Productivity Shocks, Long-Term Contracts and Earnings Dynamics*

Thibaut Lamadon †

October 21, 2016

Abstract

This paper examines how employer and worker specific productivity shocks transmit to earnings and employment in an economy with search frictions and firm commitment. I develop an equilibrium search model with worker and firm shocks and characterize the optimal contract offered by competing firms to attract and retain workers. In equilibrium risk-neutral firms offer risk-averse workers contingent contracts where payments are back-loaded in good times and front-loaded in bad ones, which only provides partial insurance against firm and worker shocks. I derive conditions for non-parametric identification of the model production function and worker and firm shock processes. I parametrize and estimate the structural model using matched employer-employee data from Sweden. Estimates suggest that firms absorb between 80% and 95% of productivity shocks, but price permanent worker differences. I evaluate the effects of re-distributive policies and find that 2/3 of government provided insurance is undone by the crowding out of firm provided insurance.

*I am thankful to Stephane Bonhomme, Jan Eeckhout, Jeremy Lise, Costas Meghir and Jean-Marc Robin for their invaluable advice and to Joseph Altonji, Richard Blundell, Hector Chade, Michael Greenstone, Jim Heckman, Lisa Laun, Rasmus Lentz, Magne Mogstad, Giuseppe Moscarini, Nicola Pavoni, Fabien Postel-Vinay and Shouyong Shi for their helpful comments and suggestions. I also thank seminar and conference participants at UCL, IFS, Yale University, Science-Po, University of Chicago, UPenn, Northwestern University, NYU, Stanford, University of Minnesota, University of Wisconsin and others. I gratefully acknowledge financial support from the cemmap, ESRC, the IFS and the Cowles Foundation. I am deeply thankful to the IFAU (www.ifau.se) for access to and help with the Swedish administrative data.

†Department of Economics, University of Chicago, lamadon@uchicago.edu, IFS and IFAU.

1 Introduction

Firms play an influential role in setting the level of earnings and employment risk for their workers when designing job contracts (Knight, 1921). Empirical work suggests that they choose to insure their workforce only partially against productivity shocks (Guiso, Pistaferri, and Schivardi, 2005). Understanding wage setting and contract formation is central to key questions in economics, including labor market attachment, wage growth over the life cycle and the earnings uncertainty that people face (Meghir and Pistaferri, 2011).

The theory of dynamic contracts provides a foundation for the use of long-term agreements between firms and risk-averse workers (Harris and Holmstrom, 1982; Thomas and Worrall, 1988; Holmstrom and Milgrom, 1991). However, combining optimal contracts with equilibrium models of job mobility with search frictions is challenging both theoretically and empirically. Such contracts do not always admit closed-form solutions (Abreu, Pearce, and Stacchetti, 1986), and history dependence makes identification difficult. In fact, the empirical literature on earnings and employment dynamics is often silent about how firms might endogenously choose the level of risk.

In this paper, I develop an empirical framework where firms optimally choose how productivity shocks transmit to their workers. I characterize the optimal contract in an equilibrium model with risk averse workers, search frictions with imperfect monitoring and both individual and firm level shocks. I establish that the model can be taken to data by providing a tractable solution and conditions for non-parametric identification of the productivity processes.

I estimate the model using matched employer-employee data from Sweden. Using the model I carry out decompositions of earnings dynamics to analyze the source of shocks faced by workers. I also quantify the re-distributive implications of government transfers in this world where firms can adjust the level of risk in the employment contracts they offer.

I consider an infinite horizon directed search model with risk averse workers[1]. The productivity of workers and jobs change over time and I introduce firm shocks by adding a correlation between job productivity shocks for co-workers. Firms remain ex-ante identical from the point of view of workers when they apply for jobs, but become ex-post heterogeneous as shocks come along. This allows for the analysis of the transmission of workers and firm specific shocks while keeping the tractability properties of the directed search equilibrium of Menzio and Shi (2010). To generate value for insurance against earnings and employment shocks in a tractable way, I assume that individuals do not have access to asset markets[2] and abstract from intensive labor supply considerations.

Firms post contracts that specify wages for each future productivity histories. Workers choose which job to apply to and an amount of effort to avoid job separation; both decisions are unobserved by their current employer. Contract flexibility is central to the goal of the paper. Piece rate contracts and Nash bargaining would impose shock transmission by construction. Firms here can choose to smooth payments, but full insurance is not a foregone conclusion since employers compete to retain workers who can privately search while employed.

[1]The pioneering work in directed search is due to (Montgomery, 1991; Peters, 1991; Moen, 1997; Shimer, 1996; Burdett, Shi, and Wright, 2001) as well as Menzio and Shi (2010), which this paper builds on.

[2]Introducing hidden savings is an active area in principal-agent environment Abraham and Pavoni (2008); Attanasio and Pavoni (2011); Doepke and Townsend (2006).

Solving numerically for this optimal contract becomes difficult in the presence of shocks with large support. The firm's problem can be written recursively using promised utility (Spear and Srivastava, 1987; Rogerson, 1985) but one still needs to solve for the promised value in each state of the world in the next period. I show that using promised marginal utility (Marcet and Marimon, 2011; Farhi and Werning, 2013) addresses this issue. It is optimal for the contract to provide insurance by equating marginal utilities across realizations of productivities. It is then sufficient to solve for one promised marginal utility, instead of a promised utility for each realization.

The optimal contract and the productivity shocks generate rich earnings dynamics. In particular, I show that wages respond to both worker and firm productivity shocks as suggested by the empirical literature (Guiso, Pistaferri, and Schivardi, 2005; Friedrich, Laun, Meghir, and Pistaferri, 2014). This is an important departure from a competitive market where earnings would track marginal productivity and not respond to firm shocks. This also implies that wages can decline even though firms have commitment power. In bad times, firms choose to lower wages of workers to incentivize them to search for a better job.

Given the complex mapping between latent productivities and realized earnings, it is important to develop conditions for the identification of the structural parameters of the model. I show that the independence assumption between worker and firm shocks allows one to separate the firm shocks from those of the worker using co-workers earnings. I then show how the joint Markov property of earnings and productivity allows one to use results on the identification of hidden Markov chains (Hu and Shum, 2012). When the preference functions are known, the production function and the productivity process can be recovered non-parametrically from data on earnings and employment of co-workers.

The model is estimated using administrative employment data from Sweden. Access to information on both employers and employees is crucial to separating worker and firm specific shocks. I first document the presence of common shocks at the firm level by looking at the co-variance of wage growth among co-workers and relate the results to the previous literature. I then estimate the model using method of moments and decompose the observed variance in earnings and earnings growth. I find that a large share (86%) of the earnings growth variance is due to dynamics associated with coming in and out of employment rather than shocks to productivity while employed. When looking at the pass-through of productivity shocks, I find that the maximum effect of the shock happens between two and four years after impact and that between 80% and 95% is insured by the contract. In contrasts, firms do not attenuate permanent productivity differences between workers.

Finally, using the estimated model, I evaluate the effect of policies that re-distribute earnings in order to affect inequality and risk faced by workers. When the government tries to provide insurance, firms respond by posting contracts with higher pass-through to create stronger incentives. Overall, two-thirds of the direct effect of such policies is undone by the firm adjusting the risk in the contracts they offer. This exercise demonstrates how, at realistic parameter values, taking into account firms' decisions to provide insurance can have important implications for policies.

Related literature. This paper contributes to the growing empirical literature that analyses how firm shocks transmit to worker earnings. Guiso, Pistaferri, and Schivardi (2005) (GPS henceforth) first documented the effect of permanent and transitory shocks. Carlsson, Messina, and Nordström Skans (2014) uses detailed information on product prices in Sweden to account for firms adjusting on their production frontier and finds results similar to the estimates on Italian data while documenting endogeneity concerns when using value added rather than firm sales. Recent work in Friedrich, Laun, Meghir, and Pistaferri (2014) develops and estimates a model with exogenous earning dynamics, including firm-level shocks, with endogenous mobility decisions. Roys (2011) uses French firm data to study the link between firm shocks, wage bill, and employment at the firm level in a model where wages are set according to Nash Bargaining. The current paper addresses the fact that the level insurance provided by firms might be endogenous.

There is an important theoretical literature on long-term contracts between firms and workers. Baily (1974) and Azariadis (1975) studied long-term contracts with commitment and developed the insurance role of the firm. Harris and Holmstrom (1982) (H&H henceforth) derived the optimal contract when workers can't commit and found that positive shocks will pass-through to workers. Thomas and Worrall (1988) looked at two-sided lack of commitment in the presence of rents. MacLeod and Malcomson (1989) develops the implications of non-verifiable output. In the search friction literature, Stevens (2004); Burdett and Coles (2003b); Shi (2008) derived optimal wage-tenure contracts and showed the presence of back loading even without shocks. Menzio and Shi (2010) derives the block-recursive properties of directed search equilibria with fixed wage and optimal contracts but does not actually characterize the link between wages and productivity. Tsuyuhara (2013) introduced the effort decision to keep the job active and demonstrated the presence of back-loading, but did not include shocks to productivity. Schaal (2010) characterizes the incentive-compatible contract in a directed search environment without risk aversion. Rudanko (2009) derives and evaluates the optimal contract with two-sided lack of commitment and aggregate shocks but without on-the-job search or private actions from the worker. Two recent working papers study the presence of search with imperfect monitoring. Lentz (2016) derives the optimal wage tenure contract while allowing for firms to respond to outside offers, but without shocks to productivity. Abraham, Alvarez-Parra, and Forstner (2016) studies a contract with moral hazard in production and its implications for cross-sectional wage dispersion. To my knowledge, the current paper is the first to characterize the long-term optimal contract offered in equilibrium by firms in an economy with search frictions, on-the-job search, firm and worker shocks and risk-averse workers.

This paper relates to the literature that studies the properties of earnings and employment dynamics. MaCurdy (1982); Abowd and Card (1986); Meghir and Pistaferri (2004); Altonji, Smith, and Vidangos (2009) study the processes of earnings and employment. Hall and Mishkin (1980); Blundell, Pistaferri, and Preston (2008); Low, Meghir, and Pistaferri (2010) evaluate how earnings shocks transmit to consumption and its implications for the role of government transfers. This paper helps understanding how the earnings process itself might change when government transfer reduces risk.

There is a long tradition of papers using structural search models to evaluate wage dispersion. Postel-Vinay and Robin (2004), Bagger, Fontaine, Postel-Vinay, and Robin (2014), and Lise, Meghir, and Robin (2008) estimate models of earnings with risk-neutral workers and sequential contracting. In this paper, I introduce risk aversion with optimal contracts to create a value for insurance explicitly.

Two important papers have evaluated empirically the presence of optimal contracts in the labor market. Chiappori, Salanie, and Valentin (1999) directly evaluates the presence of downward rigidity. Lemieux, Thomas, Bentley MacLeod, and Daniel (2009) studies the use and implications of performance pay. To my knowledge, the current paper is the first to use matched employer-employee data to estimate a search model with optimal contracting and both firm and worker shocks.

Outline. In Section 1, I present empirical evidence on the transmission of firm shocks to workers in the Swedish matched employer-employee data. In Section 2, I present the equilibrium search model, and I characterize the optimal contract in Section 3. In Section 4, I present the estimation strategy and the identification of the model. This section also reports the estimation results. In Section 5, I analyze the estimated model and evaluate the effect of the redistributive tax policy.

2 Earnings dynamics and firm-level shocks

2.1 Data and institutional background

The employer-employee matched data from Sweden links three administrative data-sets: the employment data, the firm data and the benefits data that tracks workers who are currently unemployed. The sample runs from 1993 to 2007 but I focus on five consecutive years between 2001 and 2005. On the worker side, all self-employed are dropped from the original sample, as well as some specific industries such as fisheries and the financial sector. I first de-trend the data with time dummies to remove any non stationary effects. I select individuals under 50 years of age, and, for moments computed at the firm level, I limit the data to firms with at least 10 employees. The sample is build on the data prepared in Friedrich, Laun, Meghir, and Pistaferri (2014) which also estimates a model of firm and individual shocks, but considers an exogenous process for earnings. The group of individuals considered here includes active and non active job seekers. This is important for mobility in and out of work which appears lower here than in gross accounting figures. For instance Table 1 reports 14.7% of non-employed, but unemployment in Sweden during this periods was reported around 7% and overall labor participation is around 65%. For the sample defined here and because of the age selection we end up with a figure in between.

An important aspect of the Swedish labor market is the presence of collective agreements. Indeed in the 90's many such agreements where put in place, specifying wage floors that where negotiated at the industry level or at the firm level. Fredriksson and Topel (2010) reports the share of workers that are covered by such agreements using sources from the Swedish National Mediation Office. They say that 11% of private sector workers are subjected to general pay increase and 7% bargain there wages without any restrictions. The remaining 82% is shared between workers with local bargain with fall back, local wage frame (total wage bill increase to be split between workers) and combination of these with pay increase. Overall 71% of private sector workers do not get a guaranteed general increase from their agreement. Nevertheless this are important considerations when thinking about the transmission of firm level shocks.

Table 1. Data description

	all individuals	stayers
number of workers	1,541,681	546,713
number of firms	72,676	11,150
number of observations	21,788,344	1,207,725
firm reported size for median worker	143	374
% male	66.3%	70%
% high education (some college)	29.7%	29.3%
% workers age ≤ 30	33.4%	23.8%
% workers in manufacturing	35.9%	46.4%
% workers in services	35.5%	29.2%
% workers in retail and trade	18.6%	16.9%
% workers in construction	10%	7.5%
% employed	85.3%	100%
mean log-earnings	10.09	10.17
variance of log-earnings	0.207	0.164

The two data-sets used to construct the different moments used in the introduction of the paper and in the estimation. The stayers data includes spells of at least 3 years of full year employment at firms with more than 10 workers.

2.2 A simple model of earnings and value added

In this section we look directly at the covariance between wages of co-workers and the value added at the firm level to draw a picture of the presence of shocks shared by co-workers at the firm level. We consider a simple statistical model for log value added and log earnings which will allow for a direct interpretation of some of the moments that we will use later in the structural estimation. This model is given by:

$$
\begin{aligned}
\text{log value added} \quad & y_{jt} &&= y^p_{jt} + y^t_{jt} \\
& y^p_{jt} &&= y^p_{j,t-1} + u^f_{jt} \\[1ex]
\text{log earnings} \quad & w_{it} &&= w^p_{it} + w^t_{it} \\
& w^p_{it} &&= w^p_{i,t-1} + \tau \cdot u^f_{j(i,t),t} + u^c_{j(i,t),t} + u^w_{it} \\[1ex]
\text{where} \quad & y^t_{jt}, u^f_{jt}, u^c_{j,t} && \text{iid and mean 0 across } (j,t) \\
& w^t_{it}, u^w_{it} && \text{iid and mean 0 across } (i,t)
\end{aligned}
$$

where w_{it} represents the residual log-earnings for worker i at time t, composed of a permanent part w^p_{it} and a transitory one w^t_{it}. $j(i,t)$ is the firm identifier of worker i at time t. y_{jt} is the residual log value added for firm j, also composed of a permanent y^p_{jt} and transitory part y^t_{jt}. The transitory components are assumed to be classical (independent across individuals or firms and across times). u^f_{jt} is the permanent innovation shock to value added at time t. The innovation to permanent earnings has three components. First, the permanent shock to value added u^f_{jt} enters with a scaling factor τ, next the shock $u^c_{j,t}$ represents a shock shared by co-workers in firm j which does not appear in the value added process. I allow for this term since value added can be a very noisy process, aggregated on the entire firm, and part of the shock to worker productivities might not show up in the value added process directly. The final innovation shock u^w_{it} is an individual specific shock to earnings.

The model parameters can be estimated using the moments directly (See Appendix C.1), yet one needs to be careful about the effect of smaller firms. I evaluate concerns associated with incidental parameter bias in the Appendix. The confidence intervals are computed using bootstrap. Since the unit of

Table 2. Residual earnings of stayers

		value	Conf. Interval $q = 0.025$	$q = 0.975$	Cons. Eq.
Moments					
wage growth variance		5.40e-02	5.24e-02	5.58e-02	
covariance between co-workers' wage growth		6.89e-04	5.49e-04	7.56e-04	
covariance between wage growth and VA growth		3.67e-04	1.35e-04	5.89e-04	
Passthrough paramater and shock variances					
passthrough parameter	τ	1.40e-02	5.00e-03	2.60e-02	
worker transitory	$Var[w^t_{it}]$	1.51e-02	1.46e-02	1.57e-02	
worker permanent:					
- idysyncratic	$Var[u^w_{it}]$	2.31e-02	2.22e-02	2.39e-02	-34.33%
- explained by VA	$Var[\tau u^f_{j(i,t),t}]$	5.00e-06	1.00e-06	1.40e-05	-0.01%
- other common at firm	$Var[u^c_{j(i,t),t}]$	6.84e-04	5.41e-04	7.48e-04	-1.12%
VA transitory	$Var[y^t_{jt}]$	3.12e-02	2.58e-02	3.70e-02	
VA permanent	$Var[u^f_{jt}]$	2.63e-02	1.52e-02	3.75e-02	-38.58%

Standard errors are computed using clustered re-sampling at the firm level. Earning differences are taken year on year. Firm level outcomes are weighted by number of workers per firm. Consumption equivalent use CRRA utility with $\rho = 1.5$ and $r = 0.015$ see appendix C.3. See Table 9 for full set of moments and parameters.

observation in this exercise is the firm, I sample firms with replacement and attach all spells for that firm. This replicates both the dependence between workers within the firm and the dependence over time. Table 2 reports the estimates. Finally, I run a placebo test. I randomly assign workers to firms. This is in order to test that the procedure does not generate any dependence by construction.

The standard deviation of the worker idiosyncratic permanent shock is 0.15. This is a larger value than GPS which has 0.108. The value of $Var[u^c_{j(i,t),t} + \tau u^f_{j(i,t),t}]$ is of interest as it represents the size of the permanent earning shock which is common to all co-workers within the firm. The standard deviation of the shock is 0.02 which is relatively smaller but still relevant. To get an economic sense of its size, the table also reports the consumption equivalents[3] of these socks. An individual would be willing to give up 40.3% of his consumption to avoid the idiosyncratic shock and 1.2% to avoid the shocks shared at the firm level.

This simple model distinguishes between permanent shocks to workers that are associated with value added (τu_{jt}^{f}) from permanent shocks shared at the firm level which are not (u_{jt}^{c}). In this sample, the common shock turns out be much bigger that the part associated with value added. The value added part represents a consumption equivalent drop of 0.01%. The link between wages and value added reported in Guiso, Pistaferri, and Schivardi (2005) gave consumption value of insuring against the remaining transmitted firm shocks equal to 0.045% of consumption, which is in the same order of magnitude as here. Our simple decomposition suggests that focusing on the link to value added only might understate the size of these shocks.

3 The contracting model

I present here an equilibrium model with search frictions and private worker actions. The key feature of the model is to embed the bilateral relationship between the firm and the worker, with productivity uncertainty, inside a competitive search equilibrium where firms compete to attract and retain workers. It is important to note that this framework is relevant beyond labor questions and should be helpful for studying dynamic contract in other markets with

[3] I consider a CRRA utility function $u(w) = \frac{w^{1-\rho}}{1-\rho}$ with $\rho = 1.5$ and discount rate $r = 0.015$. These values are borrowed from Low, Meghir, and Pistaferri (2010). I derive the expression for the consumption equivalent in the Appendix C.3.

bi-lateral relationships[4].

In this model, ex-ante identical firms compete by posting long-term contracts to attract heterogeneous workers. Employed and unemployed workers observe the menu of contracts offered in equilibrium and decide which one to apply to. This process forms sub-markets of workers applying to particular contracts and firms offering them. Within each queue the matching between firms and workers is random. When choosing which sub-market to participate in, both firms and workers take into account the value of matching and the probability of matching. This probability is driven by how many firms and workers participate in a particular sub-market.

When matched, the contract specifies the wage after each possible history of shocks for the firm and workers. Given his wage profile, the worker chooses which sub-market to visit while employed and chooses effort, which directly affects the probability the current match continues to exist. Both of these actions are private and so unobserved by the firm. Firms take this into account and post contracts that incentivize the worker's action in an optimal way. This will mean that in some cases the wage will adjust downward albeit in a smooth way. I now formally introduce the model.

3.1 Environment

Agents and preferences. Time is discrete, indexed by t and continues for ever. The economy is composed of a discrete uniform distribution of infinitely lived workers with ability indexed by $x \in \mathbb{X} = \{x_1, x_2 ... x_{n_x}\}$. Workers want to

[4] See Hendel and Lizzeri (2000) for commitment in insurance markets, and a recent paper Boualam (2015) for relational banking.

maximize expected lifetime utility, $\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t (u(w_t) - c(e_t))$ where utility of consumption $u : \mathbb{R} \to \mathbb{R}$ is increasing and concave and cost of effort $c : \mathbb{R} \to \mathbb{R}$ is increasing and convex with $c(0) = 0$. Worker's ability x changes over time according to Markov process $\Gamma_x(x_{t+1}|x_t)$. Unemployed workers receive flow value of unemployment $b(x)$. The other side of the market is composed of a uniform distribution of ex-ante identical firms with active jobs and vacancies. Vacancies live for one period and become active jobs if matched with a worker.

An active job is characterized by the current worker ability x and the current match quality z. The match quality z evolves with an innovation ι_t drawn at the firm level such that $z_{t+1} = g(z_t, \iota_t)$. ι_t is a firm level shock that affects all continuing workers' firm specific productivity. New hires start with a draw $z_{t+1} = g(z_0, \iota_t)$ for some fixed z_0. The function $g(\cdot, \cdot)$ is assumed to generate a monotonic transition rule. Every period a match (x_t, z_t) has access to a technology that produces $f(x_t, z_t)$. Worker's effort e affects the probability that the job continues to exists next period. This captures the idea that a negligent worker might loose a client or break the machine and cause the job to disappear. The firm cares about the total discounted expected profit of each created vacancy.

Firms here operate constant return to scale production functions and can be thought of as one worker per firm. However, empirically one cannot aggregate firms with the same output as the history of productivity shocks affects the distribution of workers. For instance whether or not a firm had a very bad shock in the last period will affect the current distribution of workers beyond the current productivity. To pin down the distribution of workers in a given firm one needs to know the entire history of shocks.

Search markets. The meeting process between workers and firms vacancies is constrained by search frictions. The labor market that matches workers to vacancies is organized in a set of queues indexed by $(x, v) \in \mathbb{X} \times \mathbb{V}$ where x is the type of the worker and v is the value promised to her in that given queue. Firms can choose in which (x, v) lines they want to open vacancies and workers can choose in which v line associated with their type x they want to queue[5]. Each visited sub-market is characterized by it's tightness represented by the function $\theta : \mathbb{X} \times \mathbb{V} \to \mathbb{R}_+$ which is the ratio of number of vacancies to workers. The tightness captures the fact that a high ratio of vacancies to workers will make it harder for firms to hire. In a directed search model like the one presented here, the tightness is queue specific which means that different worker types could be finding jobs at different rates. In queue (x, v) a worker of type x matches with probability $p(\theta(x, v))$ and receives utility v. Firms post vacancies at unit cost η and when posting in market (x, v) the vacancy is filled with probability $q(\theta(x, v))$. $\phi(x, v)$ will denote the mass of vacancies created in market (x, v).

States and actions. A worker is either employed or unemployed and enters each period with a given ability x. When unemployed, she collects benefit $b(x)$ and can search every period. When searching she chooses which sub-market (x, v) to visit, in which case she gets matched with probability $p(\theta(x, v))$ and if matched joins a job and receives lifetime utility v.

An employed worker is part of a match and starts the period with a given

[5] Menzio and Shi (2009) Theorem 3 tells us that workers will separate by type in equilibrium if markets are indexed by the value that each type x would get in a particular sub-market ($\mathbf{v} = (v(x_1), v(x_2)...v(x_{n_x})) \in \mathbb{R}^{n_x},$) and workers can apply to any. At equilibrium only a given type x visits a particular market. This market can then be represented directly by (x, v) as done in the current paper.

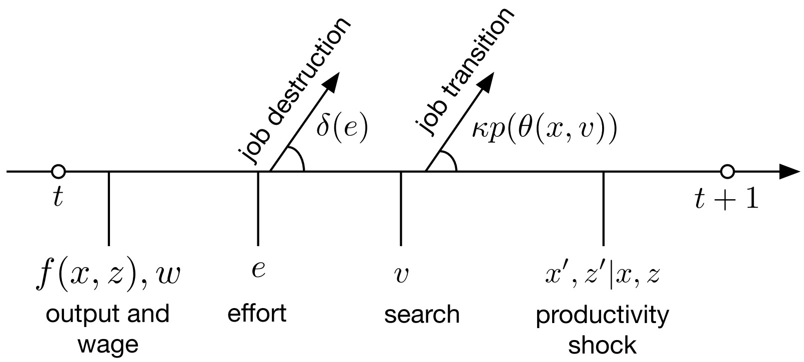


Figure 1: within period time line

ability level x and a current match quality z. The period is then divided in four stages as illustrated in Figure 1, first is production, the firm collects output $f(x,z)$ and pays the wage w to the worker. The worker cannot save, consumes all of w, chooses effort e and gets flow utility $u(w) - c(e)$. With probability $(1 - \delta(e))$, where $\delta(e)$ is decreasing in e, the employment persists to the next period. With probability $\delta(e)$ the worker moves to unemployment. In the search stage, the worker is allowed to search with efficiency κ. When searching she chooses which sub-market (x, v) to visit and gets matched with probability $\kappa p(\theta(x, v))$. If matched she moves to a new match where she will enjoy v and the current job will be destroyed. If the worker is not matched to a new job, the current job persists, a new x' is drawn conditional on the old one, and a firm level shock ι is drawn to update z. In summary, in every period an active job chooses the wage w, and the worker chooses effort e and which sub-market (x, v) to search in. Because $c(0) = 0$ the worker can quit in every period if the firm does not promise enough. By choosing v and e the worker controls his transition to other jobs and to unemployment.

Informational structure and contracts. A contract defines the transfer and actions for the worker and the firm within a match for all future histories. Call $s_\tau = (x_\tau, z_\tau) \in \mathbb{S} = \mathbb{X} \times \mathbb{R}$ the state of the match τ periods in the future and call $s^\tau = (s_1 ... s_\tau) \in \mathbb{S}^\tau$ a given history of realizations between s_1 the state today and s_τ, the state in τ periods.

The history of productivity is common knowledge to the worker and the firm and fully contractible. However the worker's actions are private information and transitions to other firms or to unemployment are assumed to be not contractible. This rules out side payments as well as countering outside offers[6]. The contract offered by the firm to the worker is then represented by:

$$\mathcal{C} := (\mathbf{w}, \boldsymbol{\zeta}); \quad \text{with } \mathbf{w} := \{w_\tau(s^\tau)\}_{\tau=0}^{\infty} \text{ , and } \boldsymbol{\zeta} := \{v_\tau(s^\tau), e_\tau(s^\tau)\}_{\tau=0}^{\infty}, \tag{1}$$

I explicitly separate the firm's choice from the worker's response. The firm chooses the wage w_τ paid at every history and the worker responds by choosing (v_τ, e_τ) the search and effort decision[7]. $\boldsymbol{\zeta}$ can be thought as the action suggested by the contract and I will focus on contracts where the recommendation is incentive compatible. The contract space is completely flexible in the way it responds to tenure and any productivity history. In particular it leaves the firm free to chose how the wage should respond to productivity shock, which is the central question of this paper.

[6] Lentz (2016) develops a model with optimal contracts and countering of outside offers, but without productivity shocks, and shows that firms continue to back-load wages.

[7] Derivations will later require a randomization which means that the contract can specify simple probability over actions instead of actions themselves. This is left implicit at this point but will be clarified in the recursive formulation of the problem.

3.2 Worker choices

An unemployed worker of type x chooses optimally which sub-market (x, v_0) she applies to. The only value she cares about is the value she will get, specifically v_0 and the tightness of the market $\theta(x, v_0)$. Higher v_0 sub-markets deliver higher values but have longer average waiting times. I can write the value $\mathcal{U}(x)$ of being unemployed as follows:

$$\mathcal{U}(x) = \sup_{v_0 \in \mathbb{R}} b(x) + \beta p(\theta(x, v_0))v_0 + \beta \left(1 - p(\theta(x, v_0))\right) \mathbb{E}_{x'|x}\mathcal{U}(x'). \quad \text{(W-BE)}$$

We follow by writing the problem of the employed worker and the firm as a recursive contract. As presented in Spear and Srivastava (1987) the state space is augmented with V, the promised utility to the worker. The recursive contract is characterized at each (x, z, V) by $\{\pi_i, w_i, e_i, v_{1i}, W_{ix'z'}\}_{i=1,2}$ where $\pi_i : \mathbb{S} \times \mathbb{V} \to [0, 1]$ is a randomization, $w_i : \mathbb{S} \times \mathbb{V} \to \mathbb{R}_+$ is the wage, $e_i : \mathbb{S} \times \mathbb{V} \to [0, \bar{e}]$ is effort choice, $v_i : \mathbb{S} \times \mathbb{V} \to [0, \bar{v}]$ is the search choice and $W_{ix'z'} : \mathbb{S} \times (\mathbb{X} \times \mathbb{R}) \to \mathbb{V}$ is the utility promised for each realization next period.

The worker optimally chooses the action (v, e), when promised next period expected utility $W = \mathbb{E}_{x'z'}W_{x'z'}$, she solves the following problem:

$$\sup_{v,e} u(w) - c(e) + \delta(e)\beta\mathbb{E}_{x'|x}\mathcal{U}(x') + (1 - \delta(e))\beta\kappa p(\theta(x, v))v + \beta(1 - \delta(e))(1 - \kappa p(\theta(x, v)))W,$$

for which we define the associated worker policies $v^* : \mathbb{X} \times \mathbb{V} \to [0, \bar{v}]$ and $e^* : \mathbb{X} \times \mathbb{V} \to [0, \bar{e}]$. Because of the properties of $p(\cdot)$, $\theta(\cdot, \cdot)$ and $c(\cdot)$, those functions are uniquely defined. Note that those policies only depend on the promised utility for next period and not on the current (z, V) as stated in the following definition.

Definition 1. *We defined the composite transition probabilities* $\tilde{p} : \mathbb{X} \times \mathbb{V} \to \mathbb{R}$ *and the utility return to the worker* $\tilde{r} : \mathbb{X} \times V \to \mathbb{R}$ *as functions of the promised utility* W *(using short-hand* $e^* = e^*(x, W)$ *and* $v_1^* = v_1^*(x, W)$*):*

$$\begin{aligned}
\tilde{p}(x, W) =& \kappa(1 - \delta(e^*))\,(1 - p(\theta(x, v_1^*)))\\
\tilde{r}(x, W) =& - c(e^*) + \beta\kappa\,(1 - \delta(e^*))\,p(\theta(x, v_1^*))\,(v_1^* - W)\\
&+ \delta(e^*)\beta\mathbb{E}_{x'|x}U(x') + \beta\,(1 - \delta(e^*))\,W.
\end{aligned}$$

These functions capture everything the firm needs to know about the consequences of setting the wage dynamically. We now turn to the firm's problem.

3.3 Firm profit, optimal contracting problem

I can now describe the firm's problem in terms of promised utilities. The firm chooses a lottery over promised values and wages which then determines the participation probabilities. The expected profit of a match to the firm can be expressed recursively as

$$\begin{aligned}
\mathcal{J}(x, z, V) = & \sup_{\pi_i, w_i, W_i, W_{ix'z'}} \sum_{i=1,2} \pi_i \Big(f(x, z) - w_i + \beta\tilde{p}(x, W_i)\mathbb{E}_{x'z'}\mathcal{J}(x', z', W_{ix'z'}) \Big)\\
s.t \qquad & V = \sum_i \pi_i\,(u(w_i) + \tilde{r}(x, W_i))\,, \qquad \text{(BE-F)}\\
& W_i = \mathbb{E}W_{ix'z'}, \quad \sum \pi_i = 1.
\end{aligned}$$

The firm chooses the current period wage w_i and the promised utilities $W_{ix'z'}$ for each lottery realization i and state (x', z') tomorrow. These control variables must be chosen to maximize expected returns subject to the promise keeping constraint. This constraint makes sure that the choices of the firm honors the promise made in previous periods to deliver the value V to the worker. The right hand side of the constraint is the lifetime utility of the worker given the choices made by the firm. The lottery is present only to insure concavity of the function. The incentive compatibility of the worker is embedded in the $\tilde{r}$ and $\tilde{p}$ functions that we defined previously.

Finally firms choose how many vacancies to open in each (x, v) market. Given vacancy creation cost φ and the fact that the match quality z starts at z_0, the return to opening a vacancy is given by:

$$\Pi_0(x, V) = \sup_{W_{0x'z'}} q(\theta(x, V))\mathbb{E}[\mathcal{J}(x', z', W_{0x'z'})|x{=}x, z{=}z_0] - \varphi$$
$$\text{s.t.} \quad \mathbb{E}[W_{0x'z'}|x_t{=}x, z_t{=}z_0] = V, \tag{BE-V}$$

and firms will open vacancies in a given market if and only if expected profit is positive. The timing here is also important. The firm is able to promise different utilities in different (x', z') realizations, and can provide insurance on the very first wage payments. The vacancy creation cost is linear, which means that if $\Pi_0(x, V)$ is positive the firm will create an infinity of vacancies, if it's negative it won't create any and if it's zero the firm is indifferent.

3.4 Equilibrium definition

Free entry condition. We now impose a free entry condition on the market. Firms will open vacancies in each markets until the the expected profit is zero or negative:

$$\forall (x, V) \in \mathbb{X} \times \mathbb{V} : \quad \Pi_0(x, V) \leq 0. \tag{EQ1}$$

This will pin down the tightness of each market. $\phi(x, v)$ will denote the total mass of vacancies posted in market $(x.v)$.

Market clearing. Markets for labor must clear, in the sense that the equilibrium distribution must be generated by the equilibrium decisions. Given an equilibrium stationary distribution $h(x,z,V)$ of workers assigned to matches with a given promised utility, given the mass $\phi(x,V)$ of vacancies, the following clearing condition must be satisfied:

$$\forall x, v \quad \phi(x,v) = \theta(x,v)\Big[u(x)\mathbf{1}[v_0^*(x) = v] \\ + \sum_{x,z}\int_{V'}\sum_i \pi_i(x,z,V')\mathbf{1}[v_{1i}^*(x,W_i) = v]\ \mathrm{d}H(x,z,V')\Big]. \quad \text{(EQ2)}$$

There is one last market clearing equation for the distribution of active jobs $h(x,z,V)$ and it states that $h(x,z,V)$ in the next period is consistent with itself, all the equilibrium decisions, and law motions such as the shocks on x, z and the endogenous separation.

Definition 2. *A* ***stationary competitive search equilibrium*** *is defined by a mass of vacancies $\phi(x,v)$ across sub-markets (x,v), a tightness $\theta(x,v) \in \mathbb{R}$, an active job distribution $h(x,z,V)$ and an optimal contract policy $\xi = \{\pi_i, w_i, e_i, v_{1i}, v_0, W_{ix'z'}, W_{0x'z'}\}_{i=1,2}$, such that:*

(a) ξ solves the firm optimal contract problem BE-F and so satisfies worker incentive compatibility, and solves the vacancy problem BE-V,

(b) $\theta(x,v)$ and $\phi(x,v)$ satisfy the free entry condition EQ1 for all (x,v),

(c) $\theta(x,v)$, $\phi(x,v)$ and $h(x,z,V)$ solve the market clearing condition EQ2,

(d) $h(x,z,V)$ is generated by $\phi(x,v)$ and ξ.

The equilibrium assigns workers to firms with contracts in a way where neither workers or firms have an incentive to deviate. The distributions ϕ and h represent the equilibrium allocation.

3.5 Contract characterization

Menzio and Shi (2010) provides the important results that a block recursive equilibrium exists in the version of this model with aggregate shocks and no worker effort or heterogeneity, and Tsuyuhara (2013) proves the existence with effort but without shocks or job heterogeneity. In this section I focus on deriving results on the properties of the optimal contract between the worker and the firm.

Lemma 1. *The Pareto frontier $\mathcal{J}(x, z, V)$ is continuously differentiable, decreasing and concave with respect to V and increasing in z.*

Proof. See appendix A.2 □

Concavity is a direct implication of the use of the lottery. I then adapt the sufficient condition from Koeppl (2006) for differentiability in two-sided limited commitment models. From the free entry condition, the tightness function is a continuously differentiable and concave function of $\mathcal{J}(x, z, V)$, which implies that the composite search function $p(\theta(x, v))$ inherits those properties for all $x \in \mathbb{X}$. We are interested in how firms decide to compensate workers over time given that they face the usual trade-off between insurance and incentives. The following proposition provides a clear prediction for how wages move dependent on the current state of the match:

Proposition 1. *For any current state (x_t, z_t, w_t), within each lottery realization i, the following relationship between wage growth and expected firm profit holds:*

$$\eta(x_t, W_{it}) \cdot \mathbb{E}_t \mathcal{J}_{i,t+1} = \frac{1}{u'(w_{i,t+1})} - \frac{1}{u'(w_t)}, \qquad \text{(EQ-FOC)}$$

where $\eta(x, W) = \frac{\partial}{\partial W} \log \tilde{p}(x, W) \geq 0$ *is the derivative of the log-probability that the relationship continues into the next period with respect to the value promised to the worker,* $\mathbb{E}_t \mathcal{J}_{i,t+1} = \mathbb{E}\mathcal{J}(x_{t+1}, z_{t+1}, W_{ix_{t+1}z_{t+1}})$ *is the expected profit for the firm next period and* $w_{i,t+1}$ *is the wage the firm will pay to the worker next period.*

Proof. See Appendix A.3. □

The optimal contract creates a link between wages and productivity. More precisely, denote $w^*(x, z)$ the wage that makes the worker the residual claimant of the firm, meaning the wage such that for a given (x, z) the expected profit for the firm is zero: $\mathbb{E}[\mathcal{J}_{t+1} | x_t{=}x, y_t{=}y, w_t{=}w^*(x, y)] = 0$. We see that whenever $w_t = w^*(x, z)$ at state (x, z), the wage will not change since the left hand side of EQ-FOC is zero. And otherwise, since $\eta(x, W_i) \geq 0$, the wage growth will have the same sign as the expected profit of the firm. Whenever the firm expects positive profits, it will be optimal to increase the wage, and whenever the profits are expected to be negative, it will be optimal to decrease it. At any state (x, z, V), the wage change $w_{t+1} - w_t$ will be in the direction of w^* (see Corollary 1 in Appendix). For all histories of shocks, the change in wage will be positive if and only if the expected profit for the firm is also positive. This implies that the realized wage will smoothly track this reference wage $w^*(x, z)$, and hence the wage will respond to both positive and negative productivity shocks, and both to firm and worker specific shocks. This wage setting generates a rich set of features that I will now describe and link to the literature.

Backloading. The firm is able to commit, but even in the absence of shocks it would choose to "tilt" the wage instead of perfectly smoothing the consumption of the worker. This feature is the well known back-loading property of long term contracts with lack of commitment on the worker side. The worker makes effort and search decisions that affect the probability that the match continues to exist next period. When some of the match surplus goes to the firm, the worker does not internalize the full future value when making these decisions (unless he is at w^*). It is then optimal for the firm to front-load some profits and back-load wages. Stevens (2004) exposed this feature in a search environment with risk neutral agents and a minimum wage constraint. In the current paper, the worker is risk averse, and so the contract optimally balances the incentive (or commitment) problem with the desire for consumption smoothing as already pointed out in Burdett and Coles (2003a); Shi (2008); Tsuyuhara (2013).

Transmission of productivity shocks. Next, to better describe how wages respond to shocks, Figure 2 draws the path of the wage under different contract arrangements in response to worker and firm level shocks, both positive and negative. The green line represents a full commitment contract, the blue line represents one sided limited commitment (as in H&H), and the red line represents the presence of imperfect monitoring (the current paper). In the presence of full commitment, the firm will be able to insure the worker against all shocks as pointed out in Azariadis (1975), hence the green line remains flat in all cases. We are now going to explore the responses to positive and negative shocks in the presence of lack commitment and hidden actions.

Positive productivity shocks. Equation EQ-FOC reveals that when the firm makes positive expected profit, the change in inverse marginal utility of the worker is positive, meaning that the wage increases. This is a common

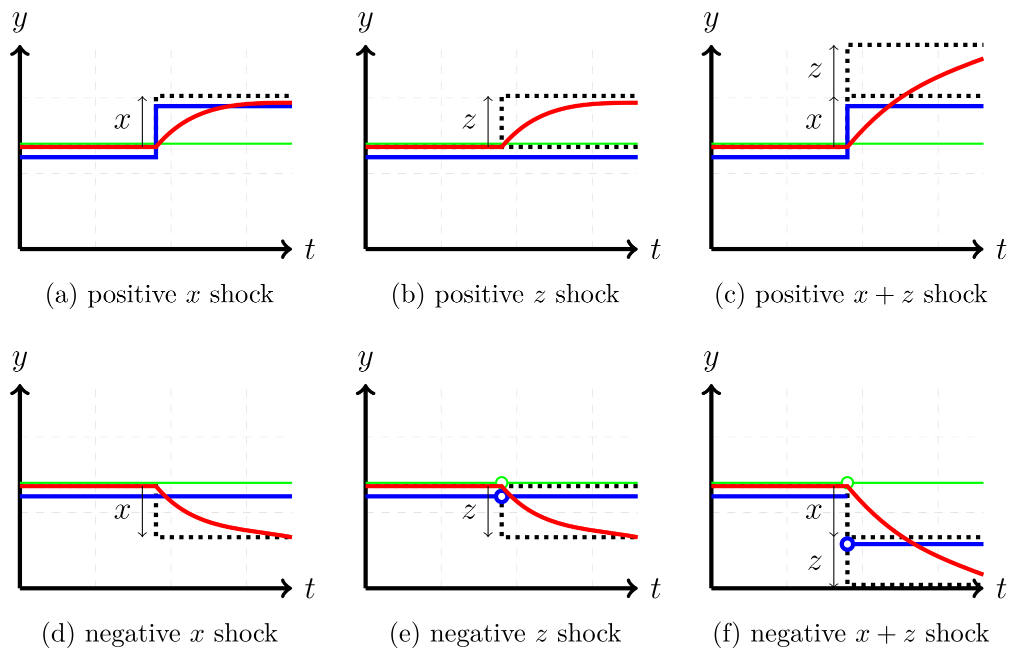


Figure 2: Contracting and productivity shocks.
This figure plots the responses of the wage in different contracting environment when faced with different underlying productivity shocks. The green line represents full insurance contract, blue line is firm commitment, red line is firm commitment and incentive. The dotted lines represent the productivities inside and outside the current job

feature of long-term contracts with firm commitment. For instance in H&H friction-less environment, the firm can't offer full insurance to the worker because if his productivity increases, the worker is unable to commit ex-ante not to take the market wage. I represent this in Figure 2a with the blue line jumping on a positive x shock. In the current paper the participation constraint is replaced with an incentive constraint, since finding a new job is probabilistic. This results in a smoother response as shown by the red line in Figure 2a. In the case of a z shock in H&H, the firm does not need to increase the wage, since the worker's outside option does not change. With an incentive problem however, when z increases, the agent's incentive is not aligned with the firm's. Here, the firm chooses to increase the wage to retain the worker. This is depicted by in Figure 2b&2c. We can see this directly in Equation EQ-FOC where the relevant quantity on the left side is the expected profit of the firm, irrespectively of whether it is driven by an increase in z or x, or both.

The mechanism in the current paper is not the lack of commitment but an incentive problem. However they are not as different as they seem. At the time where the worker gets a positive productivity shock in H&H, the worker can choose to take the spot-market offer. This decision is akin a very stark incentive problem. It can be thought of as a limiting case where $p(x, V) = 1[v \leq V^*(x)]$, where a worker of type x can get any value $v \leq V^*(x)$ with probability one but no value above. There are two important commonalities between the current paper and H&H. First, without the effort decision, there would be a similar area above the highest value promised by vacancies, without incentive constraint, where full insurance would be provided. Second, the downward rigidity appears here as an asymmetry in the size of the wage response above and below $w^*(x, z)$.

Negative productivity shocks. The contract of H&H will not respond to negative shocks to x, as described in the original paper. Since the firm can commit, and since the participation constraint of the worker does not bind in this case, the firm will provide full insurance. In the presence of imperfect monitoring however, as shown in Equation EQ-FOC, the contract will respond to each shock as depicted in Figure 2d.

A negative firm shock z is an interesting case. The commitment of the worker binds in the opposite direction of Figure 2a. Since the match in the current job is lower, it would be efficient for the worker to move to a new job, which is also what the firm would prefer. However the worker is enjoying a high wage because of initial conditions. In this case, the firm will lower the wage to make the worker leave to the new better match. In the case of a z shock only, in H&H, the worker will move immediately and keep the same wage. In the case of imperfect monitoring with mobility frictions the wage will decrease slowly, and eventually the worker will find a job (see Figure 2e).

Interestingly, the combination of a z and x shock can generate a wage decline, even in the H&H contract. In order for the worker to move to a new firm $z' = z_0 > z$, he has to be willing to take the new lower spot wage because of his lower productivity $x' < x$. In cases where this is efficient, the contract will lower the wage to force him to take the new job. It is the interaction of inefficient mobility and worker's lack of commitment that can trigger the wage decrease even with firm commitment. This never happens in the original H&H because all jobs are identical, and mobility is irrelevant from an efficiency point of view. For the same reason, the wage decreases in the presence of the incentive constraint. Effectively, the worker chooses to search for options that are too hard to get, and exerts too high of an effort to keep his job. This is because he is receiving more than the value of the job from his current contract. This dynamics are shown in Figure 2f.

Finally, we should remind ourselves why the firm chooses to lower the wage smoothly in bad times instead of fully renegotiating as in Thomas and Worrall (1988). The firm here uses its commitment power to deliver insurance in expectation to the worker at time of hiring. When allowed to commit, it finds it profitable to do so. The wage goes down in Thomas and Worrall (1988) because the firm participation constraint binds. Here, the firm is allowed to commit to histories where expected profit is negative, just as in H&H, and offering a lower wage first, but smoothing against downward shocks is the optimal way to deliver a given value V to the worker from the perspective of the time of hiring.

Amount of insurance. Finally, from Equation EQ-FOC, we note that in this environment, the level of transmission of shocks is not directly linked to the amount of mobility observed in the data, but rather has to do with how the mobility can be affected by the promised wage. This is captured by η which multiplies expected profit in Equation EQ-FOC. The incidence of the shock has to do with the severity of the agency problem.

4 Estimating the contracting model

There are two important challenges to overcome when taking the model to the data. The first difficulty has to do with identification. Individual productivity is not observed, and the way it translates into earnings and participation is very non-linear. The structural estimation will use a parametrized version of the model with a minimum distance approach. Yet, to better understand how we can hope to estimate such a model from data, I derive conditions for the non-parametric identification of the objects of interest in the model. The second main difficulty has to do with tractability. Solving directly for promised utilities in each future state realization is not feasible. I now address these two concerns, but the reader can jump to the next section without loss of continuity.

Non-parametric identification. In the context of this paper we are particularly interested in recovering the production function and the dynamics of productivity for workers and jobs. In appendix B, I derive sufficient conditions such that knowing preference functions $u(\cdot)$ and $c(\cdot)$, the production function $f(\cdot,\cdot)$, the search function $p(\cdot,\cdot)$ and the productivity processes Γ_x, Γ_z are identified from a five year panel dataset on workers' and co-workers' earnings and participation. The proof proceed in three steps. In the first step, I show how the fact that earnings path of co-workers are independent of each other conditional on firm level productivity history can be used to isolate firm shocks from worker shocks. Data on three co-workers who started at the same time is sufficient for identification. The second step, uses the Markovian and limited dependence properties of the optimal contract in wages and productivity. Using the result from Hu and Shum (2012)[8] on the identification of hidden Markov chains gives us identification of the wage process and participation conditional on latent variables x and z. The third and final step requires recovering the structural parameters. In an approach akin to conditional choice probability estimation, I show that the present value of the worker, the firm, and the production function can be expressed as a function of objects recovered from the second step. This requires expressing the value of being unemployed and wages after transitions as function of observables as well. The reader should refer to Appendix B for the formal results.

It is important to notice that such identification relies on a particular property of the optimal contract about the limited dependence between earnings and x, as well as the conditional independence of co-workers once we have

[8]The conditions of Hu and Shum (2012) are very demanding in terms of support (Connault, 2015). In the identification section, I consider a modification of the model that includes independent random commitment shocks and create spread around the w_{t+1} value conditional on x_t, z_t, w_t. I also look at what happens with the deterministic rule.

conditioned on firm shocks. Secondly, the proof does not require the use of equilibrium constraints. This suggests that one could estimate ex-post vacancy cost and the shape of the matching function without imposing it first.

A tractable solution. The main difficulty resides in solving the firm problem where tackling directly (BE-F) requires finding the promised utilities $W_{z'x'}$ in each state of the world for the next period. This becomes infeasible as soon as reasonable supports are considered for $\mathbb{X}$ and $\mathbb{Z}$. However, the first order condition with respect to W reveals that the utilities promised in different states are linked to each other and that it is optimal for the firm to promise identical marginal utilities across states tomorrow. The solution is then characterized by a single promised marginal utility. In Appendix A.4, I rewrite the contracting problem recursively using marginal utilities[9] and use this to solve the contract numerically.

4.1 Model specification and estimation

I estimate the model using simulated method of moments. I use the constant relative risk aversion utility function $u(w) = \frac{w^{1-\rho}-1}{1-\rho}$. The discount rate for the worker and the interest rate for the firm are set to an annual 5% and the model is solved quarterly. The flow value of being unemployed is set to $u(b)$ and the same for everyone. The productivity characteristics of the worker x will have two components, one fully permanent x_0 and one subjected to individual shocks x_1. The quality of a job is captured by z. (x_0, x_1, z) are discretized log normal distributions with log-mean of 0 and log-variances $\sigma^2_{x_0}$,

[9] This is known as the recursive Lagrangian approach as developed by Kocherlakota (1996); Marcet and Marimon (2011); Messner, Pavoni, and Sleet (2012); Cole and Kubler (2012), Farhi and Werning (2013) uses promised marginal utility directly.

$\sigma^2_{x_1}$ and σ^2_z. The job productivity is redrawn from the marginal at quarterly Poisson rare λ_z. The x_1 component follows an auto-regressive process with auto-correlation parameter $1-\lambda_x$. The output of a match (x_0, x_1, z) is given by $f(x_0, x_1, z) = (x_0 \cdot x_1 \cdot z)^{1/2}$. The total factor is set to 1 and the mean value left unmatched as the model can be entirely rescaled. I specify the marginal cost of choosing quit probability $q \in [0, 1]$ as $c'(q) = \gamma_0 \cdot (1 - q^{-1/\gamma_1})$, and the condition that $c(1) = 0$ pins down the intercept. The model is estimated in partial equilibrium. Closing the equilibrium could be done in a second step, using the estimated job arrival function $p(x, V)$ to recover a matching function and the vacancy cost under additional assumptions. Here, I parametrize the job value offer curves for each (x_0, x_1) to be piece-wise linear with three sections:

$$p(x_0, x_1, V) = \begin{cases} 0 & \text{if } V \leq \underline{v}(x_0, x_1) \\ \frac{\overline{v}(x_0) - V}{\overline{v}(x_1) - \underline{v}(x_0, x_1)} & \text{if } \underline{v}(x_0, x_1) \leq V \leq \overline{v}(x_0, x_1) \\ 1 & \text{if } V \geq \overline{v}(x_0) \end{cases} \tag{2}$$

where $\underline{v}(x_0, x_1) = s_0 \frac{s_2 \overline{x}_0 - s_1 x_0}{\overline{x}_0 - \underline{x}_0} f(x_0, x_1, z_0)$ is the lowest offer any individual of type (x_0, x_1) will ever apply to and $\overline{v}(x_0) = \frac{s_2 \overline{x}_0 - s_1 x_0}{\overline{x}_0 - \underline{x}_0} f(x_0, \overline{x}_1, \overline{z})$ the highest. I set the initial match quality to its median $z_0 = 1$. The s_0, s_1, s_2 parameters control the quality of jobs offered to workers of type (x_0, x_1) and how likely they are to get each of them. I fixed $s_0 = 0.95$ and estimated s_1, s_2. I add two level of education to the model as observables and use it as conditioning variables when computing some of the moments. The probability of being of high education is specified as an increasing function of the permanent component of the type x_0. $Pr[Hed = 1 | x_0] \propto \Phi\Big(\alpha_0 + \alpha_1 x_0\Big)$ where α_1 controls how much permanent productivity is linked to education level. Finally, I will use a measure of output per worker at the firm level. To remain consistent with the theory, I use a constant return to scale aggregation and directly sum the match outputs. This specification leaves us with the following parameters to estimate:

$$\vartheta = \{\rho, b, \alpha_1, \alpha_2, \lambda_x, \lambda_z, \sigma_{x_1}, \sigma_{x_2}, \sigma_z, \gamma_0, \gamma_1, s_1, s_2\}.$$

4.2 Moments and estimates

Estimation is performed using simulated method of moments. The objective function is minimized over all parameters besides the education parameters (α_0, α_1) which are picked conditional on the others to match the education share and the education wage gap moments.

Within each education group, I use quarterly transition probabilities of starting a job (U2E), loosing a job (E2U) and changing jobs (J2J). Including these moments disciplines the parameters (s_1, s_2) of the search function. In addition the difference in the E2U rates between the two education group provides information about the level of the effort function γ_0 as well as its curvature γ_1, as it affects how much it responds to promised value, which is linked to wages. Secondly, the difference between U2E and J2J rates will also determine the value of being unemployed, since individuals without jobs will choose where to apply based on current present value.

The next set of moments includes information about the earnings of job stayers. Their role is to provide information about the production function and the sources of shocks. In particular, the wage growth variance of earnings should inform us about the total earnings shock each workers is facing. On the other hand, the co-variance between wage growth of co-workers should tell us about the risk shared at the firm level. This moments will help pin down

Table 3. Moments and within sample model fit

	data	model
Low education group moments		
$Pr[U2E\|Led]$	0.142	0.141
$Pr[E2U\|Led]$	0.025	0.027
$Pr[J2J\|Led]$	0.022	0.026
$Var(\log w_{it}\|EE, Led)$	0.137	0.154
$\mathbb{E}(\Delta \log w_{it}\|EE, Led)$	-0.002	0.022
High education group moments		
$Pr[U2E\|Hed]$	0.179	0.153
$Pr[E2U\|Hed]$	0.021	0.022
$Pr[J2J\|Hed]$	0.028	0.032
$Var(\log w_{it}\|EE, Hed)$	0.220	0.147
$\mathbb{E}(\Delta \log w_{it}\|EE, Hed)$	0.003	0.017
Common moments		
$Pr[Hed]$	0.299	0.298
$\mathbb{E}(\log w_{it}\|Hed) - \mathbb{E}(\log w_{it}\|Led)$	0.290	0.286
$Var(\Delta \log w^p_{it}\|EE)$	0.024	0.017
$Cov(\Delta \log w^p_{it}, \Delta \log w^p_{it-1}\|EE)$	0.000	0.004
$Cov(\Delta \log w_{it}, \Delta \log w_{jt}\|EE)$	0.00068	0.00058
$Var(\Delta \log y^p_{it}\|EE)$	0.026	0.041

Transition rates are quarterly using full sample, wage growth are year on year using stayers. *Led* is for high school graduates or less, *Hed* is for some college. See Table 1 for more info. Earnings are net of time and age effects. The variance and auto-covariance of wage growth are net a of classical measurement error.

λ_x and λ_z. Note that I choose to match the variance and auto-covariance of earning growth net of classical measurement error as measured by the model in Section 2 of the paper. This decision is motivated by the fact that the insurance contract will not generate any such classical measurement error. The variance within education group and the difference between the groups provide

Table 4. Parameter estimates

parameter		value	s.e.
CRRA parameter	ρ	1.687	8.20e-04
flow value of unemployment	b	0.889	2.33e-04
worker permanent heterogeneity	σ_{x_0}	1.035	2.89e-04
worker sthochastic heterogeneity	σ_{x_1}	2.292	1.26e-03
worker shock auto-correlation	λ_x	0.043	6.29e-05
job stochastic heterogeneity	σ_z	1.607	1.35e-03
job shock poisson probability	λ_z	0.028	6.65e-05
effort cost parameters	γ_0	1.044	1.55e-03
	γ_1	2.369	2.15e-03
job offer curve	s_1	0.386	7.28e-05
	s_2	0.256	1.61e-04
education parameters	α_0	2.785	
	α_1	5.000	

Standard errors are computed according to the asymptotic variance formula derived in the text. The scaling factor is the square root of the number of firms among stayers. The education parameters do not have standard errors because they are derived from the other parameters, the share of high education worker and their earning premium.

information on the variances of each of the terms of the production function $\sigma^2_{x_0}$, $\sigma^2_{x_1}$ and σ^2_z. Finally, the total variance of output create a constraints on how large or small the pass through can be, and so should inform us about the level of risk aversion. Since the model is strongly parametrized, I choose the weighting matrix to reflect how informative each moments should be about the parameters of interest. The default weight is chosen to be the inverse of the level in order to minimize a distance in relative deviation. Extra weight is also put on the variance and covariance terms of wage growths (see Appendix D.1).

Table 3 reveals that the transition rates are much lower than their counterpart in the US. The U2E values give a mean duration of unemployment of 21 months for the low education group and 17 months for the high education one. As described before, this is because we consider the entire population and not only the active job seekers. The fit of the model comes short in two dimensions which suggests that it might benefit from making the search function $p(x, V)$ more flexible. First we notice that the U2E probability is too low, and the J2J too high, particularly for the high education group. This might be due to the fact that I imposed $\kappa = 1$ in estimation. The literature tend to find search efficiency lower for employed worker which would move these fits in the right directions. Overall, the model explains a good share of the overall variation in the transition probabilities across education groups and type of events.

The variance of earnings is too high for the low education group and too small for the high education group. Matching both the mean difference in wages as well as the variance with group might require a distribution over permanent heterogeneity x_0 with a bigger tail than the log-normal.

Finally we look at the common parameters on wage growth that are informative about the productivity shocks in the data. The co-variance between wage growth of co-workers is well matched. The fit on the variance and auto-covariance of wage growth is equally good. Even though the role of insurance is central to the model, it is able to generate a large variance with a small auto-covariance of wage growth. This values suggests that the model provides a good fit of the earning process, including the risk shared at the firm level. Finally, the total variance of permanent value added growth is larger in the model which suggests that it understates the level of transmission of shocks, and compensates with higher productivity shocks. It is important to note that value added in the data might be a much nosier measurement of productivity than the way we generate it in the model.

The parameter values are presented in Table 4. The flow value of unemployment is $b = 0.889$ which represents about 20% of the mean value of $f(x_0, x_1, z_0)$ among the unemployed, which seems low when compared to the literature. The variances of the productivity components can be attributed directly if we ignore the selection on z. The idiosyncratic component in productivity are very large when compared to the permanent part x_0. They are also very persistent. The firm level shock z is shocked on average only every 9 years. The auto-correlation of x_1 is equally persistent with a quarterly auto-correlation of $1 - \lambda_x = 0.957$. The remainders of the parameters will be evaluated together in the following section.

5 Results analysis

Decomposing earnings variances. I now use the model to decompose the variance of earnings and earnings growth. The model gives us the ability to decompose this moments into structural components.

Table 5 reports the decomposition of the cross-sectional variance of log-output and log-wages. I regress each outcome on additive dummies for each of the levels of x_0, x_1, z. Given that we modeled productivity as a log-normal

Table 5. Cross-sectional variance decompositions

	total	$Var[x_0]$	$Var[x_1]$	$Var[z]$	residuals
$\log(output)$	0.833	0.081 (9.7%)	0.641 (76.9%)	0.117 (14.0%)	0.000 (0.0%)
$\log(wage)$	0.169	0.085 (49.9%)	0.003 (1.7%)	0.012 (7.0%)	0.073 (42.9%)

Using data simulated from the model, I run a linear regression with dummies for each levels of each of x_0, x_1, z. I then report the variance associated with each term. The percentages are shares of variances.

Table 6. Cross-sectional growth variance decompositions

	total	U2E/E2U	J2J	x	z
$\Delta\log(output)$	0.707	0.173 (6.0%)	0.077 (1.2%)	0.643 (82.8%)	0.223 (10.0%)
$\Delta\log(wage)$	0.210	0.194 (85.4%)	0.024 (1.3%)	0.060 (8.1%)	0.048 (5.2%)

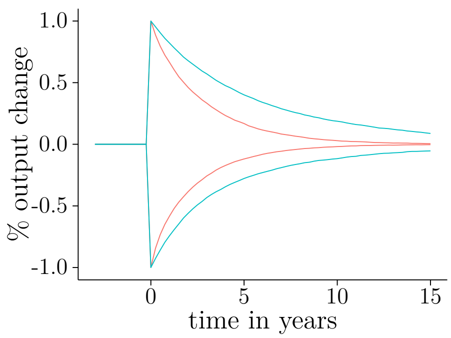
Using data simulated from the model, I sequential shut down channels in the simulation, while keeping the save agents policies. I first remove transitions in and out of unemployment, then job to job transitions, and finally the x_1 and z shocks. At each step I keep the cross-sectional productivity distribution fixed, and compute earnings growth variances.

we get perfect fit for output. We also note that the covariance between the different productivities play a small role in the total contribution to log-output variance. The results for output suggest that the main source of output variance is associated with the stochastic worker heterogeneity x_1, which explains 76.9% of the variance. I run a similar linear projection for log-wages. The decomposition is quite different. Now the permanent heterogeneity x_0 is responsible for most of the log wage variance with 49.9% of the contribution. The second larger share is in the residual with 42.9%. This residual can be interpreted as the dispersion created by the path dependency of the contract. The contracting agreement does not seem to absorb any of the permanent heterogeneity since the variances associated with x_0 are very similar in both output and wages. This numbers are very different for x_1 and z. Part of their contribution might show up in the residuals, yet the total variance is very much reduce. This shows the presence of important insurance provision from the employer.

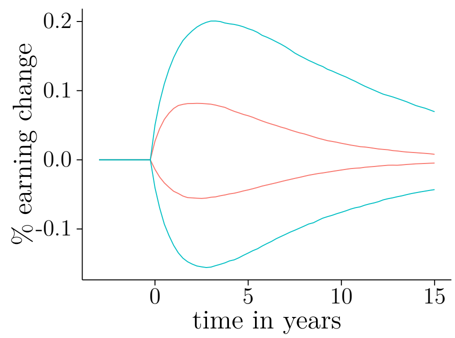
To understand further, I decompose the output and earnings growth year on year using simulated data. The decomposition of the variance is this case is not linear. To quantify the contribution of the different feature of the model, I choose to shut down sequentially the shocks and consider the long run cross-section. Unfortunately this means that the order matters. For mobility shocks, I follow the structure of the model and first shut down transitions in and out of employment, followed by job to job transitions. For the shocks to x_1 and z I report the mean between removing x_1 shocks first or removing z shocks first. One further difficulty is that removing certain types of transitions changes the productivity distributions (for example, J2J transitions allow to hop to better matches). To address that concern I rescale the cross-sectional distribution over x_1, z to match the original one. The values should then be interpreted as the variance if each worker in the current cross-section had been at his current productivity for a long period of time, washing out transitory effect due to the smoothing of the contract, and moving forward from that point with remaining shocks.

Table 6 shows the results of these decompositions. As in the level decomposition, we see that x_1 captures a large share of the productivity shocks at 82.8%. If we then look at the wage growth variance, we note that 85.4% is generated by transitory dynamics associated with transition in and out of employment. Once we remove these, the larger share is associated with x_1 at 8.1%, then z shocks at 5.2% and finally only 1.3% associated with job to job transitions. This results suggests that the dynamics of wages associated with coming in and out of jobs are an important source of the variance of earnings that we see in the data. Beyond these, individuals and firm specific shocks appear to have similar magnitudes.

Output and earnings effect of x and z shocks. I report the impulse response in this model to innovation shocks to z and x, both positive and negative. I generate a treatment group that replicates the simulated population



(a) x and z shocks on output



(b) x and z shocks on earnings

Figure 3: Average impulse repsonse of x and z innovation shocks.
Effects of a 100% increase in either x or z, either positive or negative at parameter estimates. The red line is an x shock and the blue line is the z shock.

Table 7. Impulse response analysis to x and z shocks

	output				earnings			
	max value	max time	half life	EPV	max value	max time	half life	EPV
x positive	+100.0%	0.0y	1.8y	+18.7%	+8.2%	2.2y	7.5y	+3.6%
x negative	-100.0%	0.0y	1.3y	-10.9%	-5.6%	2.5y	7.3y	-2.3%
z positive	+100.0%	0.0y	3.7y	+33.9%	+20.1%	3.2y	11.8y	+12.4%
z negative	-100.0%	0.0y	2.4y	-20.4%	-15.6%	2.8y	9.8y	-8.4%

Effects of a 100% increase in either x or z, either positive or negative at parameter estimates. This is a companion table to Figure 3. It gives for output and earnings, the percentage change at the highest change, how long after the shock the max is reached, the duration of half life and the percentage change in the present discounted sum of the flows.

and which receives at time 0 a fix percentage shock to either z or x, either positive or negative. I report in Figure 3 the percentage change in output and earnings in response to these shocks. I compute the percentage change between the control and the treatment at each time, including zeros for unemployed workers. The shocks are scaled to generate a 100% output change on impact.

In Figure 3, we see the plot in the percentage change in output and in earnings over time. The plot shows how the response in earnings is delayed when compared to the output response. This is due to the smoothing of the contract as well as the employment response in the treatment group. Table 7 shows that the highest earning point is reached 2 to 3 years after the productivity impact. The size of the response, at the highest point varies between 5% and 20% of the equivalent productivity shock. The z shock appears to have a larger impact. I did control for the size at time 0, however the stronger persistence of the z shock means that the long term effect can be larger just for this reason. However when comparing the change in expected present values (EPV) between output and earnings, we see that the z also has a stronger pass-through than x. When comparing negative to positive shocks, we see that positive shocks tend to have longer lasting and stronger overall impacts. There are two reasons for this. First negative productivity shocks gets selected out as lower wage workers move out of unemployment and move to better matches. Second the contract itself, in the spirit of Harris and Holmstrom (1982) generates some downward rigidities. To conclude the transmission effects of productivity shocks are delayed, and we see an important level of insurance provided by the firm, where only between 5% and 20% of the shock are pass-through overall.

5.1 Policy evaluation

I analyze the effect of a revenue neutral government policy that redistributes from high wages to lower wages. I parametrize the policy as $\tilde{w} = \lambda w^{\frac{1}{\tau}}$. I pick a set of values for λ and solve for τ so that the policy is revenue neutral in the estimated model. The goal of such policy is to affect the level of dispersion and the amount of shocks faced by individuals. Firms however are free to adjust contracts in the presence of this government transfers. For each policy I compute the effect on the cross-sectional variance of earnings and on the

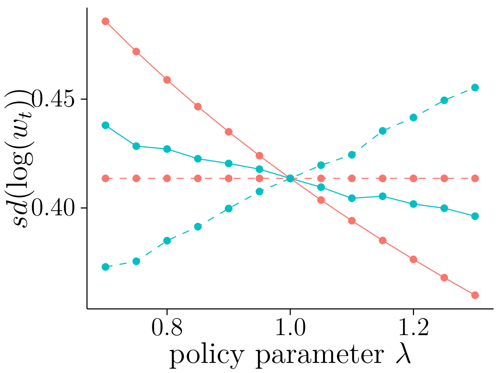
Table 8. Effect of revenue neutral policies

		keep contracts fixed		let contracts respond	
		gross	net	gross	net
Policy 1	$sd[\log(w)]$	0.41	0.38 (-18.9%)	0.44 (+13.1%)	0.40 (-5.8%)
(λ=1.2, τ=1.1)	$sd[\Delta \log(w)]$	0.21	0.19 (-18.9%)	0.22 (+13.3%)	0.20 (-5.6%)
Policy 2	$sd[\log(w)]$	0.41	0.46 (+20.8%)	0.38 (-14.3%)	0.43 (+6.4%)
(λ=0.8, τ=0.9)	$sd[\Delta \log(w)]$	0.21	0.23 (+20.8%)	0.19 (-14.8%)	0.22 (+5.9%)

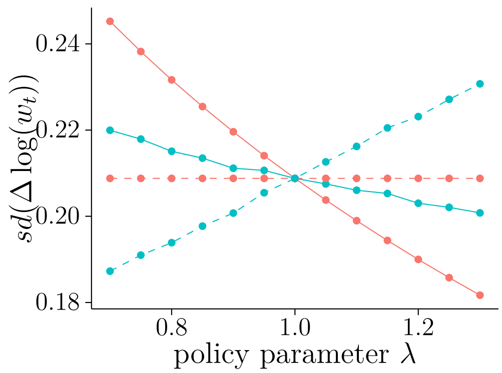
Policies of the form $\tilde{w} = \lambda w^{\frac{1}{\tau}}$ are applied to the model. For each policy the effect on log wages and log wage growth are reported. The first set of columns applies the transfer without letting individual responding. The last set of column presents results for the contract solved when all agent expects the transfers to happen. Percentage changes are in variances and relative to the first column (the estimated model). $\Delta \log(w)$ includes J2J transitions, hence is bigger than moments in Table 3.

variance of earnings growth. To make the effect clear, I report four numbers at each policy in Table 8: the model solved at the estimated parameters without any transfer (first column), use the same solution and apply transfers from the given policy without adjusting decisions (column 2). Then solve the model again with agents knowing about the transfers, and report pre-transfer (column 3) and post-transfer values (column 4).

The goal of policy 1 with parameters ($\lambda = 0.8, \tau = 0.9$) is to reduce earnings inequality and earnings growth shocks. Indeed applying directly the transfers to the population reduces the variance and growth variances by 18.9%. However when we let agent re-optimize their contracts, we see that pre-transfer variance increases by 13.3%. Understanding that the government is providing more insurance, firms choose to pass on more of the productivity risk to their workers. The result is that post transfer variances and wage growth variances are reduced by only 5.6%. We see very clearly that insurance provided by firms is crowded out by insurance provided by the government. Around 2/3 of the direct effect of transfer is undone by firm adjusting the offered contracts. We see a similar effect for the policy 2, which we could think of as reducing



(a) z shock on log-output



(b) z shock on log-earnings

Figure 4: Policy effect at different values of λ.

For each value, τ is solved for to generate 0 revenue. Dotted lines represent pre-transfer measures, and plain lines are post-transfer measures. The red line is for agent policy oblivious of the transfer, the blue line for agents aware of the transfers.

current government transfers.

Figure 4 plots the four different values of the columns of Table 8 for a range of policies. Comparing the plain lines which are post-transfer values, we see that agents that ignore the presence of transfers receive the strongest effect on both measures. The blue lines are the values of for agents who do take the transfers into account when designing the contracts. Looking to the right of $\lambda = 1$, we see that these agents increase variances pre-transfer (the dotted blue line rises above the red-dotted), which results in an attenuated effect of the transfer policy (the plain blue line ends in between the dotted and plain red lines). Approximately two third of the effect is undone by changes in the contract across values of λ. This exercise emphasize the importance of considering how firm might choose to pass on risk.

6 Conclusion

In this paper I study the different sources of uncertainty faced by workers in the labor market. Workers are subject to individual productivity shocks and their earnings may also be affected by the performance of their employer because of search frictions in the labor market. To understand the way shocks get transmitted and how this might affect welfare and labor market policy I develop an equilibrium model with search frictions, risk averse workers, firm and worker productivity shocks. In this model I show that the optimal contract pays a wage that smoothly tracks the joint match productivity. This implies that both worker and firm level shocks transmit to wages, albeit only partially. In contrast to the perfectly competitive model, on one hand firm may insure workers' productivity shocks but on the other hand they are able to transmit firm level shocks to wages.

I then use matched employer employee data to learn about the size of these different shocks. I first use a very simple model to provide evidence of the presence of shocks at the level of the firm. In the fourth part of the paper and in the Appendix, I develop the econometric properties of the model and I estimate a parametrized version using simulated method of moments.

Future work should look at relaxing some of the constraints I imposed in this paper, both in terms of contractual environment and and in terms of specification restrictions. An important extension to this model is to allow individuals to hold assets, which would allow them to self insure. The inclusion of observable assets would depart only slightly from the current version of the model but a more realistic environment would allow workers to save privately.

Finally, I hope that this insurance framework will be useful in other fields where relational long-term contracts and provision of insurance, together with search frictions are important. Relational banking, insurance markets and buyer-seller repeated relationships seem very natural candidates.

References

Abowd, J. M., and D. Card (1986): "On the covariance structure of earnings and hours changes," *Econometrica*.

Abraham, A., F. Alvarez-Parra, and S. Forstner (2016): "The Effects of Moral Hazard on Wage Inequality in a Frictional Labor Market," .

Abraham, A., and N. Pavoni (2008): "Efficient allocations with moral hazard and hidden borrowing and lending: A recursive formulation," *Rev. Econ. Dyn.*, 11(4), 781–803.

Abreu, D., D. Pearce, and E. Stacchetti (1986): "Optimal cartel equilibria with imperfect monitoring," *J. Econ. Theory*, 39(1), 251–269.

Allman, E. S., C. Matias, and J. A. Rhodes (2009): "Identifiability of parameters in latent structure models with many observed variables," *Ann. Stat.*, 37(6A), 3099–3132.

Altonji, J. G., A. Smith, and I. Vidangos (2009): "Modeling earnings dynamics," *NBER Work. Pap. Ser.*

An, Y., Y. Hu, J. Hopkins, and M. Shum (2013): "Identifiability and inference of hidden Markov models," *Working Paper*.

Attanasio, O. P., and N. Pavoni (2011): "Risk Sharing in Private Information Models With Asset Accumulation: Explaining the Excess Smoothness of Consumption," *Econometrica*, 79(4), 1027–1068.

Azariadis, C. (1975): "Implicit Contracts and Underemployment Equilibria," *J. Polit. Econ.*, 83(6), 1183–1202.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014): "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *Am. Econ. Rev.*, 104(6).

Baily, M. N. (1974): "Wages and employment under uncertain demand," *Rev. Econ. Stud.*, 41(1), 37–50.

Benveniste, L. M., and J. A. Scheinkman (1979): "On the Differentiability of the Value Function in Dynamic Models of Economics," *Econometrica*, 47(3), 727–732.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption Inequality and Partial Insurance," *Am. Econ. Rev.*, 98(5), 1887–1921.

Boualam, Y. (2015): "Bank Lending and Relationship Capital," Discussion paper, Working paper at University of Pennsylvania, last accessed 24th February 2015, http://economics. yale. edu/sites/default/files/yasser_boualam_-_jmp_-_nov2014. pdf.

Burdett, K., and M. Coles (2003a): "Equilibrium Wage-Tenure Contracts," *Econometrica*, 71(5), 1377–1404.

Burdett, K., and M. G. Coles (2003b): "Wage/tenure contracts and equilibrium search," *Econometrica*, 71(5), 1377–1404.

Burdett, K., S. Shi, and R. Wright (2001): "Pricing and Matching with Frictions," *J. Polit. Econ.*, 109(5), 1060–1085.

Carlsson, M., J. Messina, and O. Nordström Skans (2014): "Firm-level shocks and labor adjustments," Discussion Paper 2014:28.

Chiappori, P., B. Salanie, and J. Valentin (1999): "Early Starters versus Late Beginners," *J. Polit. Econ.*, 107(4), 731–760.

Cole, H., and F. Kubler (2012): "Recursive contracts, lotteries and weakly concave Pareto sets," *Rev. Econ. Dyn.*, 15(4), 479–500.

Connault, B. (2015): "Hidden Rust Models," .

Dardanoni, V. (1995): "Income distribution dynamics: monotone Markov chains make light work," *Soc. Choice Welfare*, 12(2), 181–192.

Doepke, M., and R. M. Townsend (2006): "Dynamic mechanism design with hidden income and hidden actions," *J. Econ. Theory*, 126(1), 235–285.

Farhi, E., and I. Werning (2013): "Insurance and Taxation over the Life Cycle," *Rev. Econ. Stud.*, 80(2), 596–635.

Fredriksson, P., and R. H. Topel (2010): "Wage Determination and Employment in Sweden Since the Early 1990s: Wage Formation in a New Setting," in *Reforming the Welfare State: Recovery and Beyond in Sweden*, pp. 83–126. University of Chicago Press.

Friedrich, B., L. Laun, C. Meghir, and L. Pistaferri (2014): "Earnings Dynamics and Firm-Level Shocks," .

Guiso, L., L. Pistaferri, and F. Schivardi (2005): "Insurance within the Firm," *J. Polit. Econ.*, 113(5), 1054–1087.

Hall, P., and X.-H. Zhou (2003): "Nonparametric Estimation of Component Distributions in a Multivariate Mixture," *Ann. Stat.*, 31(1), 201–224.

Hall, R. E., and F. S. Mishkin (1980): "The sensitivity of consumption to transitory income: estimates from panel data on households," .

Harris, M., and B. Holmstrom (1982): "A Theory of Wage Dynamics," *Rev. Econ. Stud.*, 49(3), 315–333.

Hendel, I., and A. Lizzeri (2000): "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance," .

Holmstrom, B., and P. Milgrom (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization*, 7, 24–52.

Hu, Y. (2008): "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *J. Econom.*, 144(1), 27–61.

Hu, Y., and M. Shum (2012): "Nonparametric identification of dynamic models with unobserved state variables," *J. Econom.*, 171(1), 32–44.

Knight, F. (1921): "H., 1921. Risk, uncertainty and profit," *Houghton Mifflin, Boston.*

Kocherlakota, N. R. (1996): "Implications of Efficient Risk Sharing without Commitment," *Rev. Econ. Stud.*, 63(4), 595–609.

Koeppl, T. (2006): "Differentiability of the efficient frontier when commitment to risk sharing is limited," *Topics in Macroeconomics*, 6(1), 1–6.

Lemieux, T., L. Thomas, W. Bentley MacLeod, and P. Daniel (2009): "Performance Pay and Wage Inequality," *Q. J. Econ.*, 124(1), 1–49.

Lentz, R. (2016): "Optimal Wage-Tenure Contracts without Search Intensity Commitment," *Working Paper*.

Lise, J., C. Meghir, and J. M. Robin (2008): "Matching, sorting and wages," *Unpublished Manuscript*.

Low, H., C. Meghir, and L. Pistaferri (2010): "Wage Risk and Employment Risk over the Life Cycle," *Am. Econ. Rev.*, 100(4), 1432–1467.

MacLeod, W. B., and J. M. Malcomson (1989): "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57(2), 447–480.

MaCurdy, T. E. (1982): "The use of time series processes to model the error structure of earnings in a longitudinal data analysis," *J. Econom.*, 18(1), 83–114.

Marcet, A., and R. Marimon (2011): "Recursive contracts," *Unpublished Manuscript*.

Meghir, C., and L. Pistaferri (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72(1), 1–32.

Meghir, C., and L. Pistaferri (2011): "Earnings, consumption and life cycle choices," *Handbook of Labor Economics*.

Menzio, G., and S. Shi (2009): "Efficient search on the job and the business cycle," *Unpublished Manuscript*.

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

Messner, M., N. Pavoni, and C. Sleet (2012): "Recursive methods for incentive problems," *Rev. Econ. Dyn.*, 15(4), 501–525.

Moen, E. R. (1997): "Competitive Search Equilibrium," *J. Polit. Econ.*, 105(2), 385–411.

Montgomery, J. D. (1991): "Equilibrium Wage Dispersion and Interindustry Wage Differentials," *Q. J. Econ.*, 106(1), 163–179.

Peters, M. (1991): "Ex Ante Price Offers in Matching Games Non-Steady States," *Econometrica*, 59(5), 1425–1454.

Postel-Vinay, F., and J.-M. Robin (2004): "To match or not to match?: Optimal wage policy with endogenous worker search intensity," *Rev. Econ. Dyn.*, 7(2), 297–330.

Rogerson, W. P. (1985): "The first-order approach to principal-agent problems," *Econometrica*, pp. 1357–1367.

Roys, N. (2011): "Persistence of shocks, employment and wages," *Unpublished Manuscript*.

Rudanko, L. (2009): "Labor market dynamics under long-term wage contracting," *J. Monet. Econ.*, 56(2), 170–183.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033.

Schaal, E. (2010): "Uncertainty, Productivity and Unemployment during the Great Recession," *unpublished, December*.

Shi, S. (2008): "Directed Search for Equilibrium Wage-Tenure Contracts." forthcoming in," *Econometrica*, 77(2), 561–584.

Shimer, R. (1996): "Contracts in a frictional labor market," Discussion paper, mimeo.

Spear, S. E., and S. Srivastava (1987): "On Repeated Moral Hazard with Discounting," *Rev. Econ. Stud.*, 54(4), 599–617.

Stevens, M. (2004): "Wage-tenure contracts in a frictional labour market: Firms' strategies for recruitment and retention," *Rev. Econ. Stud.*, 71(2), 535–551.

Thomas, J., and T. Worrall (1988): "Self-enforcing wage contracts," *Rev. Econ. Stud.*, 55(4), 541–554.

Tsuyuhara, K. (2013): "Dynamic Contracts with Worker Mobility via Directed On-the-Job Search," *Working Paper*.

A Model appendices

A.1 Properties of worker search functions

Lemma 2. *Given* (x, W), $v^*(x, W)$ *and* $e^*(W)$ *are uniquely determined,* $\tilde{p}(x, W)$ *is continuous and decreasing,* $\tilde{r}(x, W)$ *is increasing in* W*, continuously differentiable and* $\frac{\partial \tilde{r}}{\partial W}(x, W) = \beta \tilde{p}(x, W)$.

Proof. We start with the definitions:

$$
\begin{aligned}
v^*(x, W) =& \arg\max_v p(\theta(x, v))(v - W) \\
e^*(x, W) =& \arg\max_e -c(e) + \delta(e)\beta \mathbb{E} W_0(x') \\
& + \beta(1 - \delta(e)) \left(p(\theta(x, v^*))v^* + \beta(1 - \delta(e))(1 - p(\theta(x, v^*)))W\right),
\end{aligned}
$$

and the definition of the composite functions

$$
\begin{aligned}
\tilde{p}(x, W) =& (1 - \delta(e^*(x, W))) \left(1 - p(\theta(x, v_1^*(x, W)))\right) \\
\tilde{r}(x, W) =& -c(e^*(x, W)) + \beta(1 - \delta(e^*(x, W)))p(\theta(x, v_1^*(x, W))) \left(v_1^*(x, W) - W\right) \\
& + \delta(e^*(x, W))\beta \mathbb{E}_{x'|x} U(x') + \beta(1 - \delta(e^*(x, W)))(x, W)W,
\end{aligned}
$$

I first normalize $\delta(e) = 1 - e$ (or equivalently redefine c and e such that $c(e) = c(\delta^{-1}(e)))$, where $c(e)$ is increasing and concave. The maximization problem for v gives the following first order condition

$$
p'(\theta(x, v))(v - W) + p(\theta(x, v)) = 0,
$$

where given the property of p and q and the equilibrium definition of θ we have that the function $v \mapsto p(\theta(x, v))$ is decreasing and strictly concave. This gives that the maximum is unique and so $v^*(x, W)$ is uniquely defined. The first order condition for e is given by

$$c'(e) = \beta p(\theta(x, v_1^*(x,W)))\,(v_1^*(x,W) - W) + \beta W - \beta \mathbb{E}_{x'|x} U(x'),$$

and given the assumption that c is strictly convex, we get that $e^*(x,W)$ is also uniquely defined.

Finally we can use the envelope condition to compute the derivative of $\tilde{r}$ with respect to W. By definition we have

$$\tilde{r}(x,W) = \sup_{v,e} u(w) - c(e) + (1-e)\beta \mathbb{E}_{x'|x} W_0(x') + e\beta p(\theta(x,v))v + e\beta(1-p(\theta(x,v)))W,$$

and so we get

$$\frac{\partial \tilde{r}}{\partial W}(x,W) = \beta e^*(x,W)(1 - p(\theta(x, v^*(x,W)))) = \beta \tilde{p}(x,W),$$

which proves that $\tilde{r}$ is continuously differentiable as long as $\tilde{p}$ is continuous. □

A.2 Regularity properties for equilibrium functions

Lemma 1. *The Pareto frontier $\mathcal{J}(x,z,V)$ is continuously differentiable, decreasing and concave with respect to V and increasing in z.*

Proof of Lemma 1 . Consider the optimal contract equation:

$$\begin{aligned}
\mathcal{J}(x,z,V) = \sup_{\pi_i, W_i, W_{ix'y'}} & \sum \pi_i \Big(f(x,z) - w_i + \beta \tilde{p}(x,W_i) \mathbb{E} \mathcal{J}(x',z',W_{ix'y'}) \Big) \\
s.t \quad (\lambda) \quad & 0 = \sum_i \pi_i \left(u(w_i) + \tilde{r}(x,W_i) \right) - V, \\
(\gamma_i) \quad & 0 = W_i - \mathbb{E} W_{ix'y'}, \\
& \sum \pi_i = 1.
\end{aligned}$$

We already know that $\mathcal{J}$ is concave because of the two point lottery. That tells us that it is continuous and differentiable almost everywhere. Let's then show that it is differentiable everywhere. I follow the steps of the derivation presented in Koeppl (2006) where he shows that in the problem with two sided limited commitment it is sufficient to have one state realization where neither participation constraint binds to achieve differentiability of the Pareto frontier. Given that the current problem is one sided the result works almost right away, it just needs to be extended to include a search decision.

For a fixed $s = (x, z)$, let's consider a point $\tilde{V}$ where it's not differentiable and call $(\tilde{w}, \tilde{\pi}_1, \tilde{W}_{ix'z'}, \tilde{W}_i)$ the firm's action at that point. This action is by definition feasible and delivers $\tilde{v}$ to the worker. From that strategy I am going to construct a continuum that delivers any V around $\tilde{V}$. Keeping $(\tilde{\pi}_1, \tilde{W}_{ix'z'}, \tilde{W}_i)$ the same, I defined $w^*(V) = u^{-1}(V - \tilde{V})$.

I then define the function $\tilde{\mathcal{J}}(s, v)$ as the value that uses strategy $\left(w^*(V) = u^{-1}(V - \tilde{V}), \tilde{\pi}_1, \tilde{W}_{ix'y'}, \tilde{W}_i\right)$. It is the case that the strategy is feasible since all constraints remain satisfied. By definition of $\mathcal{J}$ we have that $\tilde{\mathcal{J}}(s, V) \leq \mathcal{J}(s, V)$ together with $\tilde{\mathcal{J}}(s, \tilde{V}) = J(s, \tilde{V})$. Finally because $u(\cdot)$ is concave, increasing and twice differentiable, $\tilde{\mathcal{J}}(s, \tilde{V})$ is also concave and twice differentiable.

We found a function concave, continuously differentiable, lower than $\mathcal{J}$ and equal to $\mathcal{J}$ at $\tilde{V}$ we can apply Lemma 1 from Benveniste and Scheinkman (1979) which gives us that $\mathcal{J}(s, v)$ is differentiable at $\tilde{v}$. We then conclude that $\mathcal{J}$ is differentiable everywhere.

Monotonicity. Let's show that $\mathcal{J}(x, z, v)$ is increasing in z. The intuition here is to show that the firm starting at (x, z_2) can mimic the strategy starting at (x, z_1) and still have some output left on the table when compared with starting at z_1, this will be our J_0 value later in this proof. We then show that this strategy is feasible, and has to deliver less than the very best strategy, the one implemented by $\mathcal{J}(x, z_2, v)$.

More precisely, let's consider two different values $z_1 < z_2$. Let's take the optimal contract ξ_1, the history contingent policy starting at (x, z_1, v). We have that:

$$\mathcal{J}(x, z_1, v) = \sum_{t=0}^{\infty} \sum_{h^t \in H_t(v_1)} \beta^t \Big(f(x_t, z_t) - w^t(h^t; \xi_1) \Big) \pi^t(h^t; \xi_1),$$

where I denoted $h^t = (x_1..x_t, z_1...z_t)$ and where $w^t(h^t; \xi_1)$ is the wage paid by ξ_1 at history h^t, and $\pi^t(h^t; \xi_1) = \prod_t \tilde{p}(x_t, v_1(t, \xi_1))$ is the composition of all separation probabilities on the path.

We then change the indexing from t, h^t to the realization $\omega \in [0, 1]$ in the probability space and order the histories by lexico-graphic order (a history is bigger based on comparing the first time x differ, and only if all x are different, comparing the first z where they differ). We directly sum across ω to get:

$$\mathcal{J}(x, z_1, v) = \int \sum_{t=0}^{\infty} \beta^t \Big(f(x(t; \omega), z_1(t; \omega)) - w_1(t; \omega) \Big) \pi_1(t; \omega) \, \mathrm{d}\omega,$$

We then perform a similar ordering of the histories starting from $z_2 \geq z_1$. Because of independence between z and x, we get that $x(t, \omega)$ is the same when starting from (x, z_2). However we get a different $z_2(t, \omega)$. Because the transition Matrix on z is assumed to be monotonic, we get that $\forall \omega, t$: $z_2(t, \omega) \geq z_1(t, \omega)$.

We then consider the following value:

$$J_0 = \int \sum_{t=0}^{\infty} \beta^t \Big(f(x(t; \omega), z_2(t; \omega)) - w_1(t; \omega) \Big) \pi_1(t; \omega) \, \mathrm{d}\omega,$$

and note that this will deliver the same v value to the worker since all wages, all x and all transitions are identical to ξ_1. This gives us a contract that starts at x, z_2 and delivers v, we then know that it has to be at most equal to the value of the optimal contract and so $J_0 \leq \mathcal{J}(x, z_2, v)$.

Second, we did construct the histories such that $\forall \omega, t, z_2(t, \omega) \geq z_1(t, \omega)$. This means that

$$\begin{aligned} J_0 &= \int \sum_{t=0}^{\infty} \beta^t \Big(f(x(t;\omega), z_2(t;\omega)) - w_1(t;\omega) \Big) \pi_1(t;\omega) \, \mathrm{d}\omega \\ &\geq \int \sum_{t=0}^{\infty} \beta^t \Big(f(x(t;\omega), z_1(t;\omega)) - w_1(t;\omega) \Big) \pi_1(t;\omega) \, \mathrm{d}\omega \\ &\geq \mathcal{J}(x, z_1, v), \end{aligned}$$

which gives the result. See Dardanoni (1995) for more on properties of monotonic Markov chains. □

A.3 Characterization of the optimal contract

Lemma 3. *For a given (x, z), a higher wage always means higher lifetime utility.*

Proof. This is a direct implication of the concavity of $\mathcal{J}$ and the envelope condition:

$$\frac{\partial \mathcal{J}(x, z, v)}{\partial v} = \frac{1}{u'(w)},$$

and given also the concavity of $u(\cdot)$, we get that w and v_s are always moving in the same direction. □

Proposition 1. *For any current state (x_t, z_t, w_t), within each lottery realization i, the following relationship between wage growth and expected firm profit holds:*

$$\eta(x_t, W_{it}) \cdot \mathbb{E}_t \mathcal{J}_{i,t+1} = \frac{1}{u'(w_{i,t+1})} - \frac{1}{u'(w_t)}, \qquad \text{(EQ-FOC)}$$

where $\eta(x,W) = \frac{\partial}{\partial W}\log\tilde{p}(x,W) \geq 0$ *is the derivative of the log-probability that the relationship continues into the next period with respect to the value promised to the worker,* $\mathbb{E}_t\mathcal{J}_{i,t+1} = \mathbb{E}\mathcal{J}(x_{t+1}, z_{t+1}, W_{ix_{t+1}z_{t+1}})$ *is the expected profit for the firm next period and* $w_{i,t+1}$ *is the wage the firm will pay to the worker next period.*

Proof of Proposition 1 . We start again from the list of first order conditions and we want to find a relationship for wage change.

$$\mathcal{J}(x,z,V) = \sup_{\pi_i, W_i, W_{ix'z'}} \sum \pi_i \Big(f(x,z) - w_i + \beta\tilde{p}(x,W_i)\mathbb{E}\mathcal{J}(x',z',W_{ix'z'}) \Big)$$

$$s.t \quad (\lambda) \quad 0 = \sum_i \pi_i \left(u(w_i) + \tilde{r}(x,W_i)\right) - V,$$

$$(\gamma_i) \quad 0 = W_i - \mathbb{E}W_{ix'z'},$$

$$\sum \pi_i = 1.$$

From the envelope theorem and the f.o.c. for the wage, we get that the wage in the current period is given by

$$i = 1,2 \quad u'(w_i) = \frac{1}{\lambda} = -\left(\frac{\partial \mathcal{J}}{\partial v}(x,z,v)\right)^{-1}.$$

Now that also means that the wage next period in state (x',z') will be given by

$$\frac{1}{u'(w_{ix'z'})} = -\frac{\partial \mathcal{J}}{\partial v}(x',z',W_{ix'z'}).$$

I then look at the first order condition with respect to W_i

$$\pi_i\beta\tilde{p}_v(x,W_i)\mathbb{E}\mathcal{J}(x',z',W_{ix'y'}) + \beta\lambda\pi_i r'(x,W_i) + \pi_i\gamma_i = 0,$$

where I substitute $r'(x,W) = \tilde{p}(x,W)$, derived in Lemma (A.1):

$$\pi_i\beta\tilde{p}_v(x,W_i)\mathbb{E}\mathcal{J}(x',z',W_{ix'y'}) + \beta\lambda\pi_i\tilde{p}(x,W_i) + \pi_i\gamma_i = 0.$$

Using the f.o.c. for $W_{ix'z'}$, which is

$$\beta\tilde{p}(x, W_i)\frac{\partial \mathcal{J}}{\partial v}(x', z', W_{ix'y'}) - \gamma_i = 0, \tag{3}$$

I get the following expression:

$$\pi_i\beta\tilde{p}_v(x, W_i)\mathbb{E}\mathcal{J}(x', z', W_{ix'z'}) + \beta\lambda\pi_i\tilde{p}(x, W_i) + \pi_i\beta\tilde{p}(x, W_i)\frac{\partial \mathcal{J}}{\partial v}(x', z', W_{ix'z'}) = 0.$$

Focusing on $p_1(x, W) > 0$ and $\pi_i > 0$ since otherwise, the worker is leaving the current firm and the next period wage is irrelevant, we first rewrite:

$$\frac{\tilde{p}_v(x, W_i)}{\tilde{p}(x, W_i)}\mathbb{E}\mathcal{J}(x', z', W_{ix'z'}) + \lambda + \frac{\partial J}{\partial v}(s', v_{s'}) = 0.$$

I finally use the envelope condition to extract the wage next period from the last term on the right

$$\frac{\tilde{p}_v(x, W_i)}{\tilde{p}(x, W_i)}\mathbb{E}\mathcal{J}(x', z', W_{ix'z'}) = \frac{1}{u'(w_{x'z'})} - \frac{1}{u'(w)},$$

where since $\tilde{p}_v(x, W_i) > 0$ the inverse marginal utility and consequently wages move according to the sign of expected surplus to the firm. This shows that within each realization of the lottery, the wage will move according to expected profit. □

Corollary 1 (optimal contract)**.** *For each viable match (x, z), independent of the lottery realization, the wage policy is characterized by a **target wage** $w^*(x, z)$, which is increasing in z such that:*

$$w_t \le w^*(x_t, z_t) \quad \Rightarrow \quad w_t \le w_{t+1} \le w^*(x_t, z_t) \quad \textit{incentive to search less}$$

$$w_t \ge w^*(x_t, z_t) \quad \Rightarrow \quad w^*(x_t, z_t) \le w_{t+1} \le w_t \quad \textit{incentive to search more}$$

where the target wage is characterized by the zero expected profit condition for *the firm:*

$$\forall x, z \quad \mathbb{E}_{x'z'|xz}\mathcal{J}(x', z', W_{x'z'}) = 0.$$

Proof of Corollary 1 . We start from the first oder condition that was established in the previous proof. We then establish two additional results, first that there won't be randomization over wage increase and decrease. Second we establish the monotonicity of the reference wage.

Randomizing over increase and decrease: let's check if it is ever optimal for the firm to randomize over a wage increase and a wage decrease at the same time. If the lottery is degenerate then the result holds directly. We are left with non-degenerate lotteries. In that case the first order condition with respect to π must be equal to zero (otherwise we are at a corner solution, which is degenerate). Taking the first order condition with respect to π gives:

$$\beta\tilde{p}(x, W_1)\mathbb{E}\mathcal{J}(x', z', W_{1x'z'}) + \lambda\beta\tilde{r}(x, W_1) = \\ \beta\tilde{p}(x, W_2)\mathbb{E}\mathcal{J}(x', z', W_{2x'z'}) + \lambda\beta\tilde{r}(x, W_2),$$

which we can reorder in

$$\beta\tilde{p}(x, W_1)\mathbb{E}\mathcal{J}(x', z', W_{1x'z'}) - \beta\tilde{p}(x, W_2)\mathbb{E}\mathcal{J}(x', z', W_{2x'yz'}) = \\ \lambda\beta\left(\tilde{r}(x, W_2) - \tilde{r}(x, W_1)\right).$$

Now, suppose that the randomization is over two expected profits of opposite sign for the firm where 1 is positive and 2 is negative. The left hand side is then positive. But in that case we know that $W_2 < V < W_1$ because higher wages give higher utilities in all states of the world, and so they do so also in expectation. This gives us that $\tilde{r}(x, W_2) < \tilde{r}(x, W_1)$. Given that λ is equal to inverse marginal utility it is positive. But then the right hand side is negative, so we have a contradiction. So independent of the randomization, the wage will move according to the sign of the expected profit.

Monotonicity in z: the final step is to show that the efficiency wage is increasing in z. We already know that $\mathcal{J}(x, z, V)$ is increasing in z and decreasing and concave in V. Let's consider $z_1 < z_2$ and associated efficiency wage $w^*(x, z_1)$. We want to show that $w^*(x, z_1) < w^*(x, z_2)$. Call ξ_1 the optimal policy starting at state $\mathcal{J}(x, z_1, V_1)$ where V_1 delivers $w^*(x, z_1)$ and using ξ_1 at (x, z_2), the worker receives V_1 and is paid $w^*(x, z_1)$. The firm makes more profit than at z_1 since $f(x, z)$ is increasing in z and $\mathbb{E}\mathcal{J}$ is larger as well. The optimal policy at (x, z_2, V_1) will pay a higher wage than $w^*(x, z_1)$ to trade some output for a longer expected lifespan, but continue to choose positive $\mathbb{E}\mathcal{J}$. So we found a wage $w_3^* \geq w^*(x, z_1)$ such that $\mathbb{E}\mathcal{J}$ is still positive. This last point implies that $w_3^* \leq w^*(x, z_2)$ and concludes. □

A.4 Solving the model (online publication)

The main difficulty resides in solving the firm's problem where tackling directly (BE-F) requires finding the promised utilities $W_{z'x'}$ in each state of the world for the next period. This becomes infeasible as soon as reasonable supports are considered for $\mathbb{X}$ and $\mathbb{Z}$. However, the first order condition with respect to W reveals that the utility promised in different states are linked to each other. Call $\lambda\beta p(x, W)$ the multiplier for the $W = \sum W_{z'x'}$ constraint, then the first order condition for $W_{x'z'}$ is

$$\frac{\partial \mathcal{J}}{\partial V}(x', z', W_{x',z'}) = \lambda,$$

where given λ, if $\mathcal{J}$ is strictly concave, then all the $W_{x'z'}$ are pinned down. This reduces the search to one dimension. The simplification comes from the fact that the firm always tries to insure the worker as much as possible across future states, and does this by keeping her marginal utility constant across realizations. Indeed, we know that the derivative of $\mathcal{J}$ is the inverse marginal utility. One difficulty however is that $\mathcal{J}$ might be weakly concave in some regions. In that case one needs to keep track of a set of possible feasible promised utilities $W_{x'z'}$. Given the concavity of $\mathcal{J}$ this set will be an interval fully captured by its two extremities. This means that at worst the number of the control variables is augmented by one. Using the marginal utility in the state space is known as the recursive Lagrangian approach as developed by Kocherlakota (1996); Marcet and Marimon (2011); Messner, Pavoni, and Sleet (2012); Cole and Kubler (2012). The problem of non-strict concavity persists in this formulation but Cole and Kubler (2012) show how to overcome this difficulty by keeping track of the upper and lower bound of the set of solutions. Numerically I solve the firm problem using recursive Lagrangian and do not find any such flat region. The recursive Lagrangian for the firm problem is derived in Appendix A.4 and is given by:

$$\mathcal{P}(x,z,\rho) = \inf_{\gamma} \sup_{w,W} f(x,z) - w + \rho\left(u(w_i) + \tilde{r}(x,W)\right) \\ - \beta\gamma\tilde{p}(x,W)\cdot W + \beta\tilde{p}(x,W)\mathbb{E}\mathcal{P}(x',z',\gamma), \quad (4)$$

where

$$\mathcal{P}(x,z,\rho) := \sup_{v} \mathcal{J}(x,z,v) + \rho v.$$

We now go through the details of the proof:

Proof. Ignoring the lottery, we have the following recursive formulation for $\mathcal{J}$:

$$\begin{aligned}\mathcal{J}(x,z,V) = \sup_{\pi_i, W_i, W_{ix'y'}} & f(x,z) - w_i + \beta\tilde{p}(x,W_i)\mathbb{E}\mathcal{J}(x',z',W_{ix'y'}) \\ s.t \quad (\lambda) \quad & 0 = u(w_i) + \tilde{r}(x,W_i) - V, \\ (\gamma_i) \quad & 0 = W_i - \mathbb{E}W_{ix'z'}.\end{aligned}$$

From which we can construct the Pareto problem

$$\mathcal{P}(x,z,\rho) = \sup_v \mathcal{J}(x,z,v) + \rho v.$$

We seek a recursive formulation. I first substitute the definition of $\mathcal{J}$ and the constraint on λ in $\mathcal{P}$ to get

$$\begin{aligned}\mathcal{P}(x,z,\rho) = \sup_{V,w,W,W_{x'z'}} & f(x,z) - w + \beta\tilde{p}(x,W)\mathbb{E}\mathcal{J}(x',z',W_{x'z'}) + \rho V \\ s.t \quad (\lambda) \quad & 0 = u(w_i) + \tilde{r}(x,W) - V, \\ (\gamma) \quad & 0 = W - \mathbb{E}W_{x'z'}.\end{aligned}$$

at which point I can substitute in the V constraint:

$$\begin{aligned}\mathcal{P}(x,z,\rho) = \sup_{V,w,W,W_{x'z'}} & f(x,z) - w + \beta\tilde{p}(x,W)\mathbb{E}\mathcal{J}(x',z',W_{x'z'}) + \rho\left(u(w_i) + \tilde{r}(x,W)\right) \\ s.t \quad (\gamma) \quad & 0 = W - \mathbb{E}W_{x'z'}.\end{aligned}$$

then I append the constraint $(\gamma)$with weight $\beta\gamma\tilde{p}(x,W)$

$$\begin{aligned}\mathcal{P}(x,z,\rho) \quad = \quad & \inf_{\gamma} \sup_{V,w,W,W_{x'z'}} f(x,z) - w + \rho\left(u(w_i) + \tilde{r}(x,W)\right) \\ & -\gamma\beta\tilde{p}(x,W)(W - \mathbb{E}W_{x'z'}) \\ & +\beta\tilde{p}(x,W)\mathbb{E}\mathcal{J}(x',z',W_{x'z'})\end{aligned}$$

which finally we recombine as

$$\begin{aligned}\mathcal{P}(x,z,\rho) &= \inf_{\gamma} \sup_{V,w,W,W_{x'z'}} f(x,z) - w + \rho\left(u(w_i) + \tilde{r}(x,W)\right) \\ &\quad -\beta\gamma\tilde{p}(x,W)W \\ &\quad +\beta\tilde{p}(x,W)\mathbb{E}\mathcal{J}(x',z',W_{x'z'}) + \gamma\mathbb{E}W_{x'z'}\end{aligned}$$

the final step is to move the sup to the right hand side to get:

$$\begin{aligned}\mathcal{P}(x,z,\rho) &= \inf_{\gamma} \sup_{w,W} f(x,z) - w + \rho\left(u(w_i) + \tilde{r}(x,W)\right) \\ &\quad -\beta\gamma\tilde{p}(x,W)W \\ &\quad +\beta\tilde{p}(x,W)\mathbb{E}\left[\sup_{W_{x'z'}} \mathcal{J}(x',z',W_{x'z'}) + \gamma W_{x'z'}\right]\end{aligned}$$

where we recognize the expression for $\mathcal{P}$ and so we are left with solving the following saddle point functional equation (SPFE):

$$\begin{aligned}\mathcal{P}(x,z,\rho) = \inf_{\gamma} \sup_{w,W} & f(x,z) - w + \rho\left(u(w_i) + \tilde{r}(x,W)\right) \\ & - \beta\gamma\tilde{p}(x,W)W + \beta\tilde{p}(x,W)\mathbb{E}\mathcal{P}(x',z',\gamma). \quad \text{(SPFE)}\end{aligned}$$

From the solution of this equation we can reconstruct the lifetime utility of the worker, and the profit function of the firm

$$\begin{aligned}\mathcal{V}(x,z,\rho) &= \frac{\partial\mathcal{P}}{\partial\rho}(c,z,\rho) \\ \mathcal{J}(x,z,v) &= \mathcal{P}(x,z,\rho^*(x,z,v)) - \rho^*(x,z,v)\cdot v.\end{aligned}$$

□

B Identification appendices (online publication)

This section covers the non-parametric identification of the model. I will present the format of the data, then I will state the results and describe them, and then include the formal proofs. The procedure is composed of two steps, first we show that we can recover the law motion of the underlying productivity of the worker and the firm, as well as how it is linked to wages and mobility decisions. With this information, we use a conditional choice probability argument to show that the structural parameters are identified.

B.1 Overview of identification

For reasons that will become apparent in Lemma 5 the model lacks variation in the realization of wages conditional on the past. As we have seen, w_{t+1} is a deterministic function of the x_t, z_t, w_t (plus the lottery). To address this problem, the reader should consider in this section a similar model, where in addition, both worker and firms face commitment issues with some small probabilities. More precisely, imagine that each period a commitment shocks hits either the firm or the worker with a value drawn from a x, z specific distribution. This represents a small modification of the recursive Lagrangian which will push γ up or down with some probability to match the participation constraint if necessary. The problem becomes:

$$\mathcal{P}(x,z,\rho) = \inf_{\gamma} \sup_{w,W} f(x,z) - w + \rho\left(u(w_i) + \tilde{r}(x,W)\right) - \beta\gamma\tilde{p}(x,W)W + \beta\tilde{p}(x,W)\mathbb{E}\mathcal{P}(x',z',\gamma+\epsilon). \quad (5)$$

where ϵ is random, assumed independent of x', z' conditional on x, z, but a function of x, z in a way that does not generate separation. The solution remains Markovian, nothing changes in this model besides the fact that now $w_{t+1}|x_t, z_t, w_t$ will generate a larger support. In recursive Lagrangian formulation, participation constraints affect the ρ parameter directly.

Data and econometric model. Consider a worker i observed over T periods. Call $X_{it} \in \{1..n_x\}$ his unobservable ability and $Z_{it} \in \{1..n_z\}$ his current firm level match if employed with $Z_{it} = 0$ when not employed. We denote by Y_{it} the wage and set $Y_{it} = 0$ whenever the worker is unemployed. We finally call M_{it} the mobility realization with $M_{it} = 0$ when the worker stayed in the same firm between $t-1$ and t, $M_{it} = 1$ if the worker moved to a new firm, $M_{it} = 2$ is the worker left his position into unemployment and $M_{it} = 3$ if the worker found a job out of unemployment. And finally $M_{it} = 4$ if an unemployed worker remains unemployed. Note here the timing which implies that the realization of the separation shock in the current period is reflected in M_{t+1}, not M_t. This is natural given the timing in the model where the wage is collected before separation.

In addition, for each period where the worker i is employed in a firm, we consider K co-workers who joined the firm at the same time as worker i (this might have been multiple periods before). When the worker is unemployed we keep the co-workers from the last employment. We index them by $k(i, t)$. For this co-workers we consider their wage which we write Y^c_{ikt} where $Y^c_{ikt} = 0$ if the co-worker became unemployed. Our data is then formed of a random sample of such sequences of $\{Y_{it}, M_{it}, Y^c_{i1t}, ...Y^c_{iKt}\}_{i=1..N,t=1..T}$.

We are going to focus our analysis on $K = 2$ and $T = 5$ since this is sufficient for our identification. We first focus on individual who moved coming into the first period $M_{i1} = 1$ or $M_{i1} = 3$. For these workers we are going to recover the sequence of realizations of the latent Z_t for $t = 1..4$. Next, we are going to use the Markovian property of the contract to recover the law of motion with respect to the second latent variable X again for this population of movers. Since the process is stationary, we get the law of motion for all states in the economy.

Throughout this discussion Y_{it} will be thought of as a discrete outcome since the conditions are easier to understand in this context. We can think of taking the percentiles of wages. Dealing with the continuous cases requires changing the full rank assumption into a linear independence requirements of the marginal distributions, see Allman, Matias, and Rhodes (2009) Theorem 8. Finally we assume that Z has n_z point of supports and X has n_x point of supports.

Lemma 4. *$Pr[Y_{i,1..5}, Z_{i,1..4}|M_{i,1}{=}1,3]$ is identified from $Pr[Y_{i,1..5}, Y^c_{i,k,1..5}|M_{i,1}{=}1,3]$ under the assumptions of the structural model and the following conditions:*

i) $Pr[Y_{i,1..5}|Z_{i,1..4}, M_{i,1}{=}1,3]$ *and* $Pr[Y^c_{i,k,1..5}|Z_{i,1..4}, M_{i,1}{=}1,3]$ *have rank* n_z^4.

ii) $\exists y_{1..4}, y'_{1..4}$ *s.t.* $\forall Z_{1..4}$

$$Pr[Z_{i,1..4}, Y^c_{i,1,1..5} = y_{1..5}|M_{i,1}{=}1,3] \neq Pr[Z_{i,1..4}, Y^c_{i,1,1..5} = y'_{1..5}|M_{i,1}{=}1,3],$$

iii) $Pr[M_{t+1} = 0|X_t, Z_t, Y_t] \geq Pr[M_{t+1} = 0|X_t, Z'_t, Y_t]$ *whenever* $Z_t \geq Z'_t$

Proof. See next section. □

The proof relies on the property of the model that conditional on the sequence $Z_{i,1..4}$ the realizations of wages of all co-workers are independent of each other. This is due to the fact that all common shocks have to be firm shocks. The conditional independence structure allows to apply the result for discrete mixture (Hall and Zhou, 2003) to the sequence of Z_{it}.

An additional difficulty is to be able to label correctly the different sequences. This is achieved by assuming that the probability to leave the firm is decreasing in z everything else equal (a property satisfied at the parameter values). We can then use a lexicographic ordering to recover the Z^t sequences. We then move to recovering the law of motion for X_{it} and for the wages Y_{it} as well as the mobility outcomes M_{it}.

Lemma 5. *Call $S_{it} = \{Y_{it}, Z_{it}, M_{it}\}$, under the assumptions of the structural model, $Pr[X_t|X_{t-1}]$ and $Pr[S_t|X_{t-1}, S_{t-1}]$ are identified under the following conditions:*

i) The matrix $A(s_2, s_3)$ defined as $a_{pq} = Pr[S_{i1}{=}s_p, S_2{=}s_2, S_3{=}s_3, S_4{=}s_q]$ has rank n_x for each s_2, s_3

ii) For any s_2, s_3, there exists s_2', s_3' such that for all $k \neq j$ we have $\lambda_{s_2,s_2',s_3,s_3'}(k) \neq \lambda_{s_2,s_2',s_3,s_3'}(j)$ where

$$\lambda_{s_2,s_2',s_3,s_3'}(k) = \frac{Pr[S_3{=}s_3|S_2{=}s_2, X_2{=}k] \cdot Pr[S_3{=}s_3'|S_2{=}s_2', X_2{=}k]}{Pr[S_3{=}s_3'|S_2{=}s_2, X_t{=}k] \cdot Pr[S_3{=}s_3|S_2{=}s_2', X_2{=}k]}$$

iii) $Pr[X_t|X_{t-1}]$ is diagonal dominant.

Proof. See next section. □

This applies the result from Hu and Shum (2012) about the identification of Markov-switching model. In the model, the productivity process is independent of the wage process or the match quality realization. This gives us the property of limited dependence required for the result to apply. The two rank conditions listed above can appear restrictive. Condition i) requires enough variation in the realizations of wages conditional on past wages (and inversely).

Because it requires the joint distribution with Z, it is hard to directly test in the data.

The second condition is perhaps less transparent. Connault (2015) tells us that this conditions are usually not satisfied in structural model since many transitions do not happen and some distributions are degenerated. In the version of the model without commitment shocks, this could be the case. For this reason, the case that uses Hu and Shum (2012) applies to the version of the model presented in equation 5.

As an example, let's ignore the mobility decision for a moment and consider a reduce form version of the optimal contract where $y_{t+1} = y_t + g(y_t, x_t) + \epsilon_t$. This is a version of Equation EQ-FOC with log utility, where g is the expected profit of the firm. Then we can look at $\lambda_{s_2,s'_2,s_3,s'_3}(k)$ under normality of ϵ which is given by $(\Delta y' - \Delta y)(g(y'_1, x) - g(y_1, x))$. So as long as y_1 and x interact in g, this will have different values as a function of x. In addition, any difference in the persistence, or heteroskedasticity of $y_{t+1}|y_t, x_t$ across values of x_t will be sufficient.

Finally, let's consider briefly what happens in the case where we have a strictly deterministic rule $S_{t+1} = g(S_t, X_t)$. Assuming invertibility of $x \to g(s_t, x)$, we have that given S_t, there is a one to one mapping between S_{t+1} and X_t. Next using three realizations S_1, S_2, S_3 together with diagonal dominance of $X_{t+1}|X_t$ and non-zero transitions, I can label the X across S_t, S_{t+1} pairs. This delivers directly both $X_{t+1}|X_t$ and $S_{t+1} = g(S_t, X_t)$. The current model is in between this extreme case, and the random one since Z_t, M_t are randomly drawn but the wage is a deterministic function.

In this procedure, it seems that we were able to abstract from the selection into work. The reason why we were able to do so is because the timing of the model rules out selection in the sense that the realization of Y_{t+1} does not affect the separation decision conditional on Y_t, X_t, Z_t. The agent makes the effort decision before the realization of the new variables. This is similar to the timing assumption of Rust (1987).

Recovering the structural parameters We now want to use the econometric model to estimate the structural parameters of the economic model. For this, we note that we have already directly estimated all the transition probabilities.

Lemma 6. *From the econometric model described in the previous section, with known utility function $u(\cdot)$ and cost function $c(\cdot)$, we show that*

i) the $b(\cdot)$ and $p(x, V)$ functions are identified.

ii) using the firm problem and the optimal contract assumption of Section 3.3, the production function $f(x, z)$ is identified.

Proof. See next section. □

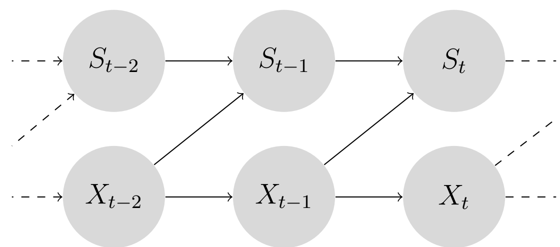
Once the transition probabilities and the flow pay-off have been recovered in the previous section, we can almost directly reconstruct the present value at each state. The complications that have to be overcome are to express the continuation values at job changes. We don't want to assume that the flow value of unemployment is observed and so we have to recover it from the transitions at our disposal. The proof shows how we can use the fact that when the worker is close to indifference between working and not working, then the probability that she quits is very close to one. By conditioning on $\delta^* \simeq 1$ the continuation value of the worker is identical to the value of begin unemployed. The second difficult value to reconstruct is the value v_1^* that the worker gets when he moves. This however can be reconstructed using the present value conditional on moving. Recovering the production function is achieved using the property of the optimal contract that $\frac{\partial \mathcal{J}}{\partial V}(x,z,v) = \frac{1}{u'(w)}$. Using the freshly recovered $W(x,z,w)$ we can integrate to get $\mathcal{J}$. The constant term is pinned down by the residual claimant wage $w^*(x,z)$ for which we know that $\mathbb{E}[\mathcal{J}(x',z',w')|x,z,w^*(x,z)] = 0$.

At this point we should ask ourself if we could use additional information to discipline the two functions $u(\cdot)$ and $c(\cdot)$. I show that even in the case where $c(\cdot)$ is not know, $W(x,z,w)$ can be written as the solution of linear integral equation of the second kind. This makes non-parametric identification hopeful, even-though so far, proving that this equation is a contraction has appeared difficult. Never-the-less the first order condition of the worker with respect to effort provides many additional restrictions that can be used to discipline a parametric cost function. As for the utility function $u(\cdot)$ we note that at this point we have not used an overall measure of pass-trough form productivity to earnings. Adding a moment as the covariance between value added growth and earning growth, similar to the first section of this paper or to Guiso, Pistaferri, and Schivardi (2005), should help measure the amount of risk aversion.

A final comment is that we have shown that the entire $p(x,V)$ function as well as the $\mathcal{J}(x,z,V)$ function can be recovered from the data without using the free entry condition. This means that one could now use these two function to recover the matching function.

B.2 Proofs

I start by providing a reduce form representation for the law of motion implied by the contract. It will be according to the following graph:



Lemma 7. *S_t and X_t satisfy:*

$$Pr[S_{t+1}, X_{t+1}|S_t, X_t, \Omega_t] = Pr[S_{t+1}|S_t, X_t]Pr[X_{t+1}|X_t]$$

Proof. The first law of motion comes directly from the assumption of the model that X_t evolves according to Γ_x and the fact that the sample follows individuals in and out of employment, hence not creating any selection.

For the second law of motion we have to look separately at the different components of $S_t = (Y_t, M_t, Z_t)$. We also recall that conditional on X_t, Z_t, we have that V_t and $Y_t = w_t$ are interchangeable since $\frac{1}{u'(w_t)} = -\frac{\mathcal{J}}{v}(x_T, z_t, V_t)$ and $\mathcal{J}(x, z, V)$ is decreasing and concave.

Let's consider the mobility decisions M_{t+1} first. For the employed worker with $M_{t-1} = 0, 1, 3$, the mobility decision is taken at the end of the period as a function of the current state space. The strategies give us the separation rates $q^*(X_t, Z_t, V_t)$ and $\tilde{p}(X_t, v_1^*(X_t, Z_t, V_t))$. And the realization of the mobility is independent of the realization of X_{t+1}. Similarly, the realization of finding a job for the unemployed realizes independently of the X_{t+1} giving that in all values of M_{t+1}, M_t:

$$Pr[M_{t+1}, X_{t+1}|S_t, X_t, \Omega_t] = Pr[M_{t+1}|S_t, X_t]Pr[X_{t+1}|X_t]$$

Let's then consider the outcomes Y_{t+1} and Z_{t+1}. Let's first address the case of job stayers. In this case the new wage is determined by Equation EQ-FOC and so we see that it is a function of the past state only, and independent of the X_{t+1} realization, and the same is true of the Z_{t+1} realization. This is a result of the fact that the firm equates marginal utilities across states and so the wage realization does not depend on X_{t+1}. Commitment shocks would not affect this independence.

For job movers and unemployed workers finding a new job, the new realization Z_{t+1} is independent by assumption in the model: it is drawn condition on $Z = Z_0$. The new wage offered by the vacancy posting will solve a similar problem to the one for stayers (see Equation BE-V). This means that the new wage will also be independent of the state realization. Finally, for workers loosing their job and unemployed not finding one, we set Y_{t+1}=0 and Z_{t+1}=0, in which case it is independent since it is degenerated.

We end up with the following expression:

$$Pr[Y_{t+1}, Z_{t+1}, |M_{t+1}, X_{t+1}, S_t, X_t, \Omega_t] = Pr[Y_{t+1}, Z_{t+1}|M_{t+1}, S_t, X_t]$$

which combined with the previous expression gives us

$$Pr[S_{t+1}, X_{t+1}|S_t, X_t, \Omega_t] = Pr[S_{t+1}|S_t, X_t]Pr[X_{t+1}|X_t].$$

□

Lemma 4. *$Pr[Y_{i,1..5}, Z_{i,1..4}|M_{i,1}{=}1,3]$ is identified from $Pr[Y_{i,1..5}, Y^c_{i,k,1..5}|M_{i,1}{=}1,3]$ under the assumptions of the structural model and the following conditions:*

i) $Pr[Y_{i,1..5}|Z_{i,1..4}, M_{i,1}{=}1,3]$ *and* $Pr[Y^c_{i,k,1..5}|Z_{i,1..4}, M_{i,1}{=}1,3]$ *have rank* n_z^4.

ii) $\exists y_{1..4}, y'_{1..4}$ *s.t.* $\forall Z_{1..4}$

$$Pr[Z_{i,1..4}, Y^c_{i,1,1..5} = y_{1..5}|M_{i,1}{=}1,3] \neq Pr[Z_{i,1..4}, Y^c_{i,1,1..5} = y'_{1..5}|M_{i,1}{=}1,3],$$

iii) $Pr[M_{t+1} = 0|X_t, Z_t, Y_t] \geq Pr[M_{t+1} = 0|X_t, Z'_t, Y_t]$ *whenever* $Z_t \geq Z'_t$

Proof. To achieve this result we apply the result of identification of mixture. The model tells us that the wage path of a given worker is a function of his own sequence of shocks, and given the firm shock history, individual specific shocks are independent across co-workers. Hence once we condition on the sequence of shock at the firm level, we achieve the conditional independence required for the identification of mixtures. It works as long as we can go back far enough to condition on the full firm shock history shared between co-workers. To achieve this we look at workers who enter in period 1. The model tells us that $Pr[Y_t, Y_{t-1}, X_{t_1}, Z_{t-1}, \Omega_t] = Pr[Y_t|Y_{t-1}, X_{t_1}, Z_{t-1}]$, hence we can write that:

$$\begin{aligned} &Pr[Y_{i,1..5}, Y^c_{i,1..2,1..5}|M_{i,1}{=}1,3] \\ &\quad = \sum_{Z_{1..4}} Pr[Z_{1..4}|M_{i,1}{=}1,3] \cdot Pr[Y_{i,1..5}, Y^c_{i,1..5}|Z_{1..4}, M_{i,1}{=}1,3] \\ &\quad = \sum_{Z_{1..4}} Pr[Z_{1..4}|M_{i,1}{=}1,3] \cdot Pr[Y_{i,1..5}|Z_{1..4}, M_{i,1}{=}1,3] \cdot \left(\prod_k Pr[Y^c_{ik,1..5}|Z_{1..4}, M_{i,1}{=}1,3] \right) \end{aligned}$$

Where we have used the conditional independence to explicitly show the mixture structure of the data. The reader should now see that with only 2 coworkers observations we get 3 independent measures conditional on the sequence $Z_{1..4}$. The objects of interest here are $Pr[Z_{1..4}|M_{i,1}{=}1,3]$ and $Pr[Y_{i,1..5}|Z_{1..4}, M_{i,1}{=}1,3]$. The first condition requires two values of wages $y_{1..5}$ and $y'_{1..5}$ for the second co-worker such that the matrices of outcome where these wages are fixed have rank n_z^4:

$B(\boldsymbol{y})$ where $b_{qp} = Pr[Y_{i,1..5} = \boldsymbol{y}_p, Y^c_{i,1,1..5} = \boldsymbol{y}, Y^c_{i,2,1..5} = \boldsymbol{y}_q|M_{i,1}{=}1,3]$ has rank n_z^4.

The second condition for the identification is that eigen values are different. This requires for the outcome for the first co-worker to be varying with the different values of $Z_{1..4}$. We need:

$$\text{for all } Z_{1..4}, \quad Pr[Y^c_{i,1..5} = \boldsymbol{y} | Z_{1..4}, M_{i,1}{=}1,3] \neq Pr[Y^c_{i,1..5} = \boldsymbol{y'} | Z_{1..4}, M_{i,1}{=}1,3]$$

Labeling. $Z_{1..4}$ is a vector of 4 values in $1..n_z$ and so take n_z^4 different values. Call $\bar{z}_p$ an indexing of this values, then following Hall and Zhou (2003) this guarantees the identification of $Pr[\bar{Z}|M_{i,1}{=}1,3]$ as well as $Pr[Y_{i,1..5}|\bar{Z}, M_{i,1}{=}1,3]$. There are two labeling issues. The first one is to group together paths that have common histories. This is straight forward since path with similar histories until t will have identical joint distributions of $(Y_{i,1..t}, Y^c_{i,k,1..t})$. The second labeling issue is to order the z within period. To achieve this, we use the monotonicity assumption with respect to z which tells us that separation will happen more often for lower values of z. We can then order the sequences using a lexico-graphic rule. Consider the following absolute and transitive ordering between values of $\bar{z}$.

$$\begin{aligned} \bar{z}_1 \prec \bar{z}_2 : \qquad & \text{if } Pr[Y^c_{1,t=1}|\bar{z}_1] < Pr[Y^c_{1,t=1}|\bar{z}_2] \\ & \text{else if } Pr[Y^c_{1,t=2} = 0|\bar{z}_1] < Pr[Y^c_{1,t=2} = 0|\bar{z}_2] \\ & \text{else if } Pr[Y^c_{1,t=3} = 0|\bar{z}_1] < Pr[Y^c_{1,t=3} = 0|\bar{z}_2] \\ & \text{else if } Pr[Y^c_{1,t=4} = 0|\bar{z}_1] < Pr[Y^c_{1,t=4} = 0|\bar{z}_2] \end{aligned}$$

This will order the $\bar{z}$ sequence in the exact same order as the lexicographic order on $Z_{1..4}$.

□

Lemma 5. *Call $S_{it} = \{Y_{it}, Z_{it}, M_{it}\}$, under the assumptions of the structural* *model,* $Pr[X_t|X_{t-1}]$ *and* $Pr[S_t|X_{t-1}, S_{t-1}]$ *are identified under the following conditions:*

i) The matrix $A(s_2, s_3)$ *defined as* $a_{pq} = Pr[S_{i1}{=}s_p, S_2{=}s_2, S_3{=}s_3, S_4{=}s_q]$ *has rank* n_x *for each* s_2, s_3

ii) For any s_2, s_3*, there exists* s'_2, s'_3 *such that for all* $k \neq j$ *we have* $\lambda_{s_2,s'_2,s_3,s'_3}(k) \neq \lambda_{s_2,s'_2,s_3,s'_3}(j)$ *where*

$$\lambda_{s_2,s'_2,s_3,s'_3}(k) = \frac{Pr[S_3{=}s_3|S_2{=}s_2, X_2{=}k] \cdot Pr[S_3{=}s'_3|S_2{=}s'_2, X_2{=}k]}{Pr[S_3{=}s'_3|S_2{=}s_2, X_t{=}k] \cdot Pr[S_3{=}s_3|S_2{=}s'_2, X_2{=}k]}$$

iii) $Pr[X_t|X_{t-1}]$ *is diagonal dominant.*

Proof. For this lemma we use the $Pr[Y^c_{i,1..5} = \boldsymbol{y}|M_{i,1}{=}1,3, Z_{1..4}]$ that we identified from the previous section and we want to recover the Markov process with respect to the underlying latent heterogeneity X which is exogenous. Given that the law of motion is stationary, I can drop the $M_{i,1}{=}1,3$ which will only affect the initial distribution.

The first step is to show that the conditional transition rule is indeed Markov and that it satisfies the limited dependence condition of Assumption 1 in Hu and Shum (2012). We first note that the timing is slightly different in the sense that in their paper Y_t is affected by X_t, whereas here it is affected by X_{t-1}, but this is only a timing assumption. We apply the proof of their paper but we label our X_t to be their X_{t+1}. The two properties we need to show are (dropping the i):

$$Pr[S_t, X_{t-1}|S_{t-1}, X_{t-2}, S_{t-2}, \Omega_{t-3}] = Pr[S_t, X_{t-1}|S_{t-1}, X_{t-2}]$$
$$Pr[S_t|S_{t-1}, X_{t-1}, X_{t-2}] = Pr[S_t|S_{t-1}, X_{t-1}].$$

We have shwon in Lemma 7 that we have the Markovian property for S_t, X_t

but here we need it for S_t, X_{t-1}.

$$
\begin{aligned}
&Pr[S_t, X_{t-1}|S_{t-1}, X_{t-2}, S_{t-2}, \Omega_{t-3}] \\
&\quad = Pr[S_t|S_{t-1}, X_{t-1}, X_{t-2}, S_{t-2}, \Omega_{t-3}] Pr[X_{t-1}|S_{t-1}, X_{t-2}, S_{t-2}, \Omega_{t-3}] \\
&\quad = \Big(\sum_{X_t} Pr[S_t, X_t|S_{t-1}, X_{t-1}, X_{t-2}, S_{t-2}, \Omega_{t-3}] \Big) Pr[X_{t-1}|S_{t-1}, X_{t-2}, S_{t-2}, \Omega_{t-3}] \\
&\quad = Pr[S_t|S_{t-1}, X_{t-1}] \frac{Pr[X_{t-1}, S_{t-1}|X_{t-2}, S_{t-2}, \Omega_{t-3}]}{Pr[S_{t-1}|X_{t-2}, S_{t-2}, \Omega_{t-3}]} \\
&\quad = Pr[S_t|S_{t-1}, X_{t-1}] Pr[X_{t-1}|X_{t-2}]
\end{aligned}
$$

where we have used the result of Lemma 7. The expression does not depend on S_{t-2} and Ω_{t-3} which gives us our result.

Next we show the property of limited dependence.

$$
\begin{aligned}
&Pr[S_t|S_{t-1}, X_{t-1}, X_{t-2}] \\
&\quad = \sum_{X_t} Pr[S_t, X_t|S_{t-1}, X_{t-1}, X_{t-2}] \\
&\quad = \sum_{X_t} Pr[S_t, X_t|S_{t-1}, X_{t-1}] \\
&\quad = Pr[S_t|S_{t-1}, X_{t-1}]
\end{aligned}
$$

Labeling. The final condition required to apply Theorem 2 from An, Hu, Hopkins, and Shum (2013) is a condition to resolve the labeling of X_t across S_{t-1} states. An, Hu, Hopkins, and Shum (2013) suggests using a monotonicity condition. As discussed in section 4.1 of Hu and Shum (2012), the monotonicity condition 4 of their paper is not necessary to recover $Pr[S_{i,t}|S_{i,t-1}, X_{t-1}]$. It is necessary to link the X across times and get the correct $Pr[X_t|X_{t-1}]$. Another sufficient restriction is then for $Pr[X_t|X_{t-1}]$ to be diagonal dominant. This is the equivalent of the ordering assumption 7 in Hu (2008). I choose to rely on that restriction on the primitives. □

Lemma 6. *From the econometric model described in the previous section, with known utility function $u(\cdot)$ and cost function $c(\cdot)$, we show that*

i) the $b(\cdot)$ and $p(x, V)$ functions are identified.

ii) using the firm problem and the optimal contract assumption of Section 3.3, the production function $f(x, z)$ is identified.

Proof. I am going to show that we have recovered all the transitions necessary to reconstruct the present value of the worker $W(x, z, w)$ at a given state (x, z, w). Recall from the model section that it is given by:

$$\begin{aligned}\forall x, z, w,\ W(x, z, w) &= \sup_{v,e} u(w) - c(e) + \delta(e)\beta \mathbb{E}_{x'|x}\mathcal{U}(x') + (1 - \delta(e))\beta\kappa p(\theta(x, v))v \\ &\quad + \beta(1 - \delta(e))(1 - \kappa p(\theta(x, v)))\mathbb{E}W(x', z', w') \\ &= u(w) - c(\delta^*) + \delta^*\beta\mathbb{E}_{x'|x}\mathcal{U}(x') + (1 - \delta^*)\beta\kappa p_1^* \cdot v_1(x, z, w) \\ &\quad + \beta(1 - \delta^*)(1 - \kappa p_1^*)\mathbb{E}W(x', z', w'),\end{aligned}$$

where the second expression substituted in the optimal policy. Here we assume that $c(\cdot)$ is known to the econometrician. We are now going to replace each of the expectations and the present values with empirical counterparts. We first note that we can rewrite $v_1(x, z, w)$ and $v_0(x)$ as functions of the empirical transitions from Lemma 5:

$$\begin{aligned}v_1(x, z, w) &= \mathbb{E}[W(x', z', Y')|x, z, w, m = 1] \\ v_0(x) &= \mathbb{E}[W(x', z_0, Y')|x, m = 3],\end{aligned}$$

where the expectations are taken with respect to $Pr[S_{i,t}|S_{i,t-1}, X_{t-1}]$. We can then rewrite

$$\forall x, z, w,\ W(x,z,w) = u(w) - c(\delta^*) + \delta^* \beta \mathbb{E}_{x'|x} \mathcal{U}(x') \\ + (1-\delta^*)\beta\kappa p_1^* \cdot \mathbb{E} W(x', z_0, w_1(x,z,w)) \\ + \beta(1-\delta^*)(1-\kappa p_1^*)\mathbb{E}W(x', z', w'),$$

In the previous expression, the unknown are W,U and $c(\cdot)$. To get $U(x)$ we note that the theory tells us that as value promised to the worker if the job does not break down goes to $\mathbb{E}U(x')$ as the probability of staying goes to 0. This value the worker gets in this case is given by

$$\widetilde{W}(x,z,w) = \kappa p_1^* \cdot \mathbb{E}[W(x', z_0, w')|x, z, w, m{=}1] \tag{6}$$

$$+ (1-\kappa p_1^*)\mathbb{E}[W(x', z', w')|x, z, w, m{=}0]. \tag{7}$$

This is an implication of $-c'(\delta^*) = \beta\widetilde{W}(x,z,w) - \beta\mathbb{E}[U(x_{t+1})|x_t = x)]$ together with $c'(1) = 0$. Call $\underline{w}(x,z)$ the wage such that

$$\underline{\mathrm{w}}(x,z) = \arg\min_w \delta^*(x,z,w) \quad \text{s.t.} \quad \delta^*(x,z,w) > 0,$$

then we have that $\mathbb{E}[U(x_{t+1})|x_t = x] = \widetilde{W}(x, z, \underline{\mathrm{w}}(x,z))$ which we can now replace in our previous expression. This gives us our first result. We have now expressed the W function as an integral equation. We substitute in the terms we just derived and get:

$$\begin{aligned}
W(x,z,w) = u(w) - c(\delta^*) \\
&+ \beta\delta^*\kappa p_1^* \cdot \mathbb{E}[W(x',z',w'|x,z,\underline{w}(x,z),m{=}1] \\
&\quad + \beta\delta^*(1-\kappa p_1^*)\mathbb{E}[W(x',z',w')|x,z,\underline{w}(x,z),m{=}0] \\
&+ (1-\delta^*)\beta\kappa p_1^* \cdot \mathbb{E}[W(x',z',w')|x,z,w,m{=}1] \\
&+ \beta(1-\delta^*)(1-\kappa p_1^*)\mathbb{E}[W(x',z',w')|m{=}0],
\end{aligned}$$

We see that this mapping satisfies the Blackwell-Boyd conditions of discounting and monotonicity.

Part 2, recovering the production function $f(x,z)$ Finally, we use the property that $\frac{\partial \mathcal{J}}{\partial V}(x,z,v) = \frac{1}{u'(w)}$ to identify the $\mathcal{J}$ function. In addition, the intercept for the $\mathcal{J}$ function is pinned down by the residual claimant wage $w^*(x,z)$ for which we know that $\mathbb{E}[\mathcal{J}(x',z',w')|x,z,w^*(x,z)] = 0$, and for which we know it satisfies that $\Delta w_t = 0$ whenever $x_t = x_{t+1}$ and $z_t = z_{t+1}$. This identifies the $\mathcal{J}$ function.

We can then use the Bellman equation of the contracting problem of the firm to recover the production function:

$$f(x,z) = w^*(x,z,V) + \mathcal{J}(x,z,V) - \tilde{p}^*(x,V)\beta\mathbb{E}\mathcal{J}(x',z',W')$$

□

B.3 When $c(\cdot)$ is not known

I develop here how we could try to recover the cost function from the data as well. We are now left with one term which remains unknown, the $c(\cdot)$ function.

To do so I show that we can write a similar integral equation. We use again the effort decision which states:

$$-c'(\delta^*(x,z,w)) = \beta\widetilde{W}(x,z,w) - \beta\mathbb{E}[U(x_{t+1})|x_t = x)]$$

we multiply both sides by $\delta^*_w(x,z,w)$, the derivative of $\delta^*(x,z,w)$ with respect to w and we then integrate from $\underline{w}(x,z)$ to a w which gives

$$\begin{aligned}
-c(\delta^*(x,z,w)) &= \beta \int_{\underline{w}(x,z)}^{w} \delta^*_w(x,z,u)\Big(\widetilde{W}(x,z,u) - \mathbb{E}[U(x_{t+1})|x_t = x]\Big)\ \mathrm{d}u \\
-c(\delta^*(x,z,w)) &= \beta \int_{\underline{w}(x,z)}^{w} \delta^*_w(x,z,u)\Big(\widetilde{W}(x,z,u) - \widetilde{W}(x,z,\underline{w}(x,z))\Big)\ \mathrm{d}u \\
&= \beta \int_{\underline{w}(x,z)}^{w} \delta^*_w(x,z,u)\widetilde{W}(x,z,u)\ \mathrm{d}u - \beta(\delta^* - 1)\mathbb{E}[U(x_{t+1})|x_t = x]
\end{aligned}$$

which can then be substituted into the main equation to get an integral equation for $W(x,z,w)$:

$$\begin{aligned}
W(x,z,w) = u(w) &+ \beta \int_{\underline{w}(x,z)}^{w} \delta^*_w(x,z,u)\Big(\widetilde{W}(x,z,u) - \widetilde{W}(x,z,\underline{w}(x,z))\Big)\ \mathrm{d}u \\
&+ \delta^*\beta\mathbb{E}U(x) \\
&+ (1-\delta^*)\beta\kappa p^*_1 \cdot \mathbb{E}[W(x',z',w')|x,z,w,m{=}1] \\
&+ \beta(1-\delta^*)(1-\kappa p^*_1)\mathbb{E}[W(x',z',w')|m{=}0],
\end{aligned}$$

$$\begin{aligned}
W(x,z,w) = u(w) &+ \beta \int_{\underline{w}(x,z)}^{w} \delta^*_w(x,z,u)\Big(\widetilde{W}(x,z,u) - \widetilde{W}(x,z,\underline{w}(x,z))\Big)\ \mathrm{d}u \\
&+ \delta^*\beta\widetilde{W}(x,z,\underline{w}(x,z)) \\
&+ (1-\delta^*)\beta\widetilde{W}(x,z,w)
\end{aligned}$$

where we note that $\Big(-\delta^*_w(x,z,u)\Big) > 0$. Even in the case where we do not know the $c(\dot{})$ we can still express $W(x,z,w)$ as the solution of an integral equation. This is a Fredholm equation of the second kind. However showing that this expression is indeed a contraction require bounding the derivative $\delta_w^*(x, z, u)$.

C Empirical appendices (online publication)

C.1 Reduce form model

Recall the auxiliary model described in the first section of the paper. Note that u_{jt}^f appears in both the worker and the firm equation. We start by re-writing this model in differences:

$$\Delta y_{j,t} = u_{jt}^f + \Delta y_{jt}^t$$
$$\Delta w_{it} = \tau \cdot u_{j(i,t),t}^f + u_{j(i,t),t}^c + u_{it}^w + \Delta w_{it}^t$$

The auxiliary model presented can be recovered from the following moments:

$$\begin{aligned} Var\left[\Delta w_{it}\right] &= \sigma_{u^c}^2 + \tau^2 \cdot \sigma_{u^f}^2 + \sigma_{u^w}^2 + 2\sigma_{w^t}^2 \\ Var\left[\Delta y_{jt}\right] &= \sigma_{u^f}^2 + 2\sigma_{y^t}^2 \\ Cov\left[\Delta w_{it}, \Delta w_{i,t-1}\right] &= -\sigma_{w^t}^2 \\ Cov\left[\Delta y_{j,t}, \Delta y_{j,t-1}\right] &= -\sigma_{y^t}^2 \\ Cov\left[\Delta y_{j(i,t),t}, \Delta w_{it}\right] &= \tau\sigma_{u^f}^2 \\ Cov\left[\mathbb{E}[\Delta w_{it}|G=1,j], \mathbb{E}[\Delta w_{it}|G=2,j]\right] &= \sigma_{u^c}^2 + \tau\sigma_{u^f}^2 \end{aligned}$$

where $\mathbb{E}[\Delta w_{it}|G = 1, j]$ represents the expectation over co-workers within firm j that were assigned randomly to a group $G = 1$. Table 9 shows all moments and parameters, together with the bootstrapped confidence intervals. Recovering the common shock at the firm level could be accomplished if all firms were large by computing the average wage growth $\mathbb{E}[\Delta w_{it}|j(i,t) = j] =$

$u^c_{jt} + \tau u^f_{jt}$. However this would suffer from incidental parameter bias as in small firms, where the error term in the average of the individual wage growth would not be negligible. Instead, I randomly assign each worker in each firm to group $G \in 1, 2$, then compute the mean wage growth within each firm, each period and each group and finally compute the covariance across firms. I compare this procedure to using the variance of the mean wage growth at the firm by varying the size of the firms.

C.1.1 Co-worker covariance and firm size

There is of course a concern that the correlation between co-workers is affected by the distribution of firm size. There are two different concerns here. First

Table 9: Reduce form model for earnings of stayers

	value	Conf. Interval	
		$q = 0.025$	$q = 0.975$
Moments			
$Var\left[\Delta w_{it}\right]$	5.40e-02	5.24e-02	5.58e-02
$Var\left[\Delta y_{jt}\right]$	8.87e-02	7.38e-02	1.05e-01
$Cov\left[\Delta w_{it}, \Delta w_{i,t-1}\right]$	-1.51e-02	-1.57e-02	-1.46e-02
$Cov\left[\Delta y_{j,t}, \Delta y_{j,t-1}\right]$	-3.12e-02	-3.70e-02	-2.58e-02
$Cov\left[\Delta y_{j(i,t),t}, \Delta w_{it}\right]$	3.67e-04	1.35e-04	5.89e-04
$Cov\left[\mathbb{E}[\Delta w_{it} \| G=1], \mathbb{E}[\Delta w_{it} \| G=2]\right]$	6.89e-04	5.49e-04	7.56e-04
Parameters			
value added transitory $\sigma^2_{y^t}$	3.12e-02	2.58e-02	3.70e-02
value added permanent $\sigma^2_{u^f}$	2.63e-02	1.52e-02	3.75e-02
worker transitory $\sigma^2_{w^t}$	1.51e-02	1.46e-02	1.57e-02
worker permanent individual $\sigma^2_{u^w}$	2.31e-02	2.22e-02	2.39e-02
worker permanent firm level $\sigma^2_{u^c}$	6.84e-04	5.41e-04	7.48e-04
τ parameter	1.39e-02	5.18e-03	2.65e-02
worker permanent value added $\tau^2 \cdot \sigma^2_{u^f}$	5.11e-06	7.28e-07	1.38e-05

the effect might be different at small and large firm. Second there might be an incidental parameter problem for small firms with few workers where the mean wage growth in G_0 and G_1 might be badly estimated.

I run two different checks. First I estimate the covariance by conditioning on firm size being larger than a certain number, and vary this number between 3 and 18. The main estimation uses 10. Second, I keep the sample of firm larger than 10 and I throw away individuals in each firms. I report the estimated co-variance for these in the following table 10.

Table 10: Effect of firm size on estimates

size	firm size		leave-out	
	$Var(\Delta w_{g_1t})$	$Cov(\Delta w_{g_1t}, \Delta w_{g_2t})$	$Var(\Delta w_{g_1t})$	$Cov(\Delta w_{g_1t}, \Delta w_{g_2t})$
3	7.62e-03	1.00e-03	3.60e-03	6.70e-04
6	5.10e-03	7.66e-04	3.74e-03	6.56e-04
9	3.75e-03	6.98e-04	3.84e-03	6.70e-04
12	3.06e-03	6.22e-04	4.06e-03	6.53e-04
15	2.62e-03	5.64e-04	4.24e-03	6.50e-04
18	2.35e-03	5.53e-04	4.30e-03	6.42e-04

The columns give the variance of the wage growth within group 1 and the co-variance between the wage growth between the two groups. When firms are very large, these two should give similar values according to the model. When firms are small however, the within group variance is inflated by the individual variance.

In the first two column I keep only firms with at least *size* workers. We see that this affects both measures very strongly. This could be due to two facts, either because the variance differ with firm size, or because the procedure is strongly affected by the small firms. To see the effect of the second kind, I keep the sample the same as in the base sample, but I throw away individual in each firms. We see that in this case, the measurements are affected but in a much smaller way. We get only a variation of less than 10% from the top of the table to the bottom for our last measure. On the other hand for the direct variance of wage growth at the firm level we get a difference of 20%.

I conclude then that first, the covariance is slightly smaller at lager firm and second that the covariance between wage growth of co-workers is the most stable measure.

C.2 Co-worker covariance - Robustness check

I run a placebo test where I draw the firm identity of each worker randomly with replacement within year and education, to stay close to the data. This removes all dependence between co-workers since they are now randomly assigned to firms. I evaluate the auxiliary earnings model on a 100 replications. At each draw I re-assigned the groups G_1 and G_2. This placebo test delivers a mean of 0.000107 with a standard deviation of 0.00097. The covariance is not significantly different from 0.

C.3 Consumption equivalent

We consider a very simple model where $c_t = exp(\sum_{\tau=0}^{t} \sigma\epsilon_\tau)$ with ϵ_τ a normal distribution. c_t is a unit root process with innovation variance σ^2. We can then compute the present value for the CRRA utility

$$
\begin{aligned}
U = \mathbb{E}\sum_t \beta^t u(c_t) &= \mathbb{E}\sum_t \beta^t \frac{c_t^{1-r}}{1-r} \\
&= \frac{1}{1-r}\mathbb{E}\sum_t \beta^t e^{(1-r)\sum_{\tau=0}^{t}\sigma\epsilon_\tau} \\
&= \frac{1}{1-r}\frac{1}{1-\beta e^{\frac{(1-r)^2\sigma^2}{2}}}
\end{aligned}
$$

which means that our consumption equivalent is given by

$$
\left(\frac{1-\beta}{1-\beta e^{\frac{(1-r)^2\sigma^2}{2}}}\right)^{\frac{1}{1-r}}
$$

Comparison with Guiso, Pistaferri, and Schivardi (2005) In their paper, they fit an auto-regressive process with auto-correlation parameter $\rho = 0.436$. They then report the variances of the permanent shock to earnings to be $\sigma^2 = 0.0091$ in Table 7. To account for the auto-regressive specification, I compute $\tilde{\sigma} = \sqrt{\frac{\sigma^2}{1-\rho^2}}$. This gives me an comparable standard deviation of $\sigma_{GPS} = 0.108$. They also report that the overall standard deviation of wage growth is 0.1245 which is a variance of 0.015 which is much smaller than the value I find in this data set of 0.054.

D Estimation Appendices

D.1 Estimation and inference

Estimation of the parameters is achieved using the method of simulated moments. Given a parameter value ϑ, I can solve for the optimal contract, then simulate a population and compute simulated moments $m_{R,S}(\vartheta)$ where R reflects the size of the simulated sample and S the number of replications. Second, I also compute the same moments $\hat{m}_N$ on the data where N reflects the data sample size. For a given weighting matrix W, the estimate is defined as:

$$\hat{\vartheta}_N = \inf_{\vartheta} \left[\hat{m}_N - m_{R,S}(\vartheta)\right]^T W \left[\hat{m}_N - m_{R,S}(\vartheta)\right].$$

Under the assumption that $\sqrt{N}\left(m_N - m(\vartheta_0)\right) \xrightarrow{d} \mathcal{N}(0,\Sigma)$ where θ_0 is the true parameter and neglecting simulation noise by assuming that N,S is large enough such that $m_{R,S}(\vartheta) = m(\vartheta)$, the asymptotic variance for the parameters is given by:

$$\sqrt{N}\left(\widehat{\vartheta}_N - \vartheta_0\right) \xrightarrow{d} \mathcal{N}(0, J^{-1}\Omega J^{-1})$$

where

$$\begin{aligned}
\Omega &= \lim_{N\to\infty} \quad \left[\frac{\partial m(\vartheta_0)}{\partial \vartheta^T}\right]^T W\Sigma W \left[\frac{\partial m(\vartheta_0)}{\partial \vartheta^T}\right] \\
J &= \lim_{N\to\infty} \quad \left[\frac{\partial m(\theta_0)}{\partial \vartheta^T}\right]^T W \left[\frac{\partial m(\vartheta_0)}{\partial \vartheta^T}\right].
\end{aligned}$$

In this context we have multiple dimensions that will grow with the sample size in the asymptotic that we are considering. The is the number of workers, the time dimension and the size of firms. The previous section established the identification in short panel, allowing us to think at fixed T. However, several of the moments of interest are defined at the firm level, such as the covariance between wage growth of co-workers, or the co-variance between wage growth and value added growth. This means that the relevant asymptotic in terms of number of firms. Hence N here refers to the number of firms.

To allow for the dependence at the firm level as well as serial correlation in earnings over time, the matrix Σ is estimated using block bootstrap where the block is a given firm, and the off diagonal terms are set to 0. The weighting matrix W is chosen to emphasize the moments of interests such as the covariance between co-worker wage growth. The default weight is the inverse of the value of the moments estimated from the data. The weights for $Var(\Delta \log w^p_{it}|EE)$, $Cov(\Delta \log w^p_{it}, \Delta \log w^p_{it-1}|EE)$, $\mathbb{E}(\log w_{it}|Hed) - \mathbb{E}(\log w_{it}|Led)$ and $Cov(\Delta \log w_{it}, \Delta \log w_{jt}|EE)$ are increased by a factor three.

The objective function is minimized using an exploration method inspired by Gibbs sampling where at each step one parameter is updated conditional on the other ones. Once an optimal value is found, $\partial m(\vartheta)/\partial\vartheta$ is evaluated locally. Finally the sandwich formula is applied.

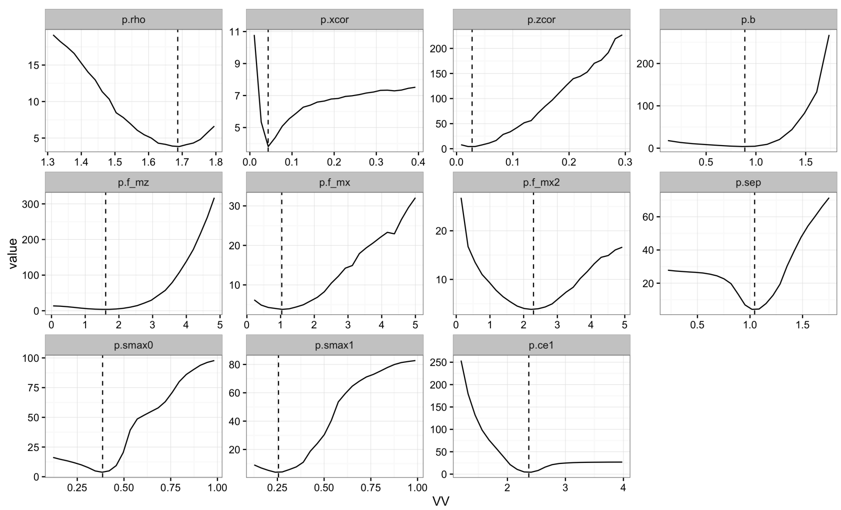
D.1.1 Numerical solution to the model

I briefly described how I solve for the solution of the problem. I use n_z=7, n_{x0}=3, n_{x_1}=6 points of support for the productivity types, this gives a total of 126 different productivities. Given a parameter value, I first solve the unemployed worker problem and the optimal application behavior given x and promised value W.

I then solve the firm problem without the incentive problem and use n_v=200 points of support for ρ at each x_0, x_1, z combination. I use value function iteration and iterate until the error comes below 10^{-6} on the mean square error for the Bellman equation. This is a very simple problem since ρ stays fixed.

Next, I use the first best solution as starting value for the optimal contract. I use value function iteration again, and at each step I have to solve for the

Figure 5: Slices of the objective function



Notes: This figure plots the objective function as each of the parameter moves away from the estimate reported in the paper.

optimal γ in the following recursive problem:

$$\mathcal{P}(x, z, \rho) = \inf_{\gamma} \sup_{w,W} f(x, z) - w + \rho\,(u(w_i) + \tilde{r}(x, W)) \\ - \beta\gamma\tilde{p}(x, W) \cdot W + \beta\tilde{p}(x, W)\mathbb{E}\mathcal{P}(x', z', \gamma), \quad (8)$$

I proceed in the following way. At each x, z, ρ I evaluate the values on a grid with 400 points of support. I then look for the point which makes the first order condition EQ-FOC the closest to zero. I then use linear interpolation to find the exact root. I iterate on the recursive problem until convergence to 10^{-6}. I then simulate a panel of length 200 quarters and compute moments on the last 25 quarters only. I keep the seed and draws the same across evaluations. I simulate $50,000$ workers, and verified that it was enough to have only a small simulation error (by re-simulating with a different seed).