job-to-job transition. In the event of losing its worker, the firm concerned disappears.

The focus of the present analysis is on one particular feature of the worker-firm interaction, embodied in the assumption that the firm cannot observe the level of effort spent by its worker. Likewise, we assume that an outside offer received on the job cannot be observed by the firm. Finally, we make three additional assumptions about the interaction between firms and workers. First, the long-term contract is a take-it-or-leave-it offer by a firm to a worker. In other words, workers have no bargaining power. Second, if an employed worker reports an outside offer to his employer, the current labor contract is annulled. Subsequently, the current and the potential future employers start competing for the worker by offering new contracts. This competition takes place in the form of a second-price auction between the two firms in terms of expected lifetime utility the respective contracts offer to the worker. And third, a labor contract signed in period t specifies, for all future dates τ and all possible histories of shock realizations until date τ (denoted by $\mathbf{A}^\tau \equiv \{A_j\}_{j=t+1}^\tau$), conditional on the worker staying in the contract, a period wage $w_\tau(\mathbf{A}^\tau)$ and a period effort level $\epsilon_\tau(\mathbf{A}^\tau)$.

Timing of events

The timing of events within a model period is as follows: A worker employed by a z -type firm spends effort ϵ and receives period wage w . Then output $y(z, A)$ is produced, revealing the period realization of A . With probability λ_e , the worker is contacted by a firm of type \tilde{z} drawn from distribution $F(\cdot)$. If he reports the outside job offer to his current employer, the two firms start a Bertrand competition by offering new labor contracts which would start from the beginning of the next period. Within the same time, an unemployed worker consumes b and is, with probability λ_u , contacted by a firm for which he may start working at the beginning of the next period. At the end of the current period, existing worker-firm matches are exogenously destroyed with probability δ . A worker dies with probability $(1 - \psi)$ and is replaced by a newborn worker who is unemployed. A firm that has lost its worker disappears and is replaced by a newly established firm.

2.2.2 The optimal contract

Firms offer to workers long-term contracts that are designed optimally with respect to the non-observability of worker effort and the fact that workers may quit or switch employers. On the one hand, firms offer wages which depend on the history of a worker's output realizations with the firm in order to provide optimal incentives for effort. On the other hand, contracts are designed such that a worker never wants to quit to unemployment and moves only to a competitor firm whose productivity is at least as high as that of the current employer's firm. However, in our framework a firm's future bidding strategy for the case that its worker receives an outside job offer is in general not part of the labor contract.⁵ When designing a contract, a firm therefore 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.

More formally, a labor contract consists of sequences of functions $\{w_\tau(\mathbf{A}^\tau), \epsilon_\tau(\mathbf{A}^\tau)\}_{\tau=1}^\infty$ specifying a period wage and effort level for all future dates τ and all possible histories of productivity realizations \mathbf{A}^τ . Thus, the actions prescribed to the worker and the payoffs to both worker and firm at any point in time depend on the whole history of previous productivity realizations, actions, and payoffs. 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 promised to the worker by the contract.

Preliminaries

In order to be able to describe, in the present setup, an optimal long-term contract between a worker and a firm, we need to expand our notation and introduce a number of additional variables. First, in the analysis different kinds of expected utility for the worker have to be considered. One of them is the expected lifetime utility of a currently unemployed worker, denoted by U^n . Another one, U , is the expected lifetime utility promised to an

⁵See 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.

employed worker under his current labor contract. This utility level, together with the productivity type z of the firm employing the worker, fully characterizes the worker's state. Other expected-utility variables of interest are the continuation values promised to an employed worker, either by his current or by a future employer. Thus, $U^i(U, z)$ denotes the continuation value of the current labor contract at a z -type firm under the conditions that the worker's current realization of specific productivity 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 from a firm of type \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 from the viewpoint of the worker to that of the firm, the value to a z -type firm of a contract delivering lifetime utility U in an optimal way⁶ 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 from a firm of productivity type \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 $V(U, z)$ to the z -type firm of a contract delivering to the worker lifetime utility U can be expressed as

$$\begin{aligned}
 V(U, z) = & z \left[A^+ \pi(\epsilon) + A^- (1 - \pi(\epsilon)) \right] - w \\
 & + \beta \psi (1 - \delta) \left\{ (1 - \lambda_e) \left[V(U^+, z) \pi(\epsilon) + V(U^-, z) (1 - \pi(\epsilon)) \right] \right. \\
 & \left. + \lambda_e \sum_{\tilde{z} \in \mathcal{Z}} \left[V_o(U^+, z, \tilde{z}) \pi(\epsilon) + V_o(U^-, z, \tilde{z}) (1 - \pi(\epsilon)) \right] f(\tilde{z}) \right\} \quad (2.3)
 \end{aligned}$$

⁶The corresponding optimization problem is defined below.

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

$$u(w) - g(\epsilon) + \beta\psi\delta U^n + \beta\psi(1 - \delta) \left\{ (1 - \lambda_e) \left[U^+ \pi(\epsilon) + U^- (1 - \pi(\epsilon)) \right] + \lambda_e \sum_{\tilde{z} \in \mathcal{Z}} \left[U_o(U^+, z, \tilde{z}) \pi(\epsilon) + U_o(U^-, z, \tilde{z}) (1 - \pi(\epsilon)) \right] f(\tilde{z}) \right\} \quad (2.4)$$

Finally, a recursive contract in the present framework can be 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 firm productivity type 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.*

Continuation values for the case of an outside offer

If and 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

$$V(U^*(z), z) = 0 \quad (2.5)$$

plays a decisive role: A firm of productivity type 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.

It should be noted, however, that in the present setup firms may be making losses at some point of the worker-firm match, that is, $U^*(z) < U^i$. In this case, if the worker receives an outside offer from a more productive firm which could still make positive profits at utility level U^i , it is in the interest of the current employer that the worker switches jobs. Moreover, transferring the worker to the competitor firm in this way leads to a Pareto-improvement. Once the outside offer has been disclosed, any bidding strategy of

⁷Since $V(U, z)$ 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 (2.5).

the current employer up to $U^*(\tilde{z})$ exclusive can achieve this objective. However, any value lower than U^i would keep the worker from reporting the offer and hence not lead to a quit. We therefore assume that, in such a situation, the current employer bids U^i , which is one of his optimal strategies.⁸

Taking into account these rules, seven different cases of relationships between values of U^i , $U^*(z)$, and $U^*(\tilde{z})$ can be outlined. The resulting classification is presented below. It includes specifications of the worker's decisions on whether to trigger firm competition, on the one hand, and on whether to stay with his current employer or move to the job offered by the competing firm, on the other. It also states the corresponding continuation values for the worker and the current employer.

The first three cases reflect the situation of the competitor firm being more productive than the incumbent firm, so that $U^*(z) < U^*(\tilde{z})$.

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 ($V_o = 0$).

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

The worker discloses the offer and the incumbent firm offers U^i . The worker moves to the new employer and gets $U_o = U^i$. The incumbent firm disappears ($V_o = 0$).

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

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 next two cases cover the situation where the competitor firm and the incumbent firm have the same productivity level.

Case 4: $U^i \leq U^*(z) = U^*(\tilde{z})$

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 ($V_o = 0$).

⁸Taking into account the worker's optimal reporting strategy, the current employer is in fact indifferent between any bidding strategies in $[U^i, U^*(\tilde{z})]$. The value chosen simply determines how the gains from the offer are split between the worker and the future employer.

Case 5: $U^*(z) = U^*(\tilde{z}) < U^i$

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 $V_o = V(U^i, z)$.

Finally, the last two cases reflect the situation where the competitor firm is less productive than the incumbent firm, so that $U^*(\tilde{z}) < U^*(z)$.

Case 6: $U^i \leq U^*(\tilde{z}) < U^*(z)$

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 7: $U^*(\tilde{z}) < U^*(z)$ 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 $V_o = V(U^i, z)$.

Based on these cases, the continuation value U_o to a worker employed by a z -type firm, who would get utility U^i under his current labor contract and has received an outside offer from a \tilde{z} -type firm, can be summarized in the following expression:

$$U_o(U^i, z, \tilde{z}) = \max \left\{ U^i(U, z), \min [U^*(z), U^*(\tilde{z})] \right\} \quad (2.6)$$

The corresponding continuation value to the current employer is given by

$$V_o(U^i, z, \tilde{z}) = \begin{cases} V(U^i(U, z), z) & \text{if } U^*(\tilde{z}) < U^i(U, z) \\ V(U^*(\tilde{z}), z) & \text{if } U^i(U, z) \leq U^*(\tilde{z}) < U^*(z) \\ 0 & \text{otherwise} \end{cases} \quad (2.7)$$

The optimization problem

The contract offered by a firm to a worker is designed optimally, given the particular features of the current setup. On the one hand, 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,

$$\begin{aligned} \epsilon \in \operatorname{argmax}_{\hat{\epsilon} \in [0, \bar{\epsilon}]} & \quad u(w) - g(\hat{\epsilon}) + \beta\psi\delta U^n \\ & \quad + \beta\psi(1 - \delta) \left\{ (1 - \lambda_e) [U^+ \pi(\hat{\epsilon}) + U^- (1 - \pi(\hat{\epsilon}))] \right. \\ & \quad \left. + \lambda_e \sum_{\tilde{z} \in \mathcal{Z}} [U_o(U^+, z, \tilde{z}) \pi(\hat{\epsilon}) + U_o(U^-, z, \tilde{z}) (1 - \pi(\hat{\epsilon}))] f(\tilde{z}) \right\} \end{aligned} \quad (2.8)$$

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

$$\begin{aligned} g'(\epsilon) = \pi'(\epsilon) \beta\psi(1 - \delta) & \left\{ (1 - \lambda_e) [U^+ - U^-] \right. \\ & \left. + \lambda_e \sum_{\tilde{z} \in \mathcal{Z}} [U_o(U^+, z, \tilde{z}) - U_o(U^-, z, \tilde{z})] f(\tilde{z}) \right\} \end{aligned} \quad (2.9)$$

We use equation (2.9) as the incentive-compatibility constraint (ICC).⁹ 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 (2.9) 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.

On the other hand, in the optimal contract, the problem of lack of commitment of workers has to be taken into account too. In order to prevent workers from quitting to unemployment, firms always promise their workers at least as much utility from the contract as they would get in unemployment.¹⁰ Hence, the participation constraint (PC)

⁹The 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.

¹⁰In 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.

requires that the value of a contract to the worker at any point in time weakly exceeds the value of unemployment U^n .

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):

$$\begin{aligned}
 U \leq & u(w) - g(\epsilon) + \beta\psi\delta U^n \\
 & + \beta\psi(1 - \delta) \left\{ (1 - \lambda_e) \left[U^+ \pi(\epsilon) + U^- (1 - \pi(\epsilon)) \right] \right. \\
 & \left. + \lambda_e \underbrace{\sum_{\tilde{z} \in \mathcal{Z}} \left[U_o(U^+, z, \tilde{z}) \pi(\epsilon) + U_o(U^-, z, \tilde{z}) (1 - \pi(\epsilon)) \right] f(\tilde{z})}_{\text{expected continuation value in case of outside offer}} \right\} \quad (2.10)
 \end{aligned}$$

The possibility of exogenous job destruction, on the one hand, and of outside job offers triggering firm competition, on the other, 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 received an outside offer is subject to only little control by the firm. Moreover, this component of the promise-keeping constraint varies between firms of different productivity types. The reason is that the productivity level of a worker's current employer 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 productivity level, the higher is the worker's expected continuation value in case an outside offer arrives. Consequently, a firm with high productivity faces a more relaxed promise-keeping constraint than a low-productivity firm 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 participation constraint, the term needs to be specified in order to be able to fully describe the contractual problem. Since workers have no bargaining power, a firm offers to an unemployed worker the value of unemployment. Therefore U^n satisfies

$$U^n = u(b) + \beta\psi \left[(1 - \lambda_u) U^n + \lambda_u U^n \right] \quad (2.11)$$

and the value of unemployment can be expressed as

$$U^n = \frac{u(b)}{1 - \beta\psi} \quad (2.12)$$

As the utility accruing to a worker from a labor contract 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 (2.4), in the following way:

$$\begin{aligned} & \underbrace{u(w) - g(\epsilon)}_{\leq 0} + \beta\psi\delta U^n + \beta\psi(1 - \delta) \left\{ (1 - \lambda_e) \left[\underbrace{U^+}_{\leq \bar{U}} \pi(\epsilon) + \underbrace{U^-}_{\leq \bar{U}} (1 - \pi(\epsilon)) \right] \right. \\ & \left. + \lambda_e \sum_{\tilde{z} \in \mathcal{Z}} \left[\underbrace{U_o(U^+, z, \tilde{z})}_{\leq \bar{U}} \pi(\epsilon) + \underbrace{U_o(U^-, z, \tilde{z})}_{\leq \bar{U}} (1 - \pi(\epsilon)) \right] f(\tilde{z}) \right\} \quad (2.13) \end{aligned}$$

Given that the whole expression (2.13) is also bounded from above by \bar{U} , this leads to the equation

$$\bar{U} = \beta\psi\delta U^n + \beta\psi(1 - \delta)\bar{U} \quad (2.14)$$

Substituting (2.12) then yields the expression

$$\bar{U} = \frac{\beta\psi\delta}{1 - \beta\psi(1 - \delta)} \frac{u(b)}{1 - \beta\psi} < 0 \quad (2.15)$$

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} .

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:

$$\begin{aligned}
V(U, z) = & \max_{\{w, \epsilon, U^+, U^-\}} z \left[A^+ \pi(\epsilon) + A^-(1 - \pi(\epsilon)) \right] - w \\
& + \beta \psi (1 - \delta) \left\{ (1 - \lambda_e) \left[V(U^+, z) \pi(\epsilon) + V(U^-, z) (1 - \pi(\epsilon)) \right] \right. \\
& \left. + \lambda_e \sum_{\tilde{z} \in \mathcal{Z}} \left[V_o(U^+, z, \tilde{z}) \pi(\epsilon) + V_o(U^-, z, \tilde{z}) (1 - \pi(\epsilon)) \right] f(\tilde{z}) \right\} \quad (2.16)
\end{aligned}$$

subject to the promise-keeping constraint (PKC)

$$\begin{aligned}
U \leq & u(w) - g(\epsilon) + \beta \psi \delta U^n \\
& + \beta \psi (1 - \delta) \left\{ (1 - \lambda_e) \left[U^+ \pi(\epsilon) + U^- (1 - \pi(\epsilon)) \right] \right. \\
& \left. + \lambda_e \sum_{\tilde{z} \in \mathcal{Z}} \left[U_o(U^+, z, \tilde{z}) \pi(\epsilon) + U_o(U^-, z, \tilde{z}) (1 - \pi(\epsilon)) \right] f(\tilde{z}) \right\} \quad (2.17)
\end{aligned}$$

the incentive-compatibility constraint (ICC)

$$\begin{aligned}
g'(\epsilon) = & \pi'(\epsilon) \beta \psi (1 - \delta) \left\{ (1 - \lambda_e) \left[U^+ - U^- \right] \right. \\
& \left. + \lambda_e \sum_{\tilde{z} \in \mathcal{Z}} \left[U_o(U^+, z, \tilde{z}) - U_o(U^-, z, \tilde{z}) \right] f(\tilde{z}) \right\} \quad (2.18)
\end{aligned}$$

the participation and feasibility constraints (PC) and (FC)

$$\begin{aligned}
U^n & \leq U^+ \leq \bar{U} \\
U^n & \leq U^- \leq \bar{U} \quad (2.19)
\end{aligned}$$

and

$$w \geq 0 \quad (2.20)$$

$$\epsilon \in [0, \bar{\epsilon}] \quad (2.21)$$

2.2.3 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 (2.16) with constraints (2.17) to (2.21), 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 (2.18) must hold. Since $g'(\epsilon) > 0$, the right-hand side of (2.18) 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 $U^+(U, z) > U^-(U, z)$. \square

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 (2.16) with constraints (2.17) and (2.19) to (2.21) – and without the incentive-compatibility constraint (2.18) – 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 B.1. \square

2.2.4 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 (2.16) subject to constraints (2.17) to (2.21).*
- (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 (2.6) and (2.7).*
- (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, $\mu_s(l, U, z) = M(\mu_s(l, U, z))$.*

2.3 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

$$\Pi_o = p[A^+ - w_o^+] + (1 - p)[A^- - w_o^-] = 0 \quad (2.22)$$

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

$$\begin{aligned} pu(w_o^+) + (1-p)u(w_o^-) - v + \beta\psi\left\{\delta U^n + (1-\delta)U_o\right\} &\geq \\ u(w_o^-) + \beta\psi\left\{\delta U^n + (1-\delta)U_o\right\} &\end{aligned} \quad (2.23)$$

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

$$p[u(w_o^+) - u(w_o^-)] = v \quad (2.24)$$

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, (2.24) 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

$$U_o = \frac{pu(w_o^+) + (1-p)u(w_o^-) - v + \beta\psi\delta U^n}{1 - \beta\psi(1 - \delta)} \quad \text{and} \quad \hat{U}_o = \frac{u(\hat{w}_o) - v + \beta\psi\delta U^n}{1 - \beta\psi(1 - \delta)} \quad (2.25)$$

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

$$U^n = pu(w_u^+) + (1-p)u(w_u^-) - v + \beta\psi\left\{\delta U^n + (1-\delta)[\lambda_e U_o + (1-\lambda_e)U^n]\right\} \quad (2.26)$$

which can be rearranged to yield

$$U^n = \frac{pu(w_u^+) + (1-p)u(w_u^-) - v + \beta\psi(1-\delta)\lambda_e U_o}{1 - \beta\psi[\delta + (1-\delta)(1-\lambda_e)]} \quad (2.27)$$

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

$$U^n = \frac{u(\hat{w}_u) - v + \beta\psi(1-\delta)\lambda_e \hat{U}_o}{1 - \beta\psi[\delta + (1-\delta)(1-\lambda_e)]} \quad (2.28)$$

Equations (2.27) and (2.28) 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 $U_o < \hat{U}_o$, (2.27) and (2.28) imply that $pu(w_u^+) + (1-p)u(w_u^-) > u(\hat{w}_u)$. From concavity of $u(\cdot)$ it follows that $pw_u^+ + (1-p)w_u^- > \hat{w}_u$. \square

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 (2.24), 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.

2.4 Quantitative Analysis

Of the two parts of this section, the first provides a detailed description of the calibration strategy, the arguments supporting the selection of empirical targets, and the calibration results. The second part presents the results of our quantitative analysis, together with an extensive discussion of the mechanisms underlying the various effects of moral hazard on the wage distribution and their implications for residual wage inequality.

2.4.1 Calibration

In this section we lay out the calibration of our model to U.S. data. Starting from some basic specifications, we move on to a description of the approach to parameter selection and the choice of targets to match, and, finally, present the parameter values obtained by our procedure. A brief closing section outlines the data background to the calibration, including the regression specification used to estimate residual wages which are central to the conceptual framework of our analysis.

Functional forms and distributions

A worker's period utility from consumption is assumed to be given by a constant relative risk aversion (CRRA) utility function, that is,

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma} \quad (2.29)$$

where $\sigma > 1$ is the coefficient of relative risk aversion. The period disutility from spending effort is given by a power function

$$g(\epsilon) = \epsilon^\gamma \quad (2.30)$$

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

$$\pi(\epsilon) = 1 - \exp\{-\rho\epsilon\} \quad (2.31)$$

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

$$\begin{aligned} A^+ &= 1 + \Delta A \\ A^- &= 1 - \Delta A \end{aligned} \quad (2.32)$$

With regard to firm-specific productivity, we set $\mathcal{Z} = \{z_1, z_2, z_3\}$, with $z_1 < z_2 < z_3$. The distribution over these values from which a newly established firm draws its productivity level is denoted by

$$f(z) = \{f(z_1), f(z_2), 1 - f(z_1) - f(z_2)\} \quad (2.33)$$

Finally, we normalize firm productivity levels in the following way:

$$\begin{aligned} z_1 &= 1 - \Delta_z \\ z_2 &= 1 \\ z_3 &= 1 + 2\Delta_z \end{aligned} \quad (2.34)$$

Strategy and results

Given the above assumptions on functional forms and on the distribution of firm-specific productivity levels, values for the following thirteen 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. γ , the power parameter in the function for disutility from spending effort,
8. ρ , the parameter in the expression for the probability of a high realization of worker productivity as a function of effort,
9. ΔA , the parameter determining the difference between high and low realizations of worker productivity,
10. b , the level of consumption while unemployed,
11. Δ_z , the parameter determining differences in firm-specific productivity levels, and
12. $f(z)$, the probability for a newly established firm to draw productivity level z .

Some of these parameters are selected in accordance with the literature. Others are chosen with the aim of matching empirical statistics on selected features of the U.S. labor market in the mid-2000s by corresponding stationary equilibrium statistics of the model. We obtain empirical statistics for our calibration from the 2004 panel of the Survey of Income and Program Participation (SIPP). The last part of the present section provides some information on the data as well as on measurement and estimation of the variables of interest.

We choose the length of a time period in the model to be one quarter.¹¹ 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 probability of exogenous destruction of a worker-firm match (δ) is chosen such that the rate of employment-to-unemployment flows (τ_{eu}) in the data is matched. Similarly, the probability for an unemployed worker to receive a job offer (λ_u) is selected with a view to matching the empirical job finding rate (τ_{ue}). In order to obtain these parameter values, we use the following two equations which, conditional on ψ , pin down the flows between employment

¹¹This is a compromise between using short time periods which correspond more closely to the high frequency of workers' labor market transitions, on the one hand, and computational feasibility of solving the model numerically, on the other.

and unemployment in stationary equilibrium:

$$\tau^{eu} = \delta + (1 - \delta)(1 - \psi) \quad (2.35)$$

$$\tau^{ue} = \psi\lambda_u \quad (2.36)$$

Accordingly, δ is set at the value of 0.018, whereas λ_u is set at 0.55.

The remaining parameters on the list are selected with the objective of matching empirical statistics on the frequency of job-to-job transitions, individuals' wage changes both within and between jobs, and the cross-sectional distribution of wages. In view of the empirical rate of job-to-job transitions (τ_{ee}), we assign to the probability for an employed worker to receive an outside job offer (λ_e) the value of 0.15. The next three parameters, γ , ρ , and ΔA , are most closely connected with costs and benefits of incentive provision when effort is not observable. As was discussed before, incentive provision under moral hazard leads to output-dependent variation of wages within a job. We therefore select values for the above three parameters with a view to matching empirical statistics on within-job wage changes. The targets used in this connection are statistics on differences between two periods in the log wages of workers who stay with the same employer. In particular, we employ as targets the following statistics on wage change within a job: the mean ($\mu(\Delta \ln w^{win})$), the standard deviation ($\sigma(\Delta \ln w^{win})$), and the fraction of negative log wage changes ($P[(\Delta \ln w^{win}) < 0]$). In this context, a complication arises from the fact that the effects of the power parameters in the function of disutility from effort (γ) and in the function relating the probability of high output realization to the level of effort (ρ) cannot be disentangled. For this reason, we choose to fix γ at a value of 2, that is, to assume quadratic disutility from effort, and to select the other two parameters with the goal of matching statistics on within-job wage changes. As a result, the parameter of the function determining effort-dependence of output realizations (ρ) is set to the value of 3. For the difference between high and low output realizations (ΔA) we select a value of 0.26.

Values for the remaining parameters are chosen in such a way that the model matches empirical statistics of the cross-sectional wage distribution and the distribution of individuals' wage changes upon job-to-job transitions. With a view to matching the average level of wages in the cross-section ($\mu(w)$), we set the level of consumption of unemployed

workers (b) to 0.8. The parameter determining differences between firm-specific productivity levels (Δ_z) as well as the productivity distribution ($f(z)$) are closely related to the overall spread of wages in the model economy. Accordingly, when choosing values for these parameters, we aim to match the standard deviation of log wages ($\sigma(\ln w)$) as a measure of wage inequality in the cross-section. Moreover, the parameters in question have strong effects on the dynamics of wages associated with job switching. Since our model framework is capable of producing only a few job-to-job transitions that are associated with wage loss to the worker, we choose to match statistics of the distribution of positive wage changes between jobs. In particular, we use as targets the mean ($\mu(\Delta \ln w_+^{bet})$) and the standard deviation ($\sigma(\Delta \ln w_+^{bet})$) of positive changes in log wages between jobs. As a result, the difference parameter of firm productivity levels (Δ_z) is set to 0.26, while the probability distribution of productivities ($\{f(z_1), f(z_2), f(z_3)\}$) is chosen as $\{0.4, 0.36, 0.24\}$.

Table 2.1 summarizes the parameter values of our calibration. The simulated model statistics together with the empirical targets are reported in Table 2.2. From the latter one can see that the calibrated economy reproduces accurately the empirical flows of workers between labor market states. Both the scale and the dispersion of cross-sectional wages are matched well, as, in particular, the model produces 86% of the empirically observed residual wage inequality. With respect to statistics on within-job wage changes, the model reproduces accurately the fraction of negative log 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 large relative to the scale of variation of wages due to the moral hazard problem. Finally, the mean and variation in positive wage changes upon job-to-job transitions are matched reasonably well. As was mentioned in the discussion on the choice of targets, the model produces a very low fraction of negative wage changes associated with job-to-job transitions. To sum up, the calibrated economy reproduces workers' transition rates and moments of the cross-sectional distribution of residual wages very well. Given the focus of our analysis, it does reasonably well in matching statistics of wage changes both within and between jobs.¹²

¹²Our numerical analyses suggest that significant improvements in matching statistics on wage changes both within and between jobs can only be achieved by a significant increase in the number of firm pro-

Table 2.1: Values of model parameters

Parameter	Value	Parameter	Value
β	0.99	ρ	3
σ	2	ΔA	0.35
ψ	0.994	b	0.8
δ	0.018	Δ_z	0.26
λ_u	0.55	$f(z_1)$	0.4
λ_e	0.15	$f(z_2)$	0.36
γ	2	$f(z_3)$	0.24

Table 2.2: Simulated vs. empirical statistics

Statistic	Model	Data
τ^{eu}	0.0236	0.0238
τ^{ue}	0.5443	0.5484
τ^{ee}	0.0387	0.0391
$\mu(w)$	1.1662	1.1324
$\sigma(\ln w)$	0.4147	0.4824
$\mu(\Delta \ln w^{win})$	0.0148	0.0060
$\sigma(\Delta \ln w^{win})$	0.0851	0.1416
$P[(\Delta \ln w^{win}) < 0]$	0.3511	0.3550
$\mu(\Delta \ln w_+^{bet})$	0.3159	0.2784
$\sigma(\Delta \ln w_+^{bet})$	0.2963	0.2580

Data background

We obtain empirical statistics for use in the calibration from the 2004 panel of the Survey of Income and Program Participation (SIPP). The data set contains information at high

ductivity levels. A modification of the quantitative model in which we assume a particular functional form for the distribution of firm productivities and calibrate the distributional parameters to statistics on between-job wage changes and cross-sectional wage dispersion is work in progress.

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. Moreover, the structure of these data 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, a quadratic in potential experience, interaction terms between education groups and experience, 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 B.2 provides additional information on the data underlying our calibration.

2.4.2 Computational experiments

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 char-

acterize the various channels through which this impact is brought to bear on the wage distribution. To this effect, we analyze different versions of a quantitative economy that are based on the model framework outlined in Section 2.2 and parametrized according to the calibration procedure described in Section 2.4.1.

The main result

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? This answer is derived from a comparison between two scenarios which differ only with respect to the assumptions made about observability of worker effort. The first scenario is identical with that described in Section 2.2 where, in particular, worker effort is not observable and, in order to extract positive effort from workers, incentives are provided through variation in future utility. We label this the moral hazard (MH) scenario. The contractual problem in this case takes the form of the optimization problem (2.16) subject to constraints (2.17) to (2.21).

For the second 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 (2.16) subject to constraints (2.17), (2.19), (2.20), and (2.21). As stated in Lemma 2 of Section 2.2.3, in this environment a firm fully insures its worker against effort-dependent productivity shocks, that is, $U' \equiv U^- = U^+$.

Table 2.3 presents a comparison of the two stationary equilibria belonging to the above scenarios in terms of various measures of wage inequality. The difference – 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 statistics shown in Table 2.3 indicate that the overall consequence of introducing a moral hazard problem into the labor contract is an increase in residual wage inequality. The size of inequality increase lies between 3.5% and 6.9% for comprehensive measures such as the standard deviation of log wages, the coefficient of variation, and the Gini coefficient. However, a closer look at changes in wage percentiles reveals that the impact of moral hazard is not even across different parts of

the wage distribution. The percentile ratios presented in Table 2.3 show that inequality between the top and bottom five percent of the wage distribution increases by nearly 15%. Moreover, the larger part of this increase comes from a rise in inequality within the lower half of the distribution, as measured by the 50th-to-5th percentile ratio. Finally, the last row in the table reports a change in the mean-min ratio of nearly 60%. Since average wages in the two scenarios are close to equal, this large increase confirms that moral hazard has a particularly strong impact on the lowest parts of the wage distribution.

Table 2.3: Measures of wage inequality: OE vs. MH

	MH	OE	Difference
Std($\ln w$)	0.389	0.415	6.52 %
CV(w)	0.340	0.352	3.50 %
Gini(w)	0.188	0.201	6.87 %
95/05(w)	2.998	3.443	14.84 %
95/50(w)	1.380	1.466	6.28 %
50/05(w)	2.173	2.348	8.06 %
Mean/min(w)	2.329	3.695	58.64 %

As shown in the analysis of the simplified model, moral hazard shapes the wage distribution through a variety of effects. In the next section, we attempt to disentangle and characterize the different channels of impact in the context of the full model, which will shed light on the mechanisms underlying the present results. As a point of departure, in the following paragraphs we discuss differences in firms' optimal policies between the two scenarios and outline their relevance for differences in the cross-sectional wage distributions.

Figure 2.1 shows policy functions of firms (of the z_3 -type) for continuation utilities U^i . 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 penalized 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 contrast, in the OE scenario the continuation utility for a worker within his current labor contract is non-stochastic ($U_{OE}^+ = U_{OE}^- \equiv U'_{OE}$). Moreover, in the case of a z_3 -type firm, this utility level is constant over time.¹³ The presence of moral hazard, by introducing stochastic dynamics in lifetime utility into labor contracts, thus adds a specific source of wage dispersion to the environment.

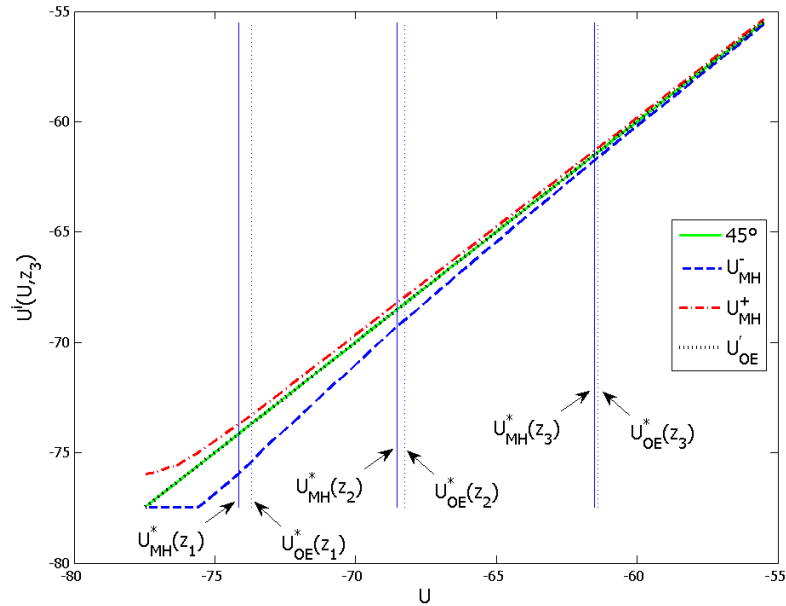


Figure 2.1: Policy functions for continuation utilities, MH vs. OE case, z_3 -firms

In parallel to the simplified model, this direct effect of moral hazard is complemented by a number of additional, indirect effects. For all of these, the cost to a firm of extracting the required level of effort plays a central role. 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. These differences in effort costs translate into differences between the wage distributions of the two scenarios through various channels.

¹³For z_1 -type and z_2 -type firms this is also true for $U < U^*(z)$. Above this value, where the firm is making losses, the potential arrival of outside offers from more productive firms introduces an externality due to which U and U' are not necessarily equal. For reasons of clarity of presentation, we use policy functions of z_3 -type firms for whom this externality is not present.

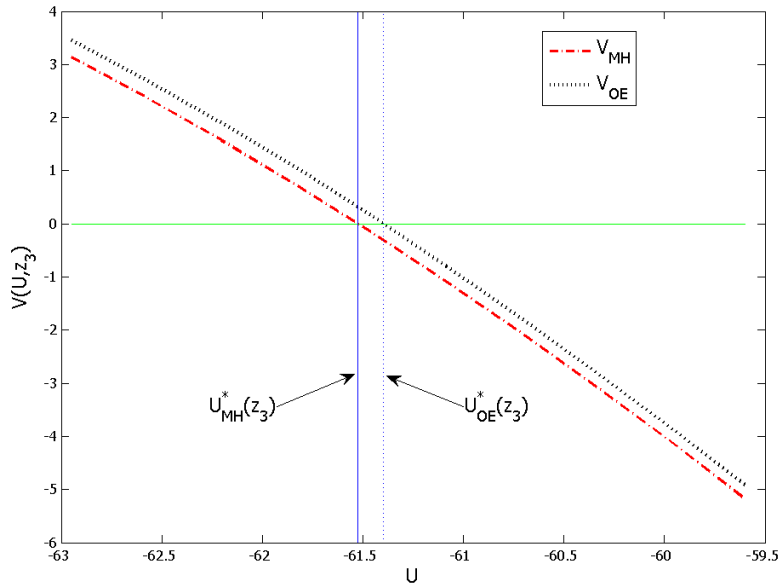


Figure 2.2: Value functions and critical utility levels, MH vs. OE case, z_3 -firms

One of the component 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 firm types the value function of the MH scenario is below that of the OE scenario, which implies that critical utility levels $U^*(z)$ are lower too. Figure 2.2 illustrates the differences between the value functions and critical utilities of z_3 -type firms corresponding to the two scenarios. Figure 2.1 depicts the differences in critical utility levels between the MH and the OE scenarios for all firm types, showing that the size of this difference is inversely related to the level of firm productivity. As regards 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 translates into lower wage gains that workers can achieve through outside offers, and therefore into lower maximum wage levels attained 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.

In contrast to the simplified model, in the present framework we allow for continuous choice of worker effort. This implies that, in the full model, another component effect of increased effort costs is associated with the levels of effort that firms seek to extract from

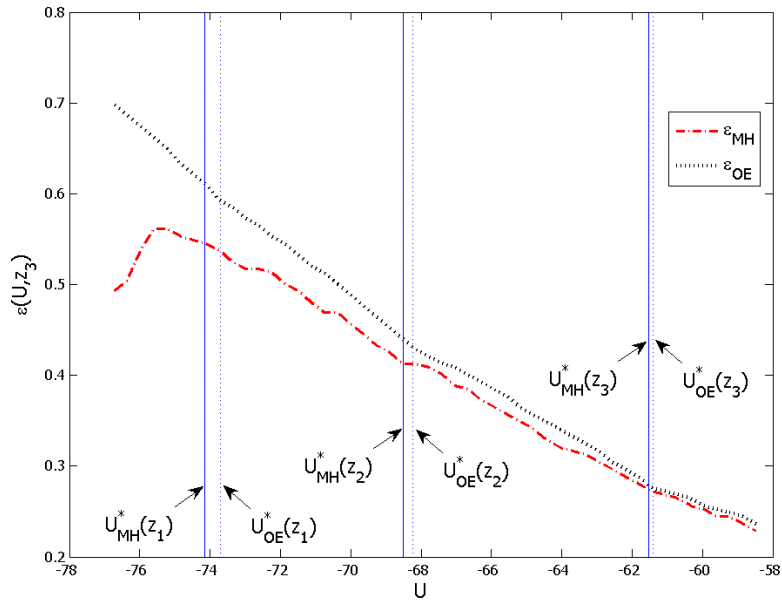


Figure 2.3: Policy functions for effort, MH vs. OE case, z_3 -firms

their workers. Figure 2.3 shows – again for the case of z_3 -type firms – the policy functions for effort in the MH and the OE case. As expected, effort levels are generally lower under moral hazard, where by and large differences between the two scenarios become smaller with increasing levels of lifetime utility. In addition, the figure exhibits a 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.

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. This additional constraint leads to a further rise of the cost of effort to firms. As lifetime utility increases, the cost of the participation constraint declines so that initially effort levels increase in utility. The range of lifetime utility over which the participation constraint is binding can be clearly seen in Figure 2.1. This region, where the function U_{MH}^- is flat, coincides with the interval over which the effort function is increasing.

Leaving everything else 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. Moreover, under moral hazard effort levels have a direct impact on the dynamics of lifetime utility of workers. As the level of effort of a worker determines the respective probabilities of an increase or decrease in utility in the next period, it affects the stochastic process of a worker's wage change within a job over time.

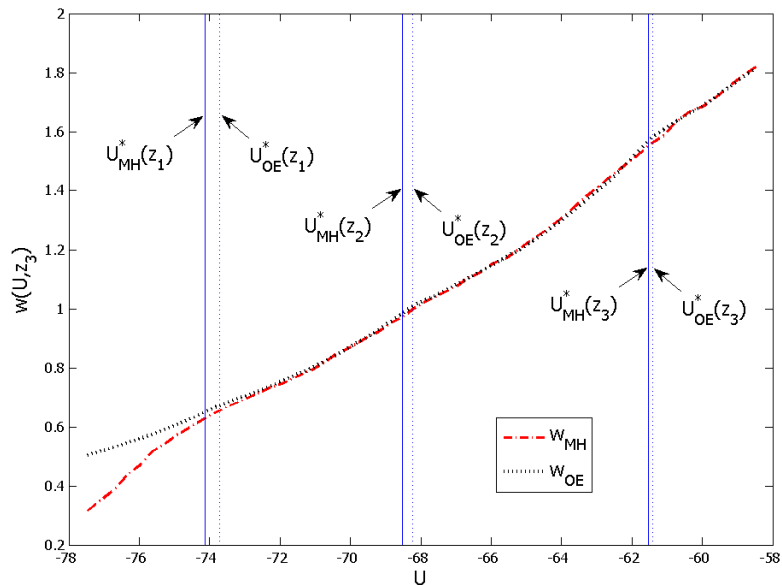


Figure 2.4: Policy functions for wage, MH vs. OE case, z_3 -firms

Finally, changes between the two scenarios in firms' policy functions for wages are related to all of the indirect effects of incentive provision. Most immediately, at a given level of lifetime utility, wages increase in order to compensate workers for the additional variation in earnings. At low values of lifetime utility, wages also increase in order to compensate workers for lower continuation values associated with outside offers. They decrease, however, as a consequence of lower effort compensation costs due to lower effort levels prescribed to workers. Finally, wages also adjust to changes in the stochastic process of a worker's lifetime utility arising from lower effort levels. Figure 2.4 presents, again for the case of z_3 -type firms, the policy functions for wages in the two scenarios. It clearly shows how the strong decrease in effort levels at low values of lifetime utility translates into significantly lower starting wages under moral hazard than under observable effort. The figure also illustrates that the various forces on firms' wage policies more or less cancel

out in the case of z_3 -type firms at higher values of lifetime utility.

Decomposition of impact and exploration of mechanisms

This section presents a series of computational experiments that aim at decomposing the overall impact of moral hazard on the wage distribution into different channels. For this purpose, we start from the observable effort scenario and add, one-by-one, the constraints as well as the components of the solution to the firms' optimal contract design problem under moral hazard. Apart from clarifying the mechanisms underlying our main result, these experiments provide an assessment of the direction of impact and the relative quantitative importance of the different effects of moral hazard on wage inequality.

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 $U_{CF}^* = U_{OE}^*(z)$, and solve (ICC) and (PKC) for the counterfactual policy functions $U_{CF}^+(U, z)$ and $U_{CF}^-(U, z)$ subject to (PC).

Table 2.4 reports changes in statistics on the wage distribution that are associated with a transition from the OE scenario to the present counterfactual scenario (EXP1). Given the setup of the experiment, these changes obviously contain the direct effect of moral hazard which consists in incentive-providing wage variation within a job. The changes in wage percentiles presented in the table show that, both at the top and the bottom of the distribution wage differences within groups of workers with the same job offer history do indeed expand as expected. At the same time, changes in the lowest wage percentiles reveal an upward compression of wages from the lower end towards the middle of the distribution. This compression arises from two forces acting at low levels of lifetime utility. On the one hand, in analogy to the simplified model, risk-averse workers need to be compensated for

wage variation by higher expected future wages. On the other hand, 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.

In the present experiment the expansion of within-group wage differences, attributable to the direct effect of moral hazard, is dominated by the upward compression arising from indirect effects. As a result, overall wage inequality decreases by nearly ten percent.

Table 2.4: Decomposition step 1: Incentive-compatibility

	OE	EXP1	Difference
Std($\ln w$)	0.389	0.351	-9.94 %
Min(w)	0.504	0.504	0.00 %
P05(w)	0.528	0.600	13.80 %
P10(w)	0.597	0.687	15.03 %
P50(w)	1.147	1.193	4.05 %
P90(w)	1.582	1.663	5.11 %
P95(w)	1.582	1.717	8.56 %
P99(w)	1.582	1.872	18.32 %

Step 2: Critical utility levels

The setting of the second experiment (EXP2) is similar to that of the first 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 $w_{CF} = w_{OE}(U, z)$, $\epsilon_{CF} = \epsilon_{OE}(U, z)$, $U_{CF}^* = U_{MH}^*(z)$, and obtain the counterfactual policy functions $U_{CF}^+(U, z)$ and $U_{CF}^-(U, z)$ by solving (PKC) and (ICC) subject to (PC).

The changes in wage percentiles reported in Table 2.5 show that the reduction in critical utility levels affects both the bottom and the top 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 an 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 from the top of the distribution. Both indirect effects at work in the present experiment thus counteract the direct effect of moral hazard. However, the overall reduction in wage inequality associated with the decrease in critical utility levels is modest.

Table 2.5: Decomposition step 2: Critical utility levels

	EXP1	EXP2	Difference
Std($\ln w$)	0.351	0.349	-0.51 %
Min(w)	0.504	0.504	0.00 %
P05(w)	0.600	0.601	0.17 %
P10(w)	0.687	0.688	0.09 %
P50(w)	1.193	1.188	-0.44 %
P90(w)	1.663	1.655	-0.45 %
P95(w)	1.717	1.712	-0.32 %
P99(w)	1.872	1.869	-0.18 %

Step 3: Effort levels

In the third experiment (EXP3), the change in effort levels between the OE and the MH scenarios is added. Since we still keep firms' policy functions for wages fixed at the observable effort solution, this counterfactual scenario 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 (PC).

As can be seen from the changes in wage percentiles reported in Table 2.6, the indirect effect working through changes in lifetime utility dynamics caused by lower effort levels also affects both the top and the bottom of the wage distribution. On the one hand, lower

probabilities of wage raises at high values of lifetime utility lead to a slight downward compression of wages from the top. On the other hand, as illustrated in Figure 2.3, the reduction in effort levels in response to moral hazard is much larger at low values of lifetime utility. Therefore, workers who have not yet received an outside job offer, on average, stay much longer at low wage levels than in the previous scenario. These changes in wage dynamics at low values of lifetime utility 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 nearly seven percent.

Table 2.6: Decomposition step 3: Effort levels

	EXP2	EXP3	Difference
Std($\ln w$)	0.349	0.372	6.77 %
Min(w)	0.504	0.504	0.00 %
P05(w)	0.601	0.591	-1.68 %
P10(w)	0.688	0.604	-12.17 %
P50(w)	1.188	1.165	-1.92 %
P90(w)	1.655	1.647	-0.50 %
P95(w)	1.712	1.692	-1.14 %
P99(w)	1.869	1.790	-4.21 %

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. As discussed in Section 2.4.2, changes in wage policies between the OE and the MH scenarios are associated with a variety of indirect effects of moral hazard. The changes in wage percentiles reported in Table 2.7 indicate that, among these effects, the downward adjustment of wages in response to lower effort compensation costs is dominant. At high levels of lifetime utility, this again leads to a downward compression of wages. However, the large decrease in wage levels at low values of lifetime utility leads to a strong expansion of wages at the bottom of

the distribution. The adjustment of wage levels thus amplifies the inequality-increasing effect of reduced effort levels in the lower parts of the wage distribution. Analogous to the previous experiment, the impact at low levels of lifetime utility is dominant, and overall wage inequality increases by around 11%.

Table 2.7: Decomposition step 4: Wage levels

	EXP3	MH	Difference
Std($\ln w$)	0.372	0.415	11.36 %
Min(w)	0.504	0.316	-37.38 %
P05(w)	0.591	0.489	-17.27 %
P10(w)	0.604	0.591	-2.12 %
P50(w)	1.165	1.149	-1.39 %
P90(w)	1.647	1.647	0.00 %
P95(w)	1.692	1.685	-0.46 %
P99(w)	1.790	1.786	-0.24 %

A brief summary of results

The results of our quantitative analysis can be summarized in a few points. First, on balance, the presence of moral hazard in labor contracts increases residual wage inequality by around six percent. Second, computational experiments intended to disentangle partial effects and to explore the underlying mechanisms suggest that the direct effect of moral hazard, attributable to incentive-providing utility variation, produces within-group wage dispersion which accounts for a moderate contribution to inequality increase.¹⁴ Third, some of the indirect effects working through channels outlined above are found to counteract the direct effect of moral hazard. Fourth, the indirect effect that is most closely linked to outside wage offers and firm competition opposes the direct effect, exerting a modest

¹⁴That the direct effect of moral hazard is quantitatively moderate is first of all suggested by the fact that in the first experiment (EXP1) it is dominated by counteracting indirect effects. Furthermore, in an additional counterfactual experiment not presented here we start from the moral hazard scenario and remove the incentive-compatibility requirement on continuation values while keeping wage and effort policy functions and critical utility levels fixed at the moral hazard solution. The difference in inequality between this counterfactual and the MH scenario is around three percent.

influence on inequality. Fifth, the lack of observability of effort, through other indirect effects, has a particularly strong impact on the lower parts of the wage distribution. It originates mainly from the fact that, under moral hazard, firms demand – in response to higher effort costs – significantly lower levels of effort from low wage workers. This results in a more than proportional increase of inequality within the lower half of the wage distribution, which contributes substantially to the overall impact of moral hazard on residual wage inequality.

2.5 Concluding Remarks

In the present paper we study moral hazard as a source of wage dispersion, using a search model that features job-to-job mobility with firm competition for workers. More specifically, we focus on the particular informational friction – internal to the firm – of worker effort being unobservable to the employer. This friction creates a moral hazard problem within worker-firm relations which eventually affects the wage distribution. Against this background, the specific goal of our analysis is two-fold: First, we quantify the overall impact of moral hazard on residual wage inequality. Second, we identify and quantify different components of the total effect and explore the underlying mechanisms, taking into account potential interactions between search frictions and the particular informational friction of unobservable effort.

An intuitive argument for why non-observability of worker effort should increase wage dispersion starts from the reduction in risk-sharing that moral hazard is known to bring about. In the context of labor contracts, this reduction takes the form of incentive-creating wage variation as the optimal response to the moral hazard problem. The implied output-dependence of wages should in turn lead to higher wage dispersion. In an environment in which wage dynamics arise from both the optimal design of long-term contracts and from job-to-job mobility under employer competition, the direct effect of incentive provision is, however, complemented by a number of indirect effects. These indirect channels of impact arise from an increase in the cost of worker effort to the firm when incentives need to be provided through spreads in future wages. Since some of the channels affect the wage distribution in opposite directions, the overall impact of moral hazard on wage dispersion

needs to be determined by a quantitative analysis.

Based on a calibration of our model to statistics on wage dispersion and individual wage dynamics in the U.S. labor market, we find that, on balance, the presence of moral hazard increases residual wage inequality by around six percent. An exploration of the underlying mechanisms suggests that the direct effect of incentive provision leads to a moderate inequality increase due to wage dispersion within groups of workers with the same job offer history. Among the indirect effects, the one most closely related to outside wage offers and firm competition counteracts the direct effect, exerting a modest influence on inequality. By contrast, other channels of impact produce particularly strong indirect effects on the lower parts of the wage distribution. More specifically, due to a large decrease of effort levels in response to a rise in effort costs, inequality increases more than proportionally within the lower half of the distribution.

The nature of our findings suggests a specific extension of the analysis and a general remark on implications for related policy studies. On the one hand, we analyze in our study only differences across identical workers. Given that the strength of impact of moral hazard depends on the level of worker productivity, which may differ between occupations or sectors, it would be a natural extension to incorporate in the model worker heterogeneity in these dimensions. This modification would allow for a joint analysis of the impact of moral hazard on both within-group and between-group wage inequality and may provide interesting implications for the differences in wage dispersion along these lines. On the other hand, the strong interaction between incentive provision, effort levels, and wages in our framework indicates that incorporating moral hazard in the estimation of equilibrium models of on-the-job search and employer competition could change the results considerably. Related to this, an important next step would be to study how these interactions change the policy implications one can draw from analyses based on such models. In particular, since the effect of moral hazard is particularly strong in the lower parts of the wage distribution, the assessment of the impact of minimum wage and unemployment insurance policies might be affected significantly.

B Appendix

B.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

$$U^*(z_i) < \tilde{U}^- < U^*(z_{i+1}) \quad \text{and} \quad U^*(z_j) < \tilde{U}^+ < U^*(z_{j+1}) \quad (2.37)$$

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

$$\eta^+ = \left(\frac{1 - \pi(\epsilon)}{\pi(\epsilon)} \frac{[1 - \lambda_e + \lambda_e F(z_i)]}{[1 - \lambda_e + \lambda_e F(z_j)]} \right) \eta^- \quad (2.38)$$

The values (η^+, η^-) satisfy conditions (2.37) and (2.38).

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

$$\tilde{V} - V = \beta\psi(1 - \delta) \left\{ [aV(\tilde{U}^+) + bV(\tilde{U}^-)] - [aV(U^+) + bV(U^-)] \right\} \quad (2.39)$$

where

$$a = \pi(\epsilon) [1 - \lambda_e + \lambda_e F(z_j)] > 0 \quad (2.40)$$

$$b = (1 - \pi(\epsilon)) [1 - \lambda_e + \lambda_e F(z_i)] > 0 \quad (2.41)$$

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

$$\alpha V(\tilde{U}^+) + (1 - \alpha)V(\tilde{U}^-) > \alpha V(U^+) + (1 - \alpha)V(U^-) \quad (2.42)$$

By construction, due to requirement (2.38),

$$\alpha \tilde{U}^+ + (1 - \alpha)\tilde{U}^- = \alpha U^+ + (1 - \alpha)U^- \quad (2.43)$$

This can be seen from

$$\alpha \tilde{U}^+ + (1 - \alpha)\tilde{U}^- = \alpha U^+ + (1 - \alpha)U^- - \alpha\eta^+ - (1 - \alpha)\eta^- \quad (2.44)$$

and

$$-\alpha\eta^+ - (1 - \alpha)\eta^- = -\frac{a}{a+b}\eta^+ + \frac{b}{a+b}\eta^- = \frac{a}{a+b} \underbrace{\left[-\eta^+ + \frac{b}{a}\eta^- \right]}_{=0 \text{ by (2.38)}} = 0 \quad (2.45)$$

Therefore, if $V(U, z)$ is concave in U , since $U^+ - U^- > \tilde{U}^+ - \tilde{U}^-$ condition (2.42) 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

$$U^*(z_i) < \tilde{U}^- < U^*(z_{i+1}) \quad \text{and} \quad U^*(z_{j-1}) < \tilde{U}^+ < U^*(z_j) < U^*(z_{j+1}) \quad (2.46)$$

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

$$\eta^+ = \left(\frac{1 - \pi(\epsilon)}{\pi(\epsilon)} \frac{[1 - \lambda_e + \lambda_e F(z_i)]}{[1 - \lambda_e + \lambda_e F(z_{j-1})]} \right) \eta^- \quad (2.47)$$

The values (η^+, η^-) satisfy conditions (2.46) and (2.47).

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

$$\begin{aligned} \tilde{V} - V = \beta\psi(1 - \delta) \left\{ [a'V(\tilde{U}^+) + bV(\tilde{U}^-)] + \underbrace{[V(U^*(z_j)I(z > z_j) - V(U^+))]}_{c \geq 0} \right. \\ \left. - [a'V(U^+) + bV(U^-)] \right\} \quad (2.48) \end{aligned}$$

where $a' = \pi(\epsilon)[1 - \lambda_e + \lambda_e F(z_{j-1})] > 0$ 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

$$\alpha'V(\tilde{U}^+) + (1 - \alpha')V(\tilde{U}^-) > \alpha'V(U^+) + (1 - \alpha')V(U^-) \quad (2.49)$$

By construction

$$\alpha'\tilde{U}^+ + (1 - \alpha')\tilde{U}^- = \alpha'U^+ + (1 - \alpha')U^- \quad (2.50)$$

Hence, given that $U^+ - U^- > \tilde{U}^+ - \tilde{U}^-$, condition (2.49) 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.

□

B.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.¹⁵

Over the length of one panel, households are interviewed every four months. 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

¹⁵<http://www.nber.org/data/sipp.html>

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 average unemployment rate is 3.6%.¹⁶ 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

¹⁶These 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.

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%.¹⁷

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.¹⁸

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 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. We use these finer categories also to calculate a person's potential labor market experience as age minus years of education

¹⁷Our 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.

¹⁸See for example Katz and Autor (1999) for similar sample restrictions.

minus five. 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, a quadratic in potential experience, interaction terms between education groups and the quadratic in experience, the non-white dummy, the region dummies as well as year dummies. The estimated standard deviation of log residuals is 0.48, while the mean and standard deviations of the exponentiated residual wages are 1.13 and 0.72, 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 0.006 and 0.14, respectively, and the fraction of negative changes is 35.5%.¹⁹

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 0.028, and the standard deviation is 0.36. Moreover, 39.7% of the log wage changes associated with job-to-job transitions are negative.²⁰ For the subsample of positive wage changes used in the calibration, the mean and standard deviation are 0.28 and 0.26, respectively.

¹⁹For comparison, before dropping outliers the mean of log wage changes within a job is 0.007, the standard deviation is 0.22, and the fraction of negative changes is 35.8%.

²⁰Before dropping outliers the mean of log wage changes between jobs is 0.023, the standard deviation is 0.45, and the fraction of negative changes is 40%.

Bibliography

- Alvarez-Parra, F. A. (2008). *Essays in Macroeconomics: Environments with Informational and Financial Frictions*. PhD thesis, University of Rochester.
- Autor, D. H. (2001). Wiring the Labor Market. *The Journal of Economic Perspectives*, 15(1):25–40.
- Bils, M., Chang, Y., and Kim, S.-B. (2008). Heterogeneity and Cyclical Unemployment. RCER Working Papers 543, University of Rochester.
- Bowlus, A. and Vilhuber, L. (2002). Displaced workers, early leavers, and re-employment wages. Technical report, Longitudinal Employer-Household Dynamics, Center for Economic Studies, U.S. Census Bureau.
- Buessing, M. and Soto, M. (2006). The State of Private Pensions: Current 5500 Data. Technical Report 42, Center for Retirement Research, Boston College.
- Burgess, S. M. (1992). A Search Model with Job Changing Costs: 'Eurosclerosis' and Unemployment. *Oxford Economic Papers*, 44:75–88.
- Cabrales, A., Calvo-Armengol, A., and Pavoni, N. (2008). Social Preferences, Skill Segregation, and Wage Dynamics. *Review of Economic Studies*, 75(1):65–98.
- Canova, F. (2007). *Methods for Applied Macroeconomic Research*. Princeton University Press, Princeton, N.J.
- Davis, S. J. (2008). The Decline of Job Loss and Why It Matters. *American Economic Review*, 98(2):263–67.

- Davis, S. J., Faberman, R. J., Haltiwanger, J., Jarmin, R., and Miranda, J. (2010). Business Volatility, Job Destruction, and Unemployment. *American Economic Journal: Macroeconomics*, 2(2):259–87.
- Dey, M. S. and Flinn, C. J. (2005). An Equilibrium Model of Health Insurance Provision and Wage Determination. *Econometrica*, 73(2):571–627.
- 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.).
- Heathcote, J., Storesletten, K., and Violante, G. L. (2010). The Macroeconomic Implications of Rising Wage Inequality in the United States. *Journal of Political Economy*, 118(4):681–722.
- Hey, J. D. and McKenna, C. J. (1979). To Move or Not to Move? *Economica*, 46(182):175–185.
- 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.
- Jolivet, G., Postel-Vinay, F., and Robin, J.-M. (2006). The Empirical Content of the Job Search Model: Labor Mobility and Wage Distributions in Europe and the U.S. *European Economic Review*, 50(4):877 – 907.
- Juhn, C., Murphy, K. M., and Pierce, B. (1993). Wage Inequality and the Rise in Returns to Skill. *The Journal of Political Economy*, 101(3):410–442.
- Kambourov, G. and Manovskii, I. (2009). Occupational Mobility and Wage Inequality. *Review of Economic Studies*, 76(2):731–759.
- 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.

- Lise, J., Meghir, C., and Robin, J.-M. (2012). Matching, Sorting, and Wages. 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.
- MacLeod, W. B. and Malcomson, J. M. (1998). Motivation and Markets. *The American Economic Review*, 88(3):388–411.
- 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.
- Michelacci, C. and Pijoan-Mas, J. (2012). Intertemporal Labour Supply with Search Frictions. *The Review of Economic Studies*, 79(3):899–931.
- Mortensen, D. T. (2003). *Wage Dispersion: Why are Similar Workers Paid Differently?* MIT Press, Cambridge, MA.
- Munnell, A., Haverstick, K., and Sanzenbacher, G. (2006). Job Tenure and the Spread of 401(k)s. Issue in Brief 55, Center for Retirement Research, Boston College.
- Munnell, A. and Perun, P. (2006). An Update on Private Pensions. Issue in Brief 50, Center for Retirement Research, Boston College.
- 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.
- Poterba, J., Venti, S., and Wise, D. A. (2007). New Estimates of the Future Path of 401(k) Assets . Working Paper 13083, National Bureau of Economic Research.
- 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.
- Sherk, J. (2008). Job-to-Job Transitions: More Mobility and Security in the Work Force. Report 08-06, The Heritage Foundation, Center for Data Analysis.
- Spear, S. E. and Srivastava, S. (1987). On Repeated Moral Hazard with Discounting. *The Review of Economic Studies*, 54(4):599–617.
- Stewart, J. (2002). Recent Trends in Job Stability and Job Security: Evidence from the March CPS. Working Paper 356, U.S. Bureau of Labor Statistics.
- Thomas, J. and Worrall, T. (1988). Self-Enforcing Wage Contracts. *The Review of Economic Studies*, 55(4):541–553.
- Tsuyuhara, K. (2011a). Repeated Moral Hazard with Worker Mobility via Directed On-the-Job Search. University of Calgary, mimeo.
- Tsuyuhara, K. (2011b). Work Incentives and Labour Mobility: Quantitative Assessment of Macroeconomic Implications. University of Calgary, mimeo.
- Violante, G. L. (2002). Technological Acceleration, Skill Transferability, And The Rise In Residual Inequality. *The Quarterly Journal of Economics*, 117(1):297–338.