



CAEPR Working Paper
#2008-006

Interfirm Mobility, Wages, and the Returns to Seniority and Experience in the U.S.

Moshe Buchinsky
UCLA, CREST-INSEE and NBER

Denis Fougère
CNRS, CREST-INSEE, CEPR and IZA

Francis Kramarz
CREST-INSEE, CEPR and IZA

Rusty Tchernis
Indiana University Bloomington

April 3, 2008

This paper can be downloaded without charge from the Social Science Research Network electronic library at: <http://ssrn.com/abstract=671022>.

The Center for Applied Economics and Policy Research resides in the Department of Economics at Indiana University Bloomington. CAEPR can be found on the Internet at: <http://www.indiana.edu/~caepr>. CAEPR can be reached via email at caepr@indiana.edu or via phone at 812-855-4050.

©2008 by Moshe Buchinsky, Denis Fougère, Francis Kramarz and Rusty Tchernis. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Interfirm Mobility, Wages, and the Returns to Seniority and Experience in the U.S.[†]

Moshe Buchinsky[‡]

UCLA, CREST-INSEE and NBER

Denis Fougère

CNRS, CREST-INSEE, CEPR and IZA

Francis Kramarz

CREST-INSEE, CEPR and IZA

Rusty Tchernis

Indiana University

March, 2008

Abstract

In this paper we follow on the seminal work of Altonji and Shakotko (1987) and Topel (1991) and reinvestigate the returns to seniority in the U.S. These papers specify a wage function, in which workers' wages can change through two channels: (a) returns to their seniority; and (b) returns to their labor market experience. We start from the same wage equation as in previous studies, and, following our theoretical model, we explicitly include a participation-employment equation and an interfirm mobility equation. The employment and mobility decisions define the individual's experience and seniority. Because experience and seniority are fully endogenized, we introduce into the wage equation a summary of the workers' entire career and past jobs. The three-equation system is estimated simultaneously using the Panel Study of Income Dynamics (PSID). For all three education groups that we study, returns to seniority are quite high, even higher than what was previously obtained by Topel. On the other hand, the returns to experience appear to be similar to those previously found in the literature.

[†]We thank Gary Chamberlain, Andrew Foster, Oded Galor, Hank Farber, Richard Freeman, Jinyong Hahn, Mark Machina, Magali Ponçon-Beffy, Jean-Marc Robin, John Rust, Bernard Salanié, Duncan Thomas, and Robert Topel for their comments and discussions. We also thank the editor as well as three referees whose comments and suggestions improved the paper considerably. Participants of seminars and conferences in Amsterdam, Brown, Crest (Malinvaud seminar), ESEM 2001 in Lausanne, Harvard, Maryland, Michigan, Pennsylvania, Prague, Tel-Aviv, Toulouse, UCLA, UCSD, Wisconsin, and Yale, provided very helpful comments. Remaining errors are ours.

[‡]Corresponding author, Department of Economics, University of California, Los Angeles, CA 90095-1477. Email: buchinsky@econ.ucla.edu.

1 Introduction

There has been an ongoing debate in the literature about the returns to human capital, particularly in the United States. Part of the debate is in assessing the magnitude of the returns to the various components of human capital, namely the returns to general and specific human capital (as defined in Becker (1964)). While there is a consensus about the role of education in the formation of human capital, there is little agreement about the role of firm-specific human capital. That is, the human capital specific to the firm at which a worker is employed. Typically in the literature firm-specific human capital is proxied by tenure on a specific job. Consequently, it is hard to disentangle the returns to firm-specific human capital from the returns to general human capital, as is measured by one's labor market experience.

In general, workers' wages can change through two, not necessarily mutually exclusive, channels. Workers can stay in the same firm for some years and collect the returns to their firm-specific human capital (seniority). Alternatively, they can change jobs because of an outside wage offer that exceeds the wage at their current job. There has been an ongoing debate in the literature as to which of the two alternatives contribute to observed wage growth among individuals in the U.S. Topel (1991) claims that there are significant returns to seniority, while Altonji and Shakotko (1987) and Altonji and Williams (2005) claim that there are very little returns to seniority, if at all, in the U.S.

In this study, we revisit the question regarding the magnitude of the returns to seniority in the U.S. and offer a new perspective. Our starting point is the same as that taken by virtually all papers in the literature on this topic. That is, we adopt the general Mincer's wage specification. To account for the endogenous decisions of participation and mobility, we extend the model of Hyslop (1999) by adding a dynamic optimization search perspective. This model gives rise to two *structural* decision equations, namely: (1) a participation-employment equation,¹ and (2) an interfirm mobility equation. In this approach, experience and seniority are fully endogenized because they are direct outcomes of the employment and interfirm mobility decisions, respectively. Hence, in order to obtain consistent estimates for the parameters of the wage function, one must estimate it jointly with the employment and mobility equations. Estimating this model accounts for the potential selection biases that stem from the (endogenous) employment and mobility decisions. Because the selection effects are quite complex, it is not possible to ex-ante assess the overall impact on estimated returns to experience and seniority; this is largely an empirical question to which we devote the analysis in this paper.

To control for other potential sources of bias, we also allow the three estimated equations to include person-specific random correlated effects as well as idiosyncratic components. To be as flexible as possible, we impose no restrictions on the covariance structure between the various stochastic elements. It is well known in the literature that modeling unobserved heterogeneity is generally very important in capturing aspects of human behavior. It turns out that this is of vital importance in our case as well. In particular, unobserved heterogeneity creates important links

¹In both the theoretical model and the empirical analysis, participation is the same as employment. We therefore use these two terms interchangeably throughout the paper.

between individuals' various decisions that, if ignored, can lead to severe biases in the estimated parameters of interest. That is, the employment and mobility decisions are correlated with the outcome of interest, i.e., wage, in part through the correlation between the person-specific unobserved components. In addition, the employment and mobility equations include lagged decisions as explanatory variables, controlling for potential state dependence that is implied by the model described below.

Our approach offers a unified framework with which one can address the re-examined results found in the literature. This is because the statistical assumptions we make—along with endogenizing seniority and experience—incorporate the most important elements from previous studies, particularly those by both Topel (1991) and Altonji and Williams (2005).²

We use data from the Panel Study of Income Dynamics (PSID) to estimate the model for three educational groups: High school dropouts, high school graduates, and college graduates. We adopt a Bayesian approach and employ Markov Chain Monte Carlo (MCMC) methods which allows us to estimate the joint posterior distribution of all the model's parameters.

The results indicate that, while the estimated returns to experience are somewhat higher than those previously found in the literature, they are of similar magnitude. In complete contrast, the estimates of the returns to seniority are much higher than those previously obtained, even higher than those obtained by Topel (1991). Moreover, they are large and significant for all education groups. Consequently, our estimates of the total within-job wage growth are significantly higher than Topel's estimates and all other results reported in the literature (e.g., Altonji and Shakotko (1987), Abraham and Farber (1987), and Altonji and Williams (2005)).

The results also demonstrate the vital importance of explicitly incorporating the employment and the mobility decisions, which, in turn, define experience and seniority. We show that the differences between our estimates and those obtained in previous studies are the direct consequence of two important factors. The first factor is insufficient control for unobserved heterogeneity prevailing in the various decisions. This factor has important implications for the implied choices of alternative career paths. The other factor is the lack of direct modeling of the job-specific component of wages. In the formulation of the wage equation in this study, we introduce a function that serves as a summary statistic for the individual's specific career path. This function permits discontinuous jumps in individuals' wages when they change jobs. We find that the estimates change markedly when this function is included. This clearly indicates, as is illustrated below, that the timing of a job change in one's career matters a great deal.

Because our findings are very different from those previously obtained in the literature, we provide a number of robustness checks. For example, we start the estimation centering the prior distributions for the relevant parameters around the results obtained by Altonji and Shakotko (1987) and by Topel (1991). It turns out that the results obtained are identical using either prior. Also, we use a somewhat different sample than the one used by Topel (1991) and Altonji and Shakotko (1987). However, we find very clear evidence that the substantially distinct results we obtain are not driven

²There are a number of other papers in the literature, including Dustmann and Meghir (2005) and Neal (1995), who analyzed similar questions but took different routes. We discuss them below.

by this fact. When we apply the methods adopted by Topel (1991) and Altonji and Shakotko (1987), we are able to very closely replicate their results using our data extract. In essence, the model we develop here gives rise to both models, even though our model does not explicitly nest the other two models. Hence, we are able to directly examine the assumptions previously made in the literature and their implications on the results obtained.

The remainder of the paper is organized as follows. In Section 2, we provide a brief review of the literature and highlight the crucial differences between the various approaches. Section 3 presents the model, which extends Hyslop’s (1999) while incorporating other vital features from the search literature. Section 4 follows with the econometric specification. We also briefly discuss the key elements of the numerical techniques used for estimating the model. A brief discussion of the data extract used in this study is provided in Section 5. Section 6 presents the empirical results and discusses their implications. Section 7 follows with some concluding remarks. There are also three appendices. Appendix A provides some mathematical details and proofs related to the model. Appendix B provides the necessary details about the numerical algorithm used in computing the posterior distribution of the model’s parameters. Appendix C presents a descriptive analysis that gives rise to the modeling approach taken here and specifically the introduction of a function summarizing an individual’s previous career path in the wage equation.

2 Literature Review

There has been a long lasting debate in the literature on the magnitude of returns to seniority in the U.S. Topel (1991)’s results, indicating that there are large returns to seniority, stood in stark contrast with previous literature. The most notable papers of Altonji and Shakotko (1987) and Abraham and Farber (1987) find virtually no returns to seniority in the U.S.

Topel (1991) finds strong evidence that the costs of displacement are highly correlated with prior job tenure. He goes on to argue that this phenomenon can stem from one of two potential explanations: (a) wages rise with seniority; or (b) tenure merely acts as a proxy for the quality of the job match.³ To examine this phenomenon, he uses the following prototype model of wage determination:

$$y_{ijt} = X_{ijt}\beta_1 + T_{ijt}\beta_2 + \epsilon_{ijt}, \tag{1}$$

where y_{ijt} is the log wage of individual i in job j at time t , X_{ijt} denotes experience, and T_{ijt} denotes seniority. The residual term ϵ_{ijt} is decomposed into three components:

$$\epsilon_{ijt} = \phi_{ijt} + \mu_i + v_{ijt}, \tag{2}$$

where ϕ_{ijt} is specific to the individual-work (or job) pair, μ_i is a term that reflects the individual’s ability, while v_{ijt} is an idiosyncratic term representing market-wide random shocks and/or measurement error. The main problem in estimating the parameters of interest β_1 and β_2 stems from

³An additional possibility, not explored by Topel, is that workers with long tenures are simply more able.

the fact that ϕ_{ijt} is likely to be correlated with experience and/or tenure. In consequence, a simple ordinary least-squares (OLS) regression on (1) would give a biased estimate, say $\widehat{\beta}_2$, for β_2 . Topel provides a convincing argument regarding the composition of “movers” and “stayers” in the data set. If the returns to seniority are positive, i.e., $\beta_2 > 0$, then $\widehat{\beta}_2$ will be downward biased.

It is crucial to note that, under Topel’s approach, experience at the entry level is assumed to be exogenous and hence uncorrelated with the error terms. Under this assumption, Topel is able to obtain an unbiased estimate for $\beta_1 + \beta_2$ and an upward biased estimate for β_1 (due to selection bias induced by not modeling the mobility decisions). Hence, he argues that his estimate of β_2 , $\widehat{\beta}_2 = .0545$, provides a lower bound for the returns to seniority.⁴

In contrast, Altonji and Shakotko (1987) (AS, hereafter) use an instrumental variables approach assuming (in Topel’s notation) that $\phi_{ijt} = \phi_{ij}$, i.e., the individual job-specific term is time-invariant. Under this assumption, deviation of seniority from its average in a specific job is a valid instrument for seniority. Since this method is (as shown by Topel (1991)) a variant of his two-step approach, it is not surprising that the estimate for $\beta_1 + \beta_2$ obtained by AS is similar to that obtained by Topel. Nevertheless, AS’s procedure appears to induce an upward bias in the IV estimate for β_1 , and hence a downward bias in the estimate for β_2 . This problem is potentially magnified by a measurement error problem in the tenure data that exists in AS. Differences in treatment of time trends in the regression is yet another factor contributing to substantial differences in the estimates of the returns to tenure.⁵

Abraham and Farber (1987) (AF, hereafter) have a different set of assumptions. In particular, they use completed tenure to proxy unobserved dimensions of the individual’s, or job’s, quality. A problem with their approach is that many of the workers in the extract they use have censored spells of employment.⁶

Altonji and Williams (2005) (AW, hereafter) specify a model closer in spirit to Topel’s (1991) model. However, the two approaches do differ in some meaningful ways, leading to substantial differences in the estimated returns to seniority. AW rely crucially on the assumption that the match effect ϕ and time are independent (i.e., $\text{Cov}(t, \phi) = 0$), conditional on experience (or experience and tenure). This assumption is problematic, especially in cases when workers have had more time to find jobs with higher ϕ . Moreover, t may also be correlated with μ in (2) because of changes in the sample composition.

One important conclusion that comes up from both Topel (1991) and AW is that individual heterogeneity is an important factor of the wage growth process. Hence, it appears that a large part of the reduction in the upward bias of the estimate found by Topel (1991) is due to a reduction in the bias that stems from individual heterogeneity.

⁴Topel (1991) also examines two additional sources of potential biases in the estimates of $\beta_1 + \beta_2$. Nevertheless, he finds that accounting for these potential biases has a very small effect on the estimate for β_2 .

⁵Topel uses a specific index for the aggregate changes in real wages using data from the CPS., while AS use a simple time trend. Consequently, growth of quality of jobs, because of better matches over time, will cause an additional downward bias in the estimate for β_2 .

⁶Another source of difference arises because AF use a quadratic in experience when estimating the log wage equation, whereas AS and Topel use a quartic specification.

In a recent paper, Dustmann and Meghir (2005) (DM, hereafter) allow for three different sources of returns due to accumulation of specific human capital, namely experience, sector-specific seniority, and firm-specific seniority.⁷ To estimate the returns to experience, they focus their attention on data of displaced workers in their new jobs, assuming that such workers cannot predict closure of an establishment (more than a year in advance).⁸ Furthermore, it is crucial to assume that displaced workers have preferences for work similar to those that induce their sectorial choices. In consequence, controlling for the endogeneity of experience also controls for the endogeneity of sector tenure. In a subsequent step of their analysis, DM estimate two reduced-form equations, one for experience and the other for participation. The residuals from these two regressions are used as regressors in the wage regression of the displaced workers (and a similar procedure is applied for seniority). This procedure allows DM to account for possible sample selection biases that are induced by restricting attention to only the individuals staying with their current employer. In this sense, our approach is similar to theirs. Using data from Germany and the U.S., DM find that the returns to tenure for both skilled and unskilled workers are large. While the estimated returns to sector-specific tenure are much smaller, they are also statistically significant.

In his survey paper, Farber (1999) notes the importance of modelling some specific features of the mobility process, an element which is generally missing from the papers discussed above. For instance, he shows that in the first few months of a job, there is an increase in the probability of separation, while this probability decreases steadily thereafter. To fully understand this process, person heterogeneity and duration dependence must be distinguished. Farber gives strong evidence that contradicts the simple model of pure unobserved heterogeneity. He also finds strong evidence that firms choose to lay off less senior workers, suggesting that specific firm-capital matters, as theory implies. Furthermore, he finds that job losses result in substantial permanent earnings losses (see also Gibbons and Katz (1991) for a further discussion of this point).

In our investigation of the returns to seniority and experience in the U.S., we adopt some of the specifications described above, while expanding the model in new directions, namely modelling of the participation and mobility decisions directly.

3 The Model

In this section, we consider a theoretical model that, under some sufficient conditions, generates first-order state dependence in the participation and mobility processes. Whether or not state dependence plays a major role is an empirical question that is examined thoroughly below. The model adopted is an extension of Hyslop's (1999) model. We augment his model with a key feature from the search literature, in that we allow a worker to move directly from one job to another. We show that job-to-job transitions may occur under some sufficient (but fairly general) conditions

⁷Neal (1995) and Parent (1999, 2000) also focus their investigation on the importance of sector- and firm-specific human capital. We abstract from the sector-specific returns in order to avoid modeling sectorial choices, as yet another equation, and focus on firm-to-firm mobility.

⁸For more on the empirical findings in the literature on displaced workers see also Addison and Portugal (1989), Jacobson, LaLonde, and Sullivan (1993), and Farber (1999).

that generate first-order dependence in the participation and mobility processes. These conditions involve the values and properties of three main parameters of the model, namely the search cost, the mobility cost, and the utility function. The individual decision to stay either non-employed or employed in the same firm for at least one more period, as well as the decision to move from one firm to another, are then shown to depend on the existence and the values of different reservation wages. More precisely, we show that, in each of three labor market states (non-employment and employment with and without interfirm mobility), there exist three threshold wage levels that generate the decision either to stay in the same state for at least one more period or to make a transition towards one of the two other states in the next period. The values of these reservation wages are state-dependent, which means that their values depend on the state actually occupied by the worker. This theoretical result allows us to build a structural econometric model with two selection decisions (one for participation and one for interfirm mobility) and a wage equation. The structure of this econometric model, which is different from a switching regression model with two distinct wage equations (one for interfirm movers and one for stayers in the firm), is a direct consequence of our specific assumptions.

The wage function: Our theoretical model builds on the same general specification of the wage function that has been adopted in the literature. That is, the observed log wage equation for individual i in job j at time t is:

$$\begin{aligned} w_{it} &= w_{it}^* \cdot \mathbf{1}(y_{it} = 1), \quad \text{and} \\ w_{ijt}^* &= x'_{wit} \delta_0 + \varepsilon_{ijt}, \end{aligned} \tag{3}$$

where $\mathbf{1}(\cdot)$ is the usual indicator function and $y_{it} = 1$ if the i th individual participates in the labor force at time t , and $y_{it} = 0$ otherwise. That is, the wage offer w_{it}^* is observed only for the individuals who decide to work. The term x_{wit} is a vector of observed characteristics—including education, labor market experience, and seniority (or tenure)—of the individual in their current job, just as in Topel (1991), AS, and AW. We differ in our modeling of the stochastic term ε_{it} . We decompose ε_{it} into three components:

$$\varepsilon_{ijt} = J_{ijt}^W + \alpha_{wi} + \xi_{it}. \tag{4}$$

The term α_{wi} is a person-specific correlated random effect, analogous to μ_i in (2), while ξ_{it} is a contemporaneous idiosyncratic error term. The term J_{ijt}^W , analogous to the term ϕ_{ijt} in (2), serves as a summary statistic for the individual's work history. More specifically, J_{ijt}^W is the sum of all discontinuous jumps in one's wage that resulted from all job changes until date t .⁹ The jumps are allowed to differ depending on the level of labor market experience and seniority at the points of

⁹In Section 6, we provide strong evidence from the descriptive statistics of the data that justifies the introduction of the J^W function and the specific modelling of the function adopted here.

job changes, namely¹⁰

$$J_{ijt}^W = (\phi_0^s + \phi_0^e e_{i0}) d_{i1} + \sum_{l=1}^{M_{it}} \left[\sum_{k=1}^4 (\phi_{k0} + \phi_k^s s_{t-1} + \phi_k^e e_{t-1}) d_{kit_l} \right], \quad (5)$$

where d_{1it_l} equals 1 if the l th job of the i th individual lasted less than a year and equals 0 otherwise. Similarly, $d_{2it_l} = 1$ if the l th job of the i th individual lasted between 2 and 5 years, and equals 0 otherwise, $d_{3it_l} = 1$ if her l th job lasted between 6 and 10 years, and equals 0 otherwise, $d_{4it_l} = 1$ if her l th job lasted more than 10 years and equals 0 otherwise. The ϕ -coefficients do not vary across individuals. Finally, M_{it} denotes the number of job changes by the i th individual who is in job j at time t (not including the individual's first sample year). If an individual changed jobs in his/her first sample year then $d_{i1} = 1$, and $d_{i1} = 0$ otherwise. The quantities e_t and s_t denote the experience and seniority in year t , respectively.¹¹

Note that the J_{ijt}^W function generalizes ϕ_{ijt} , and it captures the initial conditions specific to the individual at the start of a new job. The inclusion of job market experience in previous jobs as a determinant of the initial earnings at the time of a job change allows one to distinguish between displaced workers, who went through a period of non-employment after displacement, from workers who move directly from one job to another. Furthermore, the inclusion of the seniority level at the end of each of the past jobs allows us to control for the quality of the previous job matches. Whether or not the frequency of changing jobs and the individual's labor market attachment matters is an empirical question that we address below.¹²

Timing of the decisions:

We assume that a worker who is employed in period t receives a wage offer from his/her current firm. Given this offer, he/she may decide at the end of period t to either stay in the firm or leave and move to a different firm, or alternative to become non-employed.¹³ At the end of period t , he/she does not know with certainty the wage that the new firm will offer him/her but can form an

¹⁰In a different context Light and Ureta (1995) also address the timing issue, but with respect to experience. They discuss the role of experience in the early stages of one's career and highlight the importance of controlling for the exact timing of work experience when estimating wage regressions.

¹¹Note that the specification for the term J_{ijt}^W introduces thirteen different regressors in the wage equation. These regressors are: a dummy for job change in year 1, experience in year 0, the numbers of job changes that lasted up to one year, between 2 and 5 years, between 6 and 10 years, or more than 10 years, seniority at last job change that lasted between 2 and 5 years, between 6 and 10 years, or more than 10 years, and experience at last job change that lasted less than one year, between 2 and 5 years, between 6 and 10 years, or more than 10 years.

¹²Also, note that J_{ijt}^W is individual-job specific. In general, there are several ways for defining a job. Our definition of a job is that of a particular employment spell in one's career. Hence, it is possible that different individuals will have the same values for J_{ijt}^W , even though they may not be employed in the same firm. For example, two workers could have entered the labor market at the same date after leaving school, and both could have occupied two different jobs of similar lengths (say, six months each) during the following year, before moving to a new firm. Nevertheless, this definition of a job is consistent with our modelling approach.

¹³Strictly speaking, our model does not allow us to distinguish between layoffs and quits. Thus a more appropriate reference to use is separation. A separation happens when an unexpected low wage is offered by the current employer in period t . Such an event can be interpreted as a negative exogenous shock on the individual's productivity in the current firm. In turn, this will induce a move to the non-employment state or, alternatively, to another firm. This interpretation is consistent with the available information in our data set, which does not allow us to distinguish between involuntary and voluntary job changes.

expectation about that wage using the wage determination process given in (3)-(5). For simplicity, we assume that in this case an outside wage offer will arrive with probability 1 in the next period.¹⁴ If a worker decides to move to a new firm at the end of period t , then he/she incurs the cost c_M , paid at the beginning of period $t + 1$.¹⁵ As it turns out, this assumption induces first-order state dependence in the participation and mobility processes. While it is obvious that some components of these costs are paid at time t , it is just as obvious that other components of these costs are paid at period $t + 1$. The need for adjustments induces some moving costs that are incurred only after moving. Typically, costs that will be incurred at time $t + 1$ are transaction costs and non-monetary costs associated with reconstructing social capital in a new workplace and family environment.¹⁶

The model's structure:

At any given period t a non-participant may search for a job at a cost γ_1 per period, paid at the beginning of the next period. This search cost is assumed to be strictly lower than the mobility cost c_M . While this assumption is not testable, it is reasonable to assume that moving to a new firm entails higher costs than getting a job when unemployed. This is because an individual who works has first to quit and then to incur the moving cost which is at least as high as the cost of moving from unemployment to a new job. Indeed, the literature provides strong support for this assumption. For example, Hardman and Ioannides (2004) argue that the high moving costs, both in terms of out-of-pocket costs and the loss of location-specific social and human capital, explain why there are infrequent moves of individuals. We further assume that hours of work are constant across all jobs, so that we can concentrate only on the extensive margin of the participation process y_t , which takes the value 1 if the individual participates in period t and takes the value 0 otherwise.

Each individual is assumed to maximize the discounted present value of the infinite lifetime intertemporally separable utility function given by

$$U_t = \sum_{s=t}^{\infty} \beta^{s-t} E_t [u(C_s, y_s; X_s)], \quad (6)$$

where $u(\cdot)$ is the current period flow of utility from consumption, C_t , and leisure, $l_t = 1 - y_t$, conditional on a vector of (observed and unobserved) exogenous individual characteristics X_t . The term β is the discount factor. The notation $E_t(\cdot)$ denotes that the expectation is taken conditional on the information available at time t . Assuming neither borrowing nor lending, the period-by-

¹⁴One can also account for the costs associated with searching for an outside wage offer. Since the main theoretical results hold under this specification as well, we assume for simplicity that there are no such costs. Equilibrium search models with costly on-the-job search and outside job offers are developed, for instance, in Bontemps, Robin and Van den Berg (1998) and Postel-Vinay and Robin (2002). However, Nagypal (2005) shows that such models with on-the-job search do not generally match the extent of job-to-job transitions. Consequently, she develops an alternative theoretical framework that presents some similarities with our model. In particular, Nagypal (2005) incorporates a stochastic process that causes the value of a job to the worker to decrease at times, predicting that workers with a lower job value have a higher probability of entering unemployment. Nevertheless, further comparison between our approach and search equilibrium models is still beyond the scope of our paper.

¹⁵It is assumed that a departure from a firm has to be announced at the end of period t . Moreover, there is no possibility of recall from the current employer.

¹⁶For ease of exposition, we assume that all the costs are incurred at $t + 1$. This assumption does not change any of the conclusions derived here.

period budget constraint is given by

$$C_t = z_t + w_t y_t - \gamma_1 (1 - y_{t-1}) - c_M m_{t-1}, \quad (7)$$

where the price of consumption in each period is normalized to 1, z_t is non-labor income, w_t is the individual's wage offer, m_{t-1} is an indicator variable that takes the value 1 if the individual moved to a new job at the end of $t - 1$ and takes the value 0 otherwise.

By virtue of Bellman's optimality principle, the value function at the beginning of period t , given participation y_{t-1} and mobility m_{t-1} in period $t - 1$, is given by

$$V_t(y_{t-1}, m_{t-1}; X_t) = \max_{y_t, m_t} \{u(C_t, y_t; X_t) + \beta E_t [V_{t+1}(y_t, m_t; X_{t+1})]\}. \quad (8)$$

If the individual does not participate at $t - 1$, i.e., $y_{t-1} = 0$ (and hence $m_{t-1} = 0$), the value function at the beginning of period t is

$$V_t(0, 0; X_t) = \max_{\{y_t, m_t\}} [V_t^0(0, 0; X_t), V_t^1(0, 0; X_t), V_t^2(0, 0; X_t)], \quad \text{where} \quad (9)$$

$$V_t^0(0, 0; X_t) = u(z_t - \gamma_1, 0; X_t) + \beta E_t [V_{t+1}(0, 0; X_{t+1})],$$

$$V_t^1(0, 0; X_t) = u(z_t + w_t - \gamma_1, 1; X_t) + \beta E_t [V_{t+1}(1, 0; X_{t+1})], \quad \text{and}$$

$$V_t^2(0, 0; X_t) = u(z_t + w_t - \gamma_1, 1; X_t) + \beta E_t [V_{t+1}(1, 1; X_{t+1})],$$

where $V_t^0(0, 0; X_t)$ denotes the value of non-participation in period t , $V_t^1(0, 0; X_t)$ denotes the value of participating without moving at the end of period t , and $V_t^2(0, 0; X_t)$ denotes the value of participating and moving at the end of period t .

To better understand (9), note that:

1. Non-participation in period t implies that $y_t = 0$ and hence $m_t = 0$. Thus, from (7), the budget constraint in period t is reduced to $C_t = z_t - \gamma_1$, and the value function at the beginning of period $t + 1$ is $V_{t+1}(0, 0; X_{t+1})$.
2. Participation in period t without moving at the end of period t corresponds to $y_t = 1$ and $m_t = 0$. Thus, the budget constraint in this period is $C_t = z_t + w_t - \gamma_1$, and the value function at the beginning of period $t + 1$ is $V_{t+1}(1, 0; X_{t+1})$.
3. Participation in period t and moving at the end of this period corresponds to $y_t = 1$ and $m_t = 1$. Thus, the budget constraint in period t is $C_t = z_t + w_t - \gamma_1$, and the value function at the beginning of period $t + 1$ is $V_{t+1}(1, 1; X_{t+1})$.

For a "stayer", i.e., a participant who stays in his/her job at the end of period $t - 1$ (i.e., $y_{t-1} = 1$

and $m_{t-1} = 0$), the value function at the beginning of period t is

$$\begin{aligned} V_t(1, 0; X_t) &= \max_{y_t, m_t} \{V_t^0(1, 0; X_t), V_t^1(1, 0; X_t), V_t^2(1, 0; X_t)\}, \quad \text{where} \quad (10) \\ V_t^0(1, 0; X_t) &= u(z_t, 0; X_t) + \beta E_t V_{t+1}[(0, 0; X_{t+1})], \\ V_t^1(1, 0; X_t) &= u(z_t + w_t, 1; X_t) + \beta E_t [V_{t+1}(1, 0; X_{t+1})], \quad \text{and} \\ V_t^2(1, 0; X_t) &= u(z_t + w_t, 1; X_t) + \beta E_t [V_{t+1}(1, 1; X_{t+1})]. \end{aligned}$$

Similarly, for a “mover”, i.e., a participant who moves to a new job at the end of $t - 1$, the value function at the beginning of period t is

$$\begin{aligned} V_t(1, 1; X_t) &= \max_{y_t, m_t} \{V_t^0(1, 1; X_t), V_t^1(1, 1; X_t), V_t^2(1, 1; X_t)\}, \quad \text{where} \quad (11) \\ V_t^0(1, 1; X_t) &= u(z_t - c_M, 0; X_t) + \beta E_t [V_{t+1}(0, 0; X_{t+1})], \\ V_t^1(1, 1; X_t) &= u(z_t + w_t - c_M, 1; X_t) + \beta E_t [V_{t+1}(1, 0; X_{t+1})], \quad \text{and} \\ V_t^2(1, 1; X_t) &= u(z_t + w_t - c_M, 1; X_t) + \beta E_t [V_{t+1}(1, 1; X_{t+1})]. \end{aligned}$$

The interpretations of (10)–(11) are similar to those of (9). Transitions to the non-participation state will occur if the wage offer in period t is less than the minimum of two alternative reservation wages, namely the wage levels that equate the value function of non-participation with that of participation, without and with interfirm mobility. These two reservation wages, denoted by $w_{01,t}^*$ and $w_{02,t}^*$, respectively, are defined implicitly by

$$V_t^0(0, 0; X_t) = V_t^1(0, 0; X_t | w_{01,t}^*) = V_t^2(0, 0; X_t | w_{02,t}^*), \quad (12)$$

where

$$\begin{aligned} V_t^1(0, 0; X_t | w_{01,t}^*) &= u(z_t + w_{01,t}^* - \gamma_1, 1; X_t) + \beta E_t [V_{t+1}(1, 0; X_{t+1})] \quad \text{and} \\ V_t^2(0, 0; X_t | w_{02,t}^*) &= u(z_t + w_{02,t}^* - \gamma_1, 1; X_t) + \beta E_t [V_{t+1}(1, 1; X_{t+1})]. \end{aligned}$$

It follows immediately from (12) then that

$$w_{02,t}^* \leq w_{01,t}^* \quad \Leftrightarrow \quad E_t [V_{t+1}(1, 0; X_{t+1})] \leq E_t [V_{t+1}(1, 1; X_{t+1})].$$

From here, our analysis extends the one set forth by Burdett (1978). As in Burdett (1978), we assume that in each period t

$$\frac{d}{dw_t} V_t^2(1, k; X_t) \leq \frac{d}{dw_t} V_t^1(1, k; X_t), \quad k = 0, 1. \quad (13)$$

If $w_{02,t}^* < w_{01,t}^*$, then there exists a wage value, denoted $w_{03,t}^*$, that equates the value function of participation without interfirm mobility in period t with that of participation with interfirm

mobility in period t (see Figure A.1 in Appendix A.1). This wage threshold is defined implicitly by

$$V_t^1(0, 0; X_t | w_{03,t}^*) = V_t^2(0, 0; X_t | w_{03,t}^*). \quad (14)$$

Then the decision rule for a non-participant is to accept any wage offer greater than $w_{02,t}^*$, and either to move to another firm in the next period, if the current wage is between $w_{02,t}^*$ and $w_{03,t}^*$, or to stay in the firm for at least one more period if the current wage is greater than $w_{03,t}^*$. If $w_{01,t}^* < w_{02,t}^*$, then the optimal strategy for a non-participant is to accept any wage offer greater than $w_{01,t}^*$, and to stay in the firm for at least one more period (see Figure A.2 in Appendix A.1).¹⁷

The reservation wages for a stayer in period $t - 1$, denoted by $w_{11,t}^*$ and $w_{12,t}^*$, respectively, are the wage levels that equate the value function of non-participation with those of participation, with and without interfirm mobility, in period t . These reservation wages are defined implicitly by

$$V_t^0(1, 0; X_t) = V_t^1(1, 0; X_t | w_{11,t}^*) = V_t^2(1, 0; X_t | w_{12,t}^*). \quad (15)$$

For this stayer, if $w_{12,t}^* < w_{11,t}^*$, the reservation wage $w_{13,t}^*$ that equates the value function of participation without interfirm mobility in period t with that of participation with interfirm mobility in period t is defined by

$$V_t^1(1, 0; X_t | w_{13,t}^*) = V_t^2(1, 0; X_t | w_{13,t}^*). \quad (16)$$

Then, if $w_{12,t}^* < w_{11,t}^*$, the decision rule for a stayer is to accept any wage offer greater than $w_{12,t}^*$, and either to move to another firm in the next period, if the current wage is between $w_{12,t}^*$ and $w_{13,t}^*$, or to stay in the firm for at least one more period, if the current wage is greater than $w_{13,t}^*$. If $w_{11,t}^* < w_{12,t}^*$, then the optimal strategy for a stayer is to accept any wage offer greater than $w_{11,t}^*$, and to stay in the firm for at least one more period.

For a mover in period $t - 1$, these reservation wages in period t , denoted $w_{21,t}^*$, $w_{22,t}^*$ and $w_{23,t}^*$, are defined similarly by

$$V_t^0(1, 1; X_t) = V_t^1(1, 1; X_t | w_{21,t}^*) = V_t^2(1, 1; X_t | w_{22,t}^*) \quad (17)$$

and

$$V_t^1(1, 1; X_t | w_{23,t}^*) = V_t^2(1, 1; X_t | w_{23,t}^*), \quad (18)$$

if $w_{22,t}^* < w_{21,t}^*$. Then, if $w_{22,t}^* < w_{21,t}^*$, the decision rule for a mover is to accept any wage offer greater than $w_{22,t}^*$, and either to move to another firm in the next period, if the current wage is between $w_{22,t}^*$ and $w_{23,t}^*$, or to stay in the firm for at least one more period, if the current wage is greater than $w_{23,t}^*$. If $w_{21,t}^* < w_{22,t}^*$, then the optimal strategy for a mover is to accept any wage offer greater than $w_{21,t}^*$, and to stay in the firm for at least one more period.

¹⁷For more details see Burdett (1978, Propositions 1 and 2, p. 215).

Summary of events and decisions:

Inter-firm mobility *may* occur when $w_{02,t}^* < w_{01,t}^*$, which, empirically, is the most relevant case for our econometric analysis. In this case, the events unfold as follows:

- A worker who is not employed in period $t - 1$ has to pay the search cost γ_1 at the beginning of period t for receiving a wage offer w_t^* at time t (from the wage offer function defined in (3)–(4)). He/she will accept the wage offer at t if it is greater or equal to a given threshold value $w_{02,t}^*$ that is defined in (12), or otherwise he/she will decline the offer and continue to search.
- An employed worker who moved to another firm at the end of period $t - 1$ pays the mobility cost c_M at the beginning of period t . In period t he/she receives a wage offer from his/her new employer. If that wage offer is lower than the threshold value $w_{22,t}^*$, the worker will become a non-participant at the end of period t . If the wage offer is higher than the threshold value $w_{23,t}^*$, then the optimal strategy for this worker is to accept this (inside) offer and stay in the firm for at least one more period (i.e., he/she becomes a stayer in period t). If the wage offer falls between $w_{22,t}^*$ and $w_{23,t}^*$, then the optimal strategy is to move to another firm at the end of period t .
- An employed worker who decided to stay in the same firm at the end of period $t - 1$ also receives a wage offer from his/her current employer in period t . If that offer is lower than the threshold value $w_{12,t}^*$, the worker will become a non-participant at the end of period t . If the wage offer is higher than the reservation wage $w_{13,t}^*$, then the optimal strategy for this worker is to accept this inside offer and to stay in the firm for at least one more period. If this offer falls between $w_{12,t}^*$ and $w_{13,t}^*$, then the optimal strategy is to move to another firm at the end of period t (i.e., the worker becomes a mover in period t).

In Appendix A.2, we show that the reservation wages of a stayer in period t may be deduced from the reservation wages of a non-participant in the same period through the following relationships:

$$\begin{aligned} w_{11,t}^* &\approx w_{01,t}^* - \gamma_{11}, \\ w_{12,t}^* &\approx w_{02,t}^* - \gamma_{12}, \quad \text{and} \\ w_{13,t}^* &\approx w_{03,t}^* + \gamma_{13}, \end{aligned}$$

where

$$\begin{aligned} \gamma_{1j} &= \gamma_1 \left[1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{0j,t}^*, 1; X_t)} \right], \quad \text{for } j = 1, 2, \\ \gamma_{13} &= \frac{u(z_t + w_{13,t}^*, 1; X_t) - u(z_t + w_{03,t}^*, 1; X_t)}{u'(z_t + w_{03,t}^*, 1; X_t)}, \end{aligned}$$

and $u'(\cdot)$ denotes the per-period marginal utility of consumption. The reservation wages of a mover in period t may be deduced from the three other pairs of reservation wages through the relationships:

$$\begin{aligned} w_{21,t}^* &\approx w_{01,t}^* - \gamma_{11} + \gamma_{21} \approx w_{11,t}^* + \gamma_{21}, \\ w_{22,t}^* &\approx w_{02,t}^* - \gamma_{12} + \gamma_{22} \approx w_{12,t}^* + \gamma_{22}, \quad \text{and} \\ w_{23,t}^* &\approx w_{03,t}^* + \gamma_{13} + \gamma_{23} \approx w_{13,t}^* + \gamma_{23}, \end{aligned}$$

where

$$\begin{aligned} \gamma_{2j} &= c_M \left[\frac{u'(z_t + w_{2j,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{0j,t}^*, 1; X_t)} \right], \quad \text{for } j = 1, 2, \quad \text{and} \\ \gamma_{23} &= \frac{u(z_t + w_{23,t}^*, 1; X_t) - u(z_t + w_{13,t}^*, 1; X_t)}{u'(z_t + w_{13,t}^*, 1; X_t)}. \end{aligned}$$

Note that in principle all the γ coefficients defined above can vary across individuals. Since it is not possible to estimate these coefficients separately from their population means along with all the other parameters of the model, we only estimate the average of these coefficients. This is why it is crucial for us to control for individual correlated random effects, because essentially they capture the deviation of the individual-specific γ 's from the average in the population.

These relationships are then used to derive the two sufficient conditions below. These conditions are crucial for establishing the results of our theoretical model, since they characterize the situations in which: (1) interfirm mobility may occur; and (2) the participation and mobility processes exhibit first-order dependence. These two conditions rely on the existence of threshold values associated with the individual decision to become a non-participant, a mover, and a stayer. The proofs of these two conditions are given in Appendix A.3.

Condition 1 *When $w_{02,t}^* < w_{01,t}^*$ the participation and mobility equations exhibit first-order state dependence: (a) if, at any level of consumption, the marginal utility of consumption is either higher or lower when working; (b) if condition (13) holds; and (c) if:*

$$c_M > \gamma_1 \left[\frac{u'(z_t + w_{02,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{22,t}^*, 1; X_t) - u'(z_t, 0; X_t)} \right]. \quad (19)$$

Condition 2 *When $w_{01,t}^* < w_{02,t}^*$ the participation equation exhibits first-order state dependence: (a) if, at any level of consumption, the marginal utility of consumption is either higher or lower when working; and (b) if:*

$$c_M > \gamma_1 \left[\frac{u'(z_t + w_{01,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{21,t}^*, 1; X_t) - u'(z_t, 0; X_t)} \right]. \quad (20)$$

In Appendix A.3. we show that the terms in square brackets in (19) and (20) are larger than 1. That is, Conditions 1 and 2 are stronger than the assumption we previously made that $c_M > \gamma_1$. Therefore, the conditions stated above require that the mobility cost c_M , which is incurred in the period after the move occurred, be relatively high for the participation and mobility processes to demonstrate first-order state dependence (see Hardman and Ioannides (2004)). These sufficient conditions are more likely to be valid if, alternatively, the search cost γ_1 is relatively low, or if the utility function u is weakly concave. Nevertheless, it is worthwhile emphasizing that these conditions are only *sufficient*, and not *necessary*, in order to have $w_{22,t}^* > w_{02,t}^*$.

When Condition 2 holds there is no inter-firm mobility. A non-participant becomes employed (respectively, remains non-participant) at the end of period t if he/she is offered a wage greater (respectively, lower) than $w_{01,t}^*$. A participant becomes a non-participant (respectively, remains employed) if he/she is offered a wage less (respectively, higher) than $w_{11,t}^*$. When the marginal utility of consumption is the same in the employment and non-employment states, the solution is similar to that in which the marginal utility of consumption is lower when working. (For more details see points B and D in Appendix A.3). Thus, Conditions 1 and 2 suffice for this situation as well.

When interfirm mobility may occur, i.e., when $w_{02,t}^* < w_{01,t}^*$, we show in Appendix A.3 that a mover at the end of period $t - 1$ becomes a non-participant at the end of period t if he/she is offered a wage less than $w_{22,t}^*$. A stayer becomes a non-participant if he/she is offered a wage less than $w_{12,t}^*$. Thus, the participation decision at period t can be characterized by

$$\begin{aligned}
y_t &= \mathbf{1} [w_t^* > w_{02,t}^* - \gamma_{12} y_{t-1} + \gamma_{22} y_{t-1} m_{t-1}] \\
&= \mathbf{1} [w_t^* - w_{02,t}^* + \gamma_{12} y_{t-1} - \gamma_{22} y_{t-1} m_{t-1} > 0], \quad \text{with} \tag{21} \\
\gamma_{12} &= \gamma_1 \left[1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)} \right], \quad \text{and} \\
\gamma_{22} &= c_M \left[\frac{u'(z_t + w_{22,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)} \right].
\end{aligned}$$

A stayer at $t - 1$ may decide to move to another firm at the end of period t if he/she is offered a wage w_t^* such that $w_{12,t}^* < w_t^* < w_{13,t}^*$, but he/she will stay in his/her current job for at least one more period if $w_t^* > w_{13,t}^*$. On the other hand, a mover at $t - 1$ will decide to move again at t only if the wage offer is such that $w_{22,t}^* < w_t^* < w_{23,t}^*$ and he/she will stay in his/her current job for at least one more period if $w_t^* > w_{23,t}^*$. Consequently, the decision to stay in the current firm at the end of period t can be characterized by the indicator variable s_t defined as:

$$\begin{aligned}
s_t &= 1 - m_t = \mathbf{1} [w_t > w_{13,t}^* + \gamma_{23} y_{t-1} m_{t-1}] \\
&= \mathbf{1} [w_t - w_{13,t}^* - \gamma_{23} y_{t-1} m_{t-1} > 0], \tag{22}
\end{aligned}$$

where m_t is the indicator variable indicating whether the worker i decides to move, i.e., $m_t = 1$, or

not to move, i.e., $m_t = 0$, at the end of period t .

Substitution of the wage w_t (omitting the i subscript) from (3) into (21) and (22) gives:

$$y_t = \mathbf{1} [J_{jt}^W + x'_{wt}\delta_0 - w_{02,t}^* + \gamma_{12} y_{t-1} - \gamma_{22} y_{t-1} m_{t-1} + \alpha_w + \xi_t > 0], \quad \text{and} \quad (23)$$

$$m_t = \mathbf{1} [w_{13,t}^* + \gamma_{23} y_{t-1} m_{t-1} - J_{jt}^W - x'_{wt}\delta_0 - \alpha_w - \xi_t > 0] \cdot \mathbf{1}(y_{i,t-1} = 1, y_{it} = 1), \quad (24)$$

where the reservation wages $w_{02,t}^*$ and $w_{13,t}^*$ are determined implicitly by (12) and (16), respectively.

Generally, the analytical solutions for $w_{02,t}^*$ and $w_{13,t}^*$ are intractable. To circumvent this difficulty, we assume that they can be approximated arbitrarily closely by linear functions of the entire set of exogenous and predetermined covariates, namely

$$w_{02,t}^* = a_{02}J_{jt}^W + x'_{yt}b_{02} + c_{02} y_{t-1} + d_{02} y_{t-1} m_{t-1} + \alpha_{02} + \xi_{02,t}, \quad \text{and} \quad (25)$$

$$w_{13,t}^* = a_{13}J_{jt}^W + x'_{mt}b_{13} + d_{13} y_{t-1} m_{t-1} + \alpha_{13} + \xi_{13,t}, \quad (26)$$

where x_{yt} and x_{mt} are sets of observable covariates. For identification purposes some of the covariates in x_{yt} and x_{mt} are excluded from x_{wt} , the set of covariates in the wage function defined in (3).¹⁸ The terms α_{02} and α_{13} are person-specific correlated random effects, while $\xi_{02,t}$ and $\xi_{13,t}$ are idiosyncratic contemporaneous errors.

Substituting (25) and (26) into (23) and (24), respectively, yields the following participation and mobility equations, respectively, at period t :

$$y_t = \mathbf{1} [a_0J_{jt}^W + x'_{yt}\beta_0 + \beta_y y_{t-1} + \beta_m y_{t-1} m_{t-1} + \alpha_y + u_t > 0] \quad \text{and} \quad (27)$$

$$m_t = \mathbf{1} [a_1J_{jt}^W + x'_{mt}\lambda_0 + \lambda_m y_{t-1} m_{t-1} + \alpha_m + v_t > 0] \cdot \mathbf{1}(y_{i,t-1} = 1, y_{it} = 1), \quad (28)$$

where $a_0 \equiv 1 - a_{02}$, $a_1 \equiv a_{13} - 1$, $\beta_y \equiv \gamma_{12} - c_{02}$, $\beta_m \equiv -\gamma_{22} - d_{02}$, $\lambda_m \equiv \gamma_{23} + d_{13}$. For β_0 we have $\beta_0 \equiv \delta_0 - b_{02}$ for all variables both in x_{wt} and in x_{yt} , and $\beta_0 \equiv -b_{02}$ for the covariates in x_{yt} but excluded from x_{wt} . Similarly, for λ_0 we have $\lambda_0 \equiv b_{13} - \delta_0$ for all variables both in x_{wt} and in x_{mt} , and $\lambda_0 \equiv b_{13}$ for the covariates in x_{mt} but excluded from x_{wt} . Note that the stochastic terms are given by

$$\alpha_y \equiv \alpha_w - \alpha_{02}, \quad \alpha_m \equiv \alpha_{13} - \alpha_w, \quad u_t \equiv \xi_t - \xi_{02,t}, \quad \text{and} \quad v_t \equiv \xi_{13,t} - \xi_t.$$

Note also that, by construction, the three person-specific effects, i.e., α_y , α_m , and α_w , are correlated, as is the case for the three contemporaneous white noises, i.e., u_t , v_t , and ξ_t . In addition, the (individual) random terms α_y , α_m , and α_w affect the wage in (3) in two ways. First, the wage is affected directly (through α_w). Second, it is affected indirectly (through α_m and α_y). Hence, all these random terms also affect the reservation wages $w_{02,t}^*$ and $w_{13,t}^*$. Essentially, these random terms account for an individual's unobserved heterogeneity. This specification would be sufficient to

¹⁸The excluded variables in this study are similar to those previously used in the literature, and they include the individual's marital status, the number of children in the family and non-labor income. We return to this point later in the text.

estimate the model through classical inference methods, such as (simulated) maximum likelihood procedures. But, in a broader sense, our Bayesian econometric approach, which will be presented in the next section, enlarges the scope of individual heterogeneity by assuming that the slope coefficients (including γ_{12} , γ_{22} and γ_{23}) are randomly drawn (from normal distributions). However, these parameters are not separately identified from α_y , α_m , and α_w . So we can think of γ_{12} , γ_{22} and γ_{23} as population means.

4 Econometric Specification and Estimation

4.1 Econometric Specification

Before proceeding with the econometric specification, it is convenient to rewrite the model as follows:

The *participation (employment)* equation, at any date $t > 1$, is given by¹⁹

$$\begin{aligned} y_{it} &= \mathbf{1}(y_{it}^* \geq 0), \\ y_{it}^* &= a_0 J_{jt}^W + x'_{y_{it}} \beta_0 + \beta_y y_{i,t-1} + \beta_m m_{i,t-1} + \alpha_{yi} + u_{it}. \end{aligned} \quad (29)$$

The *interfirm mobility* equation, at any date $t > 1$, is given by²⁰

$$\begin{aligned} m_{it} &= \mathbf{1}(m_{it}^* \geq 0) \times \mathbf{1}(y_{i,t-1} = 1, y_{it} = 1), \\ m_{it}^* &= a_1 J_{jt}^W + x'_{m_{it}} \lambda_0 + \lambda_m m_{i,t-1} + \alpha_{mi} + v_{it}. \end{aligned} \quad (30)$$

The *wage* equation, for any date t , is given by:

$$\begin{aligned} w_{ijt} &= w_{ijt}^* \times \mathbf{1}(y_{it} = 1), \quad \text{where} \\ w_{ijt}^* &= x'_{w_{ijt}} \delta_0 + J_{ijt}^W + \alpha_{wi} + \xi_{it}. \end{aligned} \quad (31)$$

The terms y_{it}^* and m_{it}^* denote the two latent variables affecting the employment and mobility decisions, respectively. Note that an obvious implication of the definition of a move is that a worker cannot move at date t unless he/she participated at both $t - 1$ and t .

Initial Conditions:

The likelihood function for the i th individual, conditional on the individual's specific effects

¹⁹Note that the labor market experience is simply the sum of the individual sequence of y_{it} . As is common in the literature, we make no distinction in this specification between unemployment and non-participation in the labor force. In addition, as mentioned earlier, we equate participation and being employed.

²⁰Mobility takes place at the beginning of period t . And seniority is the sum of the individual sequence of m_{it} with his/her current employer.

$\alpha_i = (\alpha_{yi}, \alpha_{mi}, \alpha_{wi})'$ and the exogenous variables $x_{it} = (x'_{yit}, x'_{mit}, x'_{wit})'$, is given by

$$l \left\{ (y_{it}, m_{it}, w_{it})_{t=1, \dots, T} \mid \alpha_i, x_{it} \right\} = \prod_{t=2}^T l \left\{ (y_{it}, m_{it}, w_{it}) \mid \alpha_i, x_{it}, y_{it-1}, m_{it-1}, J_{it}^W \right\} \quad (32)$$

$$\times l \{ w_{i1} \mid \alpha_i, x_{i1}, y_{i1}, m_{i1} \} l \{ y_{i1}, m_{i1} \},$$

where the last term of the right hand side of (32) is the likelihood of (y_{i1}, m_{i1}) for the initial state at time $t = 1$. Following Heckman (1981), we approximate this part of the likelihood by a probit specification given by

$$y_{i1} = 1(y_{i1}^* \geq 0), \quad \text{where } y_{i1}^* = ax_{yi1} + \delta_y \alpha_{yi} + u_{i1}, \quad \text{and} \quad (33)$$

$$m_{i1} = 1(m_{i1}^* \geq 0) \times 1(y_{i1} = 1), \quad \text{where } m_{i1}^* = bx_{mi1} + \delta_m \alpha_{mi} + v_{i1},$$

and α_{yi} and α_{mi} are the individual specific effects in the participation and mobility equations defined in (29) and (30), respectively. That is, δ_y and δ_m are allowed to differ from one.

Note that in order to obtain estimates for the model's parameters one needs also to integrate the likelihood function in (32) over the distribution of the individual specific effects, which makes the estimation for some of the groups infeasible. Hence, in the analysis reported below, we adopt a Bayesian approach whereby we obtain the conditional posterior distribution of the parameters, conditional on the data, using Markov Chain Monte Carlo (MCMC) methods as explained below.²¹

Stochastic Assumptions:

We assume that the individual specific effects are stochastically independent of the idiosyncratic shocks, that is $\alpha_i \perp (u_{it}, v_{it}, \xi_{it})$. Furthermore, α_i follows the distribution given by

$$\alpha_i \sim N(f(x_{i1}, \dots, x_{iT}), \Gamma_i), \quad \text{where} \quad (34)$$

$$\Gamma_i = D_i \Delta_\rho D_i,$$

$$D_i = \text{diag}(\sigma_{yi}, \sigma_{mi}, \sigma_{wi}), \quad \text{and}$$

$$\{\Delta_\rho\}_{j,l} = \rho_{\alpha_{ji}\alpha_{li}}, \quad \text{for } j, l = y, m, w.$$

The function $f(x_{i1}, \dots, x_{iT})$ depends, in principle, on all exogenous variables in all periods. This can be seen from the expression of its statistical counterpart, that is the mean of the posterior distribution of the individual specific effects α_i (see Appendix B). The matrix Γ_i is indexed by i , since we also allow for σ_{ji} to be heteroskedastic, i.e., to depend on x_{yit} , x_{mit} , and x_{wit} , respectively.

²¹One can also use an alternative ("frequentist") approach such as Simulated Maximum Likelihood (SML) (e.g. Gouriéroux and Monfort (1996), McFadden (1989), and Pakes and Pollard (1989)). However, the maximization is rather complicated and highly time consuming. For comparison we estimated the model using the SML method only for one group (the smallest one) of college graduates. The estimation took several months. Nevertheless, the point estimates for all the model's parameters obtained by the SML are virtually the same as the mean of the joint posterior distribution of the model's parameter, inducing identical estimates for the main parameters of interest, namely the returns to seniority and experience. The MCMC method can be viewed in this sense as a mechanical device for obtaining SML estimates.

That is,

$$\sigma_{ji}^2 = \exp(h_j(x_{i1}, \dots, x_{iT})), \quad \text{for } j = y, m, w, \quad (35)$$

where each $h_j(\cdot)$ is some real valued function. The ultimate goal in doing so is to control for the possible existence of heterogeneity in a parsimonious way. Hence, we base our estimation only on the first three principle components of the sample averages of the regressors' vector, i.e., $\bar{x}_{ji} = (\sum_{t=1}^T x_{jit})/T$ (as well as a constant term) for each individual. That is, we approximate $h_j(x_{i1}, \dots, x_{iT})$ by²²

$$\hat{h}_j(x_{ji1}, \dots, x_{jiT}) = pc_i' \hat{\gamma}_j, \quad (36)$$

where pc_i is the vector containing the first three principal components. This significantly reduces the computational burden because the posterior distribution for $\gamma = (\gamma'_y, \gamma'_m, \gamma'_w)'$ is difficult to obtain, and it requires a complicated Metropolis-Hastings step.²³

Finally, the idiosyncratic error components from (29), (30), and (31), i.e., $\tau_{it} = (u_{it}, v_{it}, \xi_{it})'$, are assumed to be contemporaneously correlated white noises, with

$$\tau_{it} \sim N(0, \Sigma), \quad \text{where} \quad (37)$$

$$\Sigma = \begin{pmatrix} 1 & \rho_{uw} & \rho_{u\xi}\sigma_\xi \\ \rho_{uw} & 1 & \rho_{v\xi}\sigma_\xi \\ \rho_{u\xi}\sigma_\xi & \rho_{v\xi}\sigma_\xi & \sigma_\xi^2 \end{pmatrix}, \quad (38)$$

and for identification purposes, we have normalized the variance of u_{it} and v_{it} to be $\sigma_u^2 = \sigma_v^2 = 1$.

4.2 Estimation—Computing The Posterior Distribution

Since it is analytically intractable to compute the exact posterior distribution of the model's parameters, conditional on the observed data, our goal here is to summarize the joint posterior distribution of the parameters of the model using a Markov Chain Monte Carlo (MCMC) algorithm. Appendix B provides a detailed description of the implementation of the MCMC simulation. Here we only touch upon the key issues.

Let the prior density of the model's parameters be denoted by $\pi(\theta)$, where θ contains all the parameters of the model as defined in detail below. The posterior distribution of the parameters would then be:

$$\pi(\theta | z) \propto \Pr(z | \theta)\pi(\theta),$$

where z denotes the observed data. This posterior density cannot be easily simulated due to the intractability of $\Pr(z | \theta)$. Hence, we follow Chib and Greenberg (1998) and augment the parameter space to include the vector of latent variables, $z_{it}^* = (y_{it}^*, m_{it}^*, w_{it}^*)$, where y_{it}^* , m_{it}^* , and w_{it}^* are defined in (29), (30), and (31), respectively.

²²The first three principle components account for over 98% of the total variance of \bar{x}_{ji} , so that there is almost no loss of information by doing so.

²³We do not attempt to attribute any causal interpretation to $\hat{\gamma}_j$, $j = y, m, w$, but rather to control for possible dependence of the covariance matrix on observed characteristics of the individual.

With this addition it is easier to implement the Gibbs sampler, which iterates through the set of the conditional distributions of z^* (conditional on θ) and θ (conditional on z^*).²⁴ In all the estimations reported below we employed 50,000 MCMC repetitions after the first 5,000 were discarded. As recommended by Gelfand and Smith (1990), the performance of the algorithm and its approach to the stationary distribution are assessed by monitoring the evolution of the quantiles as the sampling proceeds.

A key element for computing the posterior distribution of the parameters is the choice of the prior distributions for the various elements of the parameter space. We adopt here conjugate, but very diffuse, priors on all the parameters of the model, reflecting our lack of knowledge about the possible values of the parameters. In all cases we use proper priors (although very dispersed) to ensure that the posterior distribution is a proper distribution.

A comprehensive sensitivity analysis that we carried out shows that the choice of the particular prior distribution hardly affects the posterior distribution. In particular, we estimated the model centering the key parameters around the results obtained by Topel (1991) and AS. The resulting posterior distributions obtained under both scenarios were virtually identical.

5 The Data

We follow Topel (1991) and Altonji and Williams (2005) and use the Panel Study of Income Dynamics (PSID). The PSID is a longitudinal study of a representative sample of individuals in the U.S. and the family units in which they reside. The survey, begun in 1968, emphasizes the dynamic aspects of economic and demographic behavior, but its content is broad, including sociological and psychological measures.

Two key features give the PSID its unique analytic power: (i) individuals are followed over very long time periods in the context of their family setting; and (ii) families are tracked across generations, where interviews of multiple generations within the same families are often conducted simultaneously. Starting with a national sample of 5,000 U.S. households in 1968, the PSID has re-interviewed individuals from those households every year, whether or not they are living in the same dwelling or with the same people. While there is some attrition rate, the PSID has had significant success in its recontact efforts.²⁵

The data extract used in this study comes from 18 waves of the PSID from 1975 to 1992. The sample is restricted to all heads of households who were interviewed for at least three years during the period from 1975 to 1992 and who were between the ages of 18 and 65 in these survey dates. We include in the analysis all the individuals, even if they reported themselves as self-employed. We also carried out some sensitivity analysis, excluding the self-employed from our sample, but the

²⁴A presentation of the theory and practice of Gibbs sampling and Markov Chain Monte Carlo methods may be found in the book written by Robert and Casella (1999), and in the survey by Chib (2001). In econometrics, recent applications to panel data include the papers by Geweke and Keane (2000), Chib and Hamilton (2002) and Fougère and Kamionka (2003).

²⁵There is a large number of studies that use this survey for many different research questions. For a more detailed description of the PSID, see Hill (1992).

results remained virtually the same. Furthermore, most of the heads of household in the sample are men. We therefore conducted an additional analysis dropping all women household heads from the estimation. Again, no visible differences have been detected in the results obtained. We excluded from the extract all the observations which came from the poverty sub-sample of the PSID.

There are some difficulties that need to be carefully addressed with some of the key variables of our analysis and especially with the tenure variable. As noted by Topel (1991), tenure on a job is often recorded in wide intervals, and a large number of observations are lost because tenure is missing. There are also a large number of inconsistencies in the data inducing tremendous spurious year-to-year variance in reported tenure on a given job. For example, between two years of a single job tenure sometimes falls (or rises) by much more than one year. There are many years with missing tenure followed by years in which a respondent reports more than 20 years of seniority.

Since the errors can basically determine the outcome of the analysis, we reconstructed the tenure and experience variables using the exact procedure suggested by Topel (1991). Specifically, for jobs that begin in the panel, tenure is started at zero and is incremented by one for each additional year in which the person works for the same employer. For jobs that started before the first year a person was in the sample, a different procedure was followed. The starting tenure was inferred according to the longest sequence of consistent observations. If there was no such sequence then we started from the maximum tenure on the job, provided that the maximum was less than the age of the person minus his/her education minus 6. If this was not the case then we started from the second largest value of recorded tenure. Once the starting point was determined, tenure was incremented by one for each additional year with the same employer. The initial experience was computed according to similar principles. Once the starting point was computed, experience was incremented by one for each year in which the person worked. Using this procedure we managed to reduce the number of inconsistencies to a minimum.²⁶

Summary statistics of some of the key variables in the extract used here are reported in Table 1. By the nature of the PSID data collection strategy, the average age of the sample individuals does not increase much over time. The average education rises by more than half a year between 1975 and 1992, because the new individuals entering the sample have, on average, higher education than those in the older cohorts. Also, experience and seniority tend to increase over the sample from 20.8 and 5.1 years in 1975 to 24.2 and 7.3 in 1992, respectively.

AW provide summary statistics for the samples used in the various studies (see Table 1 of AW). The experience variable (line 2) used here seems to be very much in line with that reported in other studies. However, there is a substantial difference between the tenure variable used here and that used by Topel and especially AW. The main reason for this discrepancy is the fact that we restrict the sample to individuals for whom we can obtain consistent data for three consecutive years. Individuals with lower seniority are also those that tend to leave the PSID sample more frequently. Nevertheless, we provide below strong evidence that the differences in the results are

²⁶We also took a few other additional cautionary measures. For example, we checked that: (i) the reported unemployment matches the change in the seniority level; and (ii) there are no peculiar changes in the reported state of residence and region of residence, etc. The programs are available from the corresponding author upon request.

not due to selection of inherently different data extracts, but rather because of the approach.

The mobility variable (line 5) indicates that, in each of the sample years, between 6.5% and 14.1% of the individuals changed jobs. The mobility is very large in the first year of the sample, probably due to measurement error. Since it is difficult to identify whether or not a person actually moved in those early years, it is necessary to control for initial conditions as was explained above.

Consistent with other data sources, the average real wage increased slightly over the sample years. This is mainly because the individuals entering the sample have wages that are increasing over time in real terms, while those who leave the sample have wages that decrease somewhat over the sample years. More importantly, note that the wage dispersion increases across years. Also consistent with other data sources, the participation rate of individuals in the sample (almost exclusively men) decreases steadily over the sample years from about 92% in 1975 to 86% in 1992. The PSID oversamples non-whites. However, since the results change very little when we use a representative sample, we keep all the individuals satisfying the conditions specified above.

Approximately 20% of the sample have children who are two years old and below, although this fraction slightly decreases over the sample years, as does the fraction of the sample that has children between the ages of 3 and 5. The total number of children also declines somewhat over the sample period. This is consistent with the general finding in the literature about the decline in the size of the typical American family.

Note that there are substantial changes in the distribution of cohorts over the sample period. The fraction of people in the youngest cohort increases steadily, in particular between 1988 and 1990, as does the number of observations, and the fraction of Hispanics.²⁷ In contrast, the fraction of people in the oldest cohorts decreases over the sample years.

6 The Results

The estimation is carried out for three separate education groups. The first group includes all individuals with less than 12 years of education, referred to as *high school (HS) dropouts*. The second group consists of those who are high school graduates, who may have acquired some college education, and/or who may have earned a degree higher than high school diploma, but have not completed a four-year college. We refer to this group as *high school graduates*. The third group, the *college graduates*, consists of those that have at least 16 years of education and who have earned a college degree. Below we present the results from the simultaneous estimation of participation-employment, mobility, and wage equations, together with the initial conditions, for each group. For brevity we do not report the estimates for the initial conditions' equations.

Specification:

The participation-employment equation includes the following variables: a constant, education, quartic in lagged labor market experience, a set of three regional dummy variables, a dummy vari-

²⁷Those in charge of the PSID made a special effort to collect information for those who left the sample in the previous years. The changes in the age and race structure are due to strong geographic mobility of these young workers.

able for residence in an SMSA, family unearned income, dummy variables for being an African American and for being Hispanic, county of residence unemployment rate, a set of variables providing information about the children in the family, a dummy variable for being married, a set of four dummy variables for the cohort of birth, and a full set of year indicators. In addition we also include the J^W function, as explained above.²⁸

The mobility equation includes all the variables that are included in the participation equation. In addition it includes a quartic in lagged seniority on the current job and a set of nine industry indicator variables.

The dependent variable in the (log) wage equation is the log of the deflated annual wage. Instead of the quartic in lagged experience and lagged seniority, we include in the wage equation a quartic in experience and a quartic in seniority on the current job. Several variables which appear in the participation and mobility equations are excluded from the wage equation. These are the family unearned income, the number of children in the family, the number of children aged between 1 and 2, the number of children aged between 3 and 5, and a marital status dummy variable that takes the value 1 when the individual is married and zero otherwise. The excluded variables from the wage equation listed above are assumed to be exogenous with respect to the model. These exclusion restrictions provide an additional source for identifying the wage regression parameters, in addition to the identification that comes from the non-linearity of the model.²⁹

It is worthwhile noting that the identification of the various parameters of the model comes from several sources of variation in the data. First, given that experience and seniority have been explicitly modeled, the cross-sectional dimension makes it possible for one to estimate all the parameters in the three equations if there are no individual specific effects. Nevertheless, the time dimension of our data extract also helps in identifying the key parameters, namely the returns to experience, and, even more importantly, the returns to seniority. That is, there are a considerable number of individuals who changed jobs at some point during the sample period and did so at different points of their life-cycle. The time dimension in the data set also allows us to identify the individual specific effects specified in Section 3. This is why we restrict our data extract to include only individuals who have at least three consecutive years in which they are observed during the sample period. Since in our model experience and seniority are fully endogenized, and since we also explicitly control for the individuals' initial conditions, one need not impose further restrictions on the data extract (e.g. to include only exogenously displaced workers) as is done in Dustmann and Meghir (2005) or Topel (1991).

The main results are provided in Tables 2 through 8. In Tables 2 and 3, we briefly describe the estimates for the participation and mobility equations, respectively.³⁰ We then concentrate our

²⁸In Appendix C we present descriptive statistics from the raw data that justify the introduction of the J^W function, as well as our preferred specification.

²⁹We follow the large literature on labor supply in doing so. For extensive discussion of these exclusion restrictions, see Mroz (1987).

³⁰It turns out that the marginal distributions of all the parameters of the model, and all the marginal and cumulative returns to experience and seniority, are very close to those of normal random variables. Therefore, it is sufficient to report, as we do, the mean and the standard deviation for all the relevant posterior distributions. For that very reason we also do not provide any figures for the posterior marginal distributions. In addition, for brevity we do not

discussion on the returns to human capital, namely education, experience, and, most importantly for this study, the returns to seniority. In Table 4 we provide the results for the wage equation from which the various marginal and cumulative returns can be computed. These are provided in Tables 5 and 6 for experience and seniority, respectively. Table 7 presents a comparison of our results with those of AW. Finally, Table 8 reports estimates of some of the key parameters associated with the stochastic terms of the model.

Results we obtain are very different, especially for the returns to seniority, from those previously presented in the literature. A natural question to ask then is whether or not these apparent differences stem simply from a different data extract. To examine this question we re-estimated the models of Topel (1991) and AW using our data extract. The results of this investigation are clear-cut: the differences between our results and theirs are not due to the use of different data extracts, or auxiliary decisions that have little to do with the methods employed. The differences stem from three elements: (a) the specific modelling strategy adopted here (namely the joint estimation of participation, mobility and wage outcomes); (b) the explicit control for individual-specific effects in the three equations; and (c) the introduction of the J^W function that captures past employment spells.

6.1 Participation-Employment and Mobility

As explained above, these two equations take into account both duration dependence and unobserved heterogeneity. In the participation equation we include lagged participation and lagged mobility, whereas the mobility equation only includes lagged mobility. In principle, the adopted specification requires the inclusion of the J^W function as defined in (5) in both the employment and mobility equations.

However, inclusion of the J^W function in these two equations does not change the results in any meaningful way. First, all the coefficients associated with the elements of the J^W function are statistically and economically insignificant.³¹ Second, the inclusion of the J^W function has no effect on any of the other model's parameters. For these reasons we simply exclude the J^W function from both participation and mobility equations.

Participation-employment:

The estimates are in line with those previously obtained in the literature and are consistent with the classic human capital theory. Education is an important factor in the participation-employment decision for the more educated and has no effect on high school dropouts. Similar qualitative results are also obtained for experience, which seems to be more important for the more educated individuals.

Lagged employment has a positive and highly significant effect for all three education groups

explicitly report the estimates of the coefficients on the year indicators (all tables). We also omit the results for the industry indicator variables in Table 3.

³¹This result implies that slope parameters a_1 and a_2 associated with the J^W function in the reduced-form equations (27) and (28) are not statistically different from zero. This is equivalent to accepting the null hypothesis that structural parameters a_{02} and a_{13} associated with J^W in equations (25) and (26) are both equal to 1.

in the participation-employment equation. Similar results are also obtained for lagged mobility, namely individuals who move from one job to another in any given period are more likely to be in employment in the following period. Moreover, as predicted by human capital theory, the latter effect is stronger for the least educated individuals.

The results for the family variables are, in general, consistent with economic theory (in particular marital status and the children variables). The coefficients for the race indicators are particularly striking. While African-Americans are less likely to participate at all education levels, the negative effect appears to be particularly large for the high school dropouts and high school graduates. On the other hand, Hispanics appear to have very similar behavior to the whites. Finally, younger cohorts are more likely to be employed than older ones.

Mobility:

For all education groups the level of experience appears to be largely irrelevant for the mobility decision. In contrast, the effect of seniority is very strong and negative. This result implies that more senior workers tend to move less because the returns to seniority, which represent returns to firm-specific human capital, are lost once a person leaves his/her job. Moreover, mobility in any given period considerably reduces the probability of a move in the subsequent period. This effect is particularly strong for the more educated individuals, potentially because the signalling effect of repeated frequent moves may be more severe for the more educated individuals.

Interestingly, among the family variables, family unearned income seems to be the only consistent factor affecting the likelihood of a move. Moreover, when evaluated at the average level of unearned income for the group, the marginal effect of unearned family income on the probability of a move is roughly the same for all groups. Finally, race does not seem to play a major role in mobility decisions (except for the very educated blacks).

6.2 The Returns to Education

In this subsection and those that follow we turn to the estimates from the wage equation. The (within-education groups) returns to education, reported on line (2) of Table 4, are broadly consistent with the human capital theory. Declining marginal returns to education lead to lower returns to education for the college graduates than for the high school graduates. However, the marginal returns are larger for the high school graduates than for the high school dropouts, 4.8% and 2.5%, respectively. These results appear to be consistent with recent findings in the literature (e.g. Card (2001)) about the within-education group returns to education.³² Nevertheless, the estimates obtained here are somewhat lower than those reported by Card (2001). It turns out that, once one controls for the selection effects, due to the participation and mobility decisions, the estimated returns to education are substantially reduced, potentially also explaining that the returns are larger for high school graduates than for high school dropouts and lower for college graduates than for the high school graduates. We return to this effect below when discussing the estimates associated

³²There is substantial variation within each group of education as we defined them. This variation allows one to separately identify the within-group returns to education for the three groups.

with the stochastic terms of the model.

6.3 The Returns to Experience

The results for the returns to experience, which are based on the estimates reported in Table 4, are provided in Table 5. The results in Panel A of Table 5 are based on the quartic model in both experience and seniority. In Panel B of Table 5 we provide an additional set of estimates based on the quadratic model in experience and seniority.

The results in Panel A of the table are close, yet generally larger, than those previously obtained in the literature. In particular, they are slightly larger than those obtained by Topel (1991), AS and AW. Topel's estimate, i.e., Topel (1991, Tables 1 and 2), of the cumulative returns to experience at 10 years of experience is .354. The estimates of AS range between .372 and .442, while those of AW range between .310 and .374. Our estimates are .362 for the high school dropouts, .402 for the high school graduates, and .661 for the college graduates. These results also indicate that there are substantial differences across the various educational groups at all levels of experience. Even only 5 years after entering the labor market, the cumulative returns for college graduates are almost twice as large as those for the other two groups.

Note also that the cumulative returns to experience are substantially lower for the quadratic models, as is indicated by the results reported in Panel B of Table 5. For example, at 10 years of experience the cumulative returns for the three educational groups are .246, .253, and .446, respectively, substantially lower than for the quartic model.

In Panels C and D of Table 5 we report the results for the quartic and quadratic model when the J^W function defined in (5) is omitted. Omitting the J^W function induces a huge reduction in the estimated cumulative returns to experience for all educational groups. Given that the estimates of the parameters associated with the J^W function (see rows 11–23 of Table 4) are jointly highly significant, this clearly highlights the importance of controlling for past job market history. Indeed, failing to account for past changes in employment status and job mobility induces a severe upward bias on the estimated returns to experience.

It is essential to see that, unlike in Topel (1991), AS, and AW, we allow for the possibility of non-zero correlations between the various individual-specific effects in the participation, mobility and wage equations. It turns out that failure to control for these non-zero correlations also induces significant biases on the estimated returns to experience. This becomes clear from the discussion below regarding the estimates of the parameters that correspond to stochastic elements of the model.

6.4 The Returns to Seniority

The returns to seniority reported in Panel A of Table 6 (for the quartic model) are higher than those previously reported in the literature, including those reported by Topel (1991). This is true for all educational groups at all levels of seniority. In fact, the support of the marginal posterior distributions for the returns to seniority is entirely in the positive segment of the real line for all

groups. The returns also increase dramatically with the level of seniority: At 2 years of seniority they are around 13%, rising to about 50% at 10 years of seniority. The results for the quadratic model, reported in Panel B of Table 6, are qualitatively similar to those obtained with the quartic model. One difference is that for the least educated group the cumulative returns tend to flatten at 10 years of seniority.

In Panels C and D of Table 6 we report the estimated returns to seniority for the two models when the J^W function is omitted from the wage equation. The resulting estimates, especially those reported in Panel C, are much lower than those reported in Panels A and B. This clearly highlights the need for one to control for past labor market history, as we do via the J^W function. Overall, the average cumulative returns over the three educational groups are, as one might expect, remarkably close to those obtained by Topel.

One key reason for the difference between our results and Topel's (1991) results is the treatment of experience. As mentioned above, Topel treats the experience level at new jobs as exogenous. There is strong evidence in the literature (e.g. Farber (1999)) indicating that this is not the case. The results presented above for the mobility equation indicate that experience plays a major role in the mobility decision and hence is highly correlated with the level of experience in a new job, for those who move. Indeed, when we treat experience as exogenous (and hence omit the participation equation from the estimation), the resulting returns to seniority are reduced. Nevertheless, they are still significantly higher than those obtained by Topel. If, in addition, we omitted the J^W component from the wage equation, the resulting estimates for the returns to seniority are reduced even further.

The cumulative returns to experience and seniority presented above for the various model specifications (i.e., quartic versus quadratic and with and without the J^W function) indicate that the sum of the returns from experience and seniority are quite similar across the various specifications (but larger than previously obtained in the literature). Indeed, since seniority and experience are positively correlated, and more so for the more educated individuals, larger returns to one component are therefore associated with lower returns to the other.³³

To summarize our results and to compare them with the literature, Table 7 presents results from AW (their Table 2) in its first panel. This panel provides the returns to seniority at 2, 5, 10, and 15 years of seniority using OLS and IV estimates based on AW's methodology using the *Topel replication sample* for the period 1968-1983. The second panel presents the results using *our sample* for the period 1975-1992. We present two sets of columns, the first one for high school dropouts and the second for college graduates. The first two rows of this panel provide the returns at 2, 5, 10, and 15 years of seniority using OLS and AW's methodology. The last four rows of the panel present the returns (already reported in Table 6) from our analyses. The results are extremely clear. The OLS returns to seniority in the two samples are very close (slightly lower for college graduates). A similar conclusion can be drawn from the IV estimates based on AW's methodology. Hence, the differences in results are due to the differences in methodology, not data.

³³The correlation coefficients between experience and seniority are: .22 for the high school dropouts, .36 for the high school graduates, and .46 for the college graduates.

Finally, the last four rows of the lower panel indicate that the different results obtained in our analyses are not consequences of using different samples. Quite the contrary, our sample yields very similar results when we use the same methodology as has been previously used in the literature. It clearly appears that the joint modelling of wages, employment, and mobility, as well as the introduction of the J^W function, strongly affects the estimated returns to seniority.

6.5 The J^W Function

As explained above, the J^W function is an individual-job specific function that parsimoniously summarizes changes in one's wage that correspond to a particular career path. The parameter estimates of the J^W function, provided in rows 11 through 23 of Table 4, clearly indicate the importance of controlling for the individual's career path. Even though some of the individual parameters are statistically insignificant, the null hypothesis that all coefficients are zeros is rejected at any reasonable significance level for all three educational groups.

Some important implications emerge from these results. First, the timing of a move in a worker's career matters. For all three groups the J^W function increases with seniority in the previous job. While at the time of a move a person loses the accumulated returns to seniority, he/she also receives the associated premium coming from the relevant component of the J^W function. Consider, for example, a high school dropout with 10 years of seniority and 10 year of experience moving to a new job. The results in Panel A of Table 6 indicate that he/she loses .507 in log-wage. But from Table 4 we see that he/she gains $.026 \times 10 = .26$ (line 15), for an overall substantial wage loss of about 25%. This loss captures the so-called *displaced worker effect* analyzed in the literature. The effect of unemployment spells during one's career is captured through the (actual) labor market experience.

Note that for high school dropouts the components of the J^W function associated with experience are not significantly different from zero (see lines 20 through 23 of Table 4). By contrast, a college-educated worker changing jobs after 10 years of seniority would lose .477 of accumulated returns to seniority (see Table 6, Panel A), but he/she also regains .517 from the J^W function (Table 4, line 16). A college-educated worker changing jobs after only two years in a job loses .136 log wage points (Table 6, Panel A) but gains through the J^W function .157 log points from the constant component (line 14) and additional .108 log points from the part proportional to seniority (line 17). That is, the effect of seniority at these low levels is economically large and positive. In contrast, there is a small negative effect associated with the experience component of the J^W function for the most highly educated workers. Finally, the high school graduates do not seem to lose overall from seniority for spells of at least 10 years in a firm, but the experience component in the J^W function seems to have a large negative effect on the wage changes at the time of a job change.

It is worth noting that the estimation of the J^W function is made possible only because we explicitly endogenize both experience and seniority. Moreover, the J^W function allows us to control for the indirect effects of experience and seniority that stem from previous jobs and introduce a

non-linear interaction between seniority and experience. In particular, it allows us to explicitly model the first few years on the labor market. The results indicate that this plays a crucial role in explaining wage growth (see also Topel and Ward (1992)).

Our model consists of a complex combination of factors affecting the estimated returns to experience and seniority, so it is hard to be able to assess the overall effect on the estimated returns. The results generally support the results and conjectures of Topel (1991). However, we also find that there is a clear need for controlling for the highly non-linear role of the individual's career path, as is demonstrated by the vastly different results obtained from the quartic models without the J^W function (see Panels A and C of Tables 5 and 6). Additional differences in the estimated returns are induced by the joint modeling of the three outcome equations as is discussed in the next subsection.

6.6 Estimates of the Stochastic Elements

In Table 8 we provide estimates of the key parameters associated with the stochastic elements of the model, that is the parameters of Σ in (38) and the correlation parameters of Δ_ρ in (34). Recall that we impose very few restrictions on either Σ or Δ_ρ . Hence, it is crucial to examine the estimated correlations of the random elements across the various equations. In general, whether or not the estimated correlations are different from zero determines whether it is necessary to control for the participation and mobility decisions when estimating the wage equation.

For the correlations between the various person-specific components, i.e., the elements of Δ_ρ , we see that most of them are highly significant and very large in magnitude. This is especially true, as one might expect, for the high school and college graduates. Since higher participation rates imply faster accumulation of labor market experience, the estimates of $\rho_{\alpha_y \alpha_w}$ (in row 6) imply that, all other things equal, high-wage workers (i.e., workers with a large person-specific component in the wage equation) tend to be more experienced workers. Hence, omission of the person-specific effect from the wage equation would induce an upward bias in the estimated returns to experience. However, because of the positive correlation between the person-specific effects in the participation and wage equations, failing to control for the participation decision would induce a downward bias on the (positive) returns to experience. The net effect of these two conflicting effects is *a priori* not clear. Overall, we find that these two effects almost cancel each other out, leading to estimated returns to experience which are only somewhat larger than those previously obtained in the literature (e.g. Topel (1991), AS, and AW).

The correlations between the participation-employment and the mobility components, i.e., $\rho_{\alpha_y \alpha_m}$ (in row 5), and between the mobility and the wage components, i.e., $\rho_{\alpha_m \alpha_w}$ (in row 7), are very large and negative for high school graduates and college graduates. That is, high-mobility workers also tend to be low-wage workers and low-participation workers. Moreover, high-wage workers have higher seniority than low-wage workers. Hence, omitting the person-specific effect from the wage equation, in absence of any control for mobility (and participation) decisions, would induce an upward bias in the estimated returns to seniority. Also, because of the negative correlation between

the two person-specific effects, failing to control for the mobility decision would induce a downward bias in the returns to seniority. Consequently, the net effect is not clear. Empirically we find that the overall effect of controlling for the mobility and participation decisions, and the inclusion of person-specific effects in the various equations, as well as the *presence* of the J^W function, is huge, leading to estimated returns to seniority which are larger than those estimated by Topel (1991).

Finally, examining the estimated Σ , we see that $\rho_{u\xi}$, i.e., the correlation between the idiosyncratic errors in the participation-employment and wage equations, is negative and highly significant for all education groups. By contrast, and mostly for the college-educated workers, positive shocks to mobility are associated with positive shocks to wages. Other correlations in Σ are not statistically significant and are generally small in magnitude.

6.7 Implications for Long-run Wage Growth

Because of the complexity of the various components at play in our model, it is somewhat difficult to assess the overall effects of various participation and mobility episodes on wage growth. To better address this issue, we follow the wage growth for two types of hypothetical workers with two distinct career paths. One group consists of those working through the entire sample period in one firm, while the other includes those working in one firm for the first four years, then accepting a new job in which the worker stays for the remainder of the sample period (with no intervening unemployment episode). Figures 1 and 2 depict the results for the high school graduate and college graduate groups, respectively.³⁴ Results for new entrants, i.e., those who in the first sample year have 5 years of experience and 2 years of seniority, for the two alternative career paths are given in Figures 1a, 1c, 2a and 2c, respectively. Results for workers that start (first sample year) with 15 years of experience and 6 years in the employing firm are presented similarly in Figures 1b, 1d, 2b, and 2d. In all figures we decompose the wage growth that stems from returns to human capital into the part that comes from rising experience and the part that comes from rising seniority. The terms associated with the J^W function are also included in the initial wage of the worker at the entry point to a new firm.

First, consistent with human capital theory, wage growth early in one's career is large, and more significantly so for the more educated individuals. This effect is reinforced when the components coming from the J^W function are incorporated. For example, wage growth over the 18-year period for new entrants with a high school degree (Figure 1a) is 0.74 log points, whereas for new entrants with a college degree it is 0.94 log points. This is in addition to the fact that the starting wage for college graduates is higher (Figure 2a).

Furthermore, for both education groups, wage growth that stems from increased seniority is larger than wage growth that stems from increased experience. For the college graduate group (workers with no job change) the increases due to seniority and experience are 0.59 and 0.35 log points, respectively, for the new entrants, and 0.50 and -0.08, for the experienced workers. The

³⁴Since the pattern of the results for the high school dropouts is very similar to that of the high school graduates, we omit it for brevity.

differences for the high school graduates (workers with no job change) are even larger. For the new entrants, the respective increases are 0.59 and 0.16 log points, while for the experienced workers, the respective increases are 0.48 and -0.04 log points. These results imply that the loss of firm-specific skills, as represented by seniority, leads to a substantial decrease in one's wage.

In general there are two effects that come into play when there is a job change. On one hand, there are direct losses due to lost returns to accumulated seniority. On the other hand, there are discrete jumps in the entry wage level at the new job, which are summarized in the J^W function. For high school graduates with little labor market experience, the loss of accumulated seniority is compensated almost entirely by a discrete increase in the base wage (Figure 1c). This is even more pronounced for new entrants with a college degree, for whom the loss of accumulated seniority is more than compensated by the change in the entry wage level (Figure 2c). Hence, movements early in a career appear to be favored by employers. In contrast, workers with substantial labor market experience (15 to 20 years) after a relatively long spell within a firm incur substantial wage loss when they change employers. For example, the implied loss for high school graduates with 9 years of seniority due to loss of accumulated seniority is 0.45 log wage points. Nevertheless, the discrete change in the base wage at the new job is composed of: (a) a fixed increase of 0.19 log wage points (Table 4, line 15); (b) a variable increase proportional to seniority at point of change of 0.16 log wage points (Table 4, line 18); and (c) a variable loss of 0.06 (Table 4, line 22) due to *career effect* (i.e., experience at the moment of mobility). The overall effect is a loss of approximately 0.16 log points (Figure 1d). The loss for college graduates with substantial labor market experience and a long tenure spell is somewhat smaller (Figure 2d).

7 Summary and Conclusions

Various theories in economics, and especially the well-known human capital theory, predict that wage compensation should rise with seniority in a firm. The existence of firm-specific human capital may explain the prevalence of long-term relationships between employees and employers. While the various theories have very similar implications, there is much disagreement about the empirical evidence. In a seminal paper, Topel (1991) examines this fundamental question and finds that there are significant returns to seniority. This finding provides strong support for the theoretical literature on human capital and optimal labor contracts. Furthermore he argues that, if anything, his estimates are likely to be downward biased. This finding stands in stark contrast with most of the previous studies in the literature. By and large, most studies have concluded that, at best, the evidence supporting positive returns to seniority is weak. In particular, Altonji and Williams (2005) confirm earlier findings of Altonji and Shakotko (1987) and find negligible, and largely insignificant, returns to seniority.

In this study we re-open the debate on the returns to seniority, while adopting new methodological advancements that allow us to take a new, and more convincing, look at this issue. In contrast to other studies, we explicitly model the joint decisions of participation and job mobility, along with the key outcome, namely the wage outcome. We start with the same general Mincer's wage

specification as previously adopted by the literature, while allowing for potential sample selection biases that stem from the participation and mobility decisions of individuals. These estimating equations are the structural counterpart of our theoretical model, an extension of Hyslop (1999) with the addition of elements from the job search theory. Our strategy allows us to more rigorously examine the question about the magnitude of the returns to seniority in the United States. Moreover, it allows us to demonstrate the reasons for the differences between the results previously obtained in the literature and compare them with our findings. In particular, we introduce random correlated person-specific effects in all three equations. We also allow these terms to be correlated across equations. This specification, along with the fact that we allow the idiosyncratic terms in the three equations to be correlated, makes it possible to trace the implied biases that stem from estimating the wage equation by itself.

We use a data extract similar to that used by both Topel (1991) and Altonji and Williams (2005), namely the Panel Study of Income Dynamics (PSID). We resort to a Bayesian estimation, which extensively uses Markov Chain Monte Carlo methods, allowing one to compute the posterior distribution of the model’s parameters. Whenever possible we use uninformative prior distributions for the parameters and hence rely heavily on the data to determine the posterior distributions of these parameters. We perform our analysis on three educational groups: (a) *high school dropouts*; (b) *high school graduates*; and (c) *college graduates*.

The results are unequivocal: We find that the returns to seniority are large and highly significant for all three educational groups. In fact, the support of the posterior distribution of the returns to seniority is entirely in the positive segment of the real line for all three educational groups. Our results clearly demonstrate the importance of jointly estimating the wage equation along with the participation-employment and mobility decisions. These two decisions have significant effects on the observed annual earnings. Moreover, we find that in addition to the direct effects of experience and seniority on one’s wages, there is a significant effect on the individual’s specific career path through the affect on the starting wage at a new job. Because careers differ widely across workers with otherwise similar backgrounds, it seems crucial to take into account this feature when estimating a wage equation. Overall, we find very strong evidence supporting Topel’s (1991) claims, even though some aspects of our modeling strategy are closer to Altonji and Williams (2005).

We find that several factors affect the results considerably. First, estimating the model with and without taking into account the workers’ careers paths provides very different results. Nevertheless, the various specifications only change the magnitude of the results but do not change the qualitative results. That is, the returns to seniority are large and significant for all groups considered. Our estimates of the returns to experience are somewhat larger, though at the same general order of magnitude than those previously obtained in the literature. However, these findings are not uniform across education groups; they are much higher for the college graduates than for the other two education groups. In addition, the pattern of job-to-job mobility—as summarized by the non-linear function J^W —has a strong impact on wages at the entry stage in a new firm. This impact of the career on entry wages differs markedly across education groups. In particular, we find that the

timing of a move in a career matters. While early moves are most beneficial to college-educated workers, late moves are more detrimental to workers with lower education. Overall, wage growth is achieved through a combination of wage increases within the firm and by interfirm mobility. The former is more important for wage growth of high school dropouts because of their lower returns to experience. The latter is more important for college graduates, both because the returns to seniority are larger during the first years in a firm and because there is no penalty implied by job-to-job mobility.

8 References

- Abraham, K.G., and H.S. Farber (1987): “Job Duration, Seniority and Earnings,” *American Economic Review*, 77, 278–97.
- Addison, J.T., and P. Portugal (1989): “Job Displacement, Relative Wage Changes, and Duration of Unemployment,” *Journal of Labor Economics*, 7, 281–302.
- Altonji, J., and R. Shakotko (1987): “Do Wages Rise with Job Seniority?” *Review of Economic Studies*, 54, 437–59.
- Altonji, J., and N. Williams (2005): “Do Wages Rise with Job Seniority? A Reassessment,” *Industrial and Labor Relations Review*, 58, 370–97.
- Becker, G. (1964): *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, NBER, Cambridge, MA.
- Bontemps, C., Robin, J. M., and G. Van den Berg (1999): “An Empirical Equilibrium Search Model with Continuously Distributed Heterogeneity of Workers’ Opportunity Cost of Employment and Firms’ Productivities, and Search on the Job,” *International Economic Review*, 40, 1039–74.
- Buchinsky, M., Fougère, D., Kramarz F. and R. Tchernis (2005): “Interfirm Mobility, Wages, and the Returns to Seniority and Experience in the U.S.,” IZA Discussion Paper No. 1521.
- Burdett, K. (1978): “A Theory of Employee Job Search and Quit Rates,” *American Economic Review*, 68, 212–20.
- Card, D. (2001): “Estimating the Return to Schooling: Progress on Some Persistent Econometrics Problem,” *Econometrica*, 69, 1127–60.
- Chib, S. (2001): “Markov Chain Monte Carlo Methods: Computation and Inference,” Chapter 57 in *Handbook of Econometrics*, Volume 5, ed. by J. J. Heckman and E. Leamer. Amsterdam: North-Holland, 3569–49.
- Chib, S. and B.P. Carlin (1999): “On MCMC Sampling in Hierarchical Longitudinal Models,” *Statistics and Computing*, 9, 17–26.
- Chib, S. and E. Greenberg (1995): “Understanding the Metropolis-Hastings Algorithm,” *American Statistician*, 49, 327–35.
- Chib, S. and E. Greenberg (1996): “Markov Chain Monte Carlo Simulation Methods in Econometrics,” *Econometric Theory*, 12, 409–31.
- Chib, S. and E. Greenberg (1998): “Analysis of Multivariate Probit Models,” *Biometrika*, 85, 347–61.
- Chib, S. and B. H. Hamilton (2002): “Semiparametric Bayes Analysis of Longitudinal Data Treatment Models,” *Journal of Econometrics*, 110, 67–89.
- Dustmann, C. and C. Meghir (2005): “Wages, Experience and Seniority,” *Review of Economic Studies*, 72, 77–108.

- Farber, H. (1999): "Mobility and Stability," Chapter 37 in *Handbook of Labor Economics*, Volume 3B, ed. by O. Ashenfelter and D. Card. New-York: North-Holland, 2441–83.
- Fougère, D. and T. Kamionka (2003): "Bayesian Inference for the Mover-Stayer Model in Continuous Time, with an Application to Labour Market Transition Data," *Journal of Applied Econometrics*, 18, 697–723.
- Gelfand, A. E. and A. F. M. Smith (1990): "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.
- Geweke, J. and M. Keane (2000): "An Empirical Analysis of Earnings Dynamics Among Men in the PSID: 1968–1989," *Journal of Econometrics*, 96, 293–356.
- Gibbons, R. and L. Katz (1991): "Layoffs and Lemons," *Journal of Labor Economics*, 9, 351–80.
- Gouriéroux, C. and A. Monfort (1996): *Simulation-Based Econometric Methods*, CORE Lectures. Oxford and New York: Oxford University Press.
- Hardman, A. and Y.M. Ioannides (2004): "Neighbors' Income Distribution: Economic Segregation and Mixing in US Urban Neighborhoods," *Journal of Housing Economics*, 13, 368–82.
- Heckman, J. (1981): "Heterogeneity and State Dependence," in *Studies in Labor Markets*, ed. by S. Rosen. Chicago: University of Chicago Press.
- Hill, M. S. (1992): *The Panel Study of Income Dynamics: A User's Guide*. Newbury Park, CA: Sage Publications.
- Hyslop D. R. (1999): "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica*, 67, 1255–94.
- Jacobson, L., Lalonde, L., and D. Sullivan (1993): "Earnings Losses of Displaced Workers," *American Economic Review*, 83, 685–709.
- Light A. and M. Ureta (1995): "Early-Career Work Experience and Gender Wage Differentials," *Journal of Labor Economics*, 3, 121–54.
- McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57, 995–1026.
- Mincer, J. (1988): "Job Training, Wage Growth, and Labor Turnover," National Bureau of Economic Research Working Paper: 2690.
- Mroz, T.A. (1987): "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, 55, 765–99.
- Nagypal E. (2005): "On the Extent of Job-to-Job Transitions," mimeo, Department of Economics, Northwestern University.
- Neal, D. (1995): "Industry-Specific Human Capital: Evidence from Displaced Workers," *Journal of Labor Economics*, 13, 653–77.
- Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–57.

- Parent, D. (1999): “Wages and Mobility: The Impact of Employer-Provided Training,” *Journal of Labor Economics*, 17, 298-317.
- Parent, D. (2000): “Industry-Specific Capital and the Wage Profile: Evidence from the National Longitudinal Survey of Youth and the Panel Study of Income Dynamics,” *Journal of Labor Economics*, 18, 306-23.
- Postel-Vinay, F. and J. M. Robin (2002): “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity,” *Econometrica*, 70, 1295-350.
- Robert, C. and G. Casella (1999): *Monte Carlo Statistical Methods*. New-York: Springer-Verlag.
- Topel, R.H. (1991): “Specific Capital, Mobility and Wages: Wages Rise with Job Seniority,” *Journal of Political Economy*, 99, 145–76.
- Topel, R.H. and M.P. Ward (1992): “Job Mobility and the Careers of Young Men,” *Quarterly Journal of Economics*, 107, 439–479.

Appendix A—Mathematical Derivation

A.1 Graphical Representation of Value Functions

Enter Figure A.1 and Figure A.2 here.

A.2 Relationships Between Reservation Wages

The reservation wages for non-participants and stayers, in (12) and (15), imply that

$$\begin{aligned}
 & u(z_t + w_{11,t}^*, 1; X_t) - u(z_t + w_{01,t}^* - \gamma_1, 1; X_t) \\
 &= u(z_t, 0; X_t) - u(z_t - \gamma_1, 0; X_t) \\
 &= u(z_t + w_{12,t}^*, 1; X_t) - u(z_t + w_{02,t}^* - \gamma_1, 1; X_t).
 \end{aligned} \tag{39}$$

First-order Taylor series expansions of both sides of (39) around $z_t + w_{01,t}^*$, $z_t + w_{02,t}^*$, and z_t , respectively, give

$$w_{1j,t}^* \approx w_{0j,t}^* - \gamma_1 \left[1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{0j,t}^*, 1; X_t)} \right] = w_{0j,t}^* - \gamma_{1j}, \tag{40}$$

for $j = 1, 2$, where $u'(\cdot)$ denotes the marginal utility of consumption, and

$$\gamma_{1j} = \gamma_1 \left[1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{0j,t}^*, 1; X_t)} \right].$$

If the utility function is concave with respect to consumption, then $u'(z_t + w_{0j,t}^*, 0; X_t) < u'(z_t, 0; X_t)$. However, if the marginal utility of consumption is greater when working, i.e., $u'(z_t + w_{0j,t}^*, 0; X_t) < u'(z_t + w_{0j,t}^*, 1; X_t)$, then γ_{1j} may be positive or negative, dependent on whether $u'(z_t + w_{0j,t}^*, 1; X_t)$ is greater or smaller than $u'(z_t, 0; X_t)$. In contrast, if the marginal utility of consumption is lower when working, then γ_{1j} is always negative, since $u'(z_t + w_{0j,t}^*, 1; X_t) < u'(z_t, 0; X_t)$.

A similar inspection of the reservation wages for non-participants and movers, in (12) and (17), gives

$$\begin{aligned}
 & u(z_t + w_{21,t}^* - c_M, 1; X_t) - u(z_t + w_{01,t}^* - \gamma_1, 1; X_t) \\
 &= u(z_t - c_M, 0; X_t) - u(z_t - \gamma_1, 0; X_t) \\
 &= u(z_t + w_{22,t}^* - c_M, 1; X_t) - u(z_t + w_{02,t}^* - \gamma_1, 1; X_t).
 \end{aligned} \tag{41}$$

Again, first-order Taylor series expansions of the left and right hand sides of (41) around $z_t + w_{0j,t}^*$,

$z_t + w_{2j,t}^*$, and z_t , respectively, give (for $j = 1, 2$):

$$\begin{aligned}
u(z_t + w_{2j,t}^* - c_M, 1; X_t) &\simeq u(z_t + w_{2j,t}^*, 1; X_t) - c_M u'(z_t + w_{2j,t}^*, 1; X_t), \\
u(z_t + w_{0j,t}^* - \gamma_1, 1; X_t) &\simeq u(z_t + w_{0j,t}^*, 1; X_t) - \gamma_1 u'(z_t + w_{0j,t}^*, 1; X_t), \\
u(z_t - c_M, 0; X_t) &\simeq u(z_t, 0; X_t) - c_M u'(z_t, 0; X_t), \\
u(z_t - \gamma_1, 0; X_t) &\simeq u(z_t, 0; X_t) - \gamma_1 u'(z_t, 0; X_t), \quad \text{and} \\
u(z_t + w_{2j,t}^*, 1; X_t) - u(z_t + w_{0j,t}^*, 1; X_t) &\simeq (w_{2j,t}^* - w_{0j,t}^*) u'(z_t + w_{0j,t}^*, 1; X_t).
\end{aligned} \tag{42}$$

Substitution of the expressions from (42) into (41) gives (for $j = 1, 2$):

$$\begin{aligned}
w_{2j,t}^* &\approx w_{0j,t}^* - \gamma_{1j} + \gamma_{2j} \approx w_{1j,t}^* + \gamma_{2j}, \quad \text{where} \\
\gamma_{2j} &= c_M \left[\frac{u'(z_t + w_{2j,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{0j,t}^*, 1; X_t)} \right], \quad \text{and} \\
\gamma_{1j} &= \gamma_1 \left[1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{0j,t}^*, 1; X_t)} \right].
\end{aligned} \tag{43}$$

Note that if the marginal utility of consumption is greater when working, i.e., $u'(z_t + w_{2j,t}^*, 0; X_t) < u'(z_t + w_{2j,t}^*, 1; X_t)$, then γ_{2j} can be positive or negative:

$$\gamma_{2j} \geq 0 \quad \Leftrightarrow \quad u'(z_t + w_{2j,t}^*, 1; X_t) \geq u'(z_t, 0; X_t).$$

However, if the marginal utility of consumption is lower when working, then γ_{2j} must be negative, since

$$u'(z_t + w_{2j,t}^*, 1; X_t) < u'(z_t + w_{2j,t}^*, 0; X_t) < u'(z_t, 0; X_t).$$

If the marginal utility of consumption is the same in both states (employment and non-employment), then

$$u'(z_t + w_{2j,t}^*, 1; X_t) = u'(z_t + w_{2j,t}^*, 0; X_t) < u'(z_t, 0; X_t),$$

and hence, $\gamma_{1j} < 0$ and $\gamma_{2j} < 0$.

First-order Taylor series expansions of $u(z_t + w_{13,t}^*, 1; X_t)$ and $u(z_t + w_{23,t}^*, 1; X_t)$ around $z_t + w_{03,t}^*$ and $z_t + w_{13,t}^*$ respectively, give

$$w_{13,t}^* \approx w_{03,t}^* + \frac{u(z_t + w_{13,t}^*, 1; X_t) - u(z_t + w_{03,t}^*, 1; X_t)}{u'(z_t + w_{03,t}^*, 1; X_t)} = w_{03,t}^* + \gamma_{13}$$

and

$$\begin{aligned}
w_{23,t}^* &\approx w_{13,t}^* + \frac{u(z_t + w_{23,t}^*, 1; X_t) - u(z_t + w_{13,t}^*, 1; X_t)}{u'(z_t + w_{13,t}^*, 1; X_t)} = w_{13,t}^* + \gamma_{23} \\
&\approx w_{03,t}^* + \gamma_{13} + \gamma_{23},
\end{aligned}$$

where

$$\gamma_{13} = \frac{u(z_t + w_{13,t}^*, 1; X_t) - u(z_t + w_{03,t}^*, 1; X_t)}{u'(z_t + w_{03,t}^*, 1; X_t)}$$

and

$$\gamma_{23} = \frac{u(z_t + w_{23,t}^*, 1; X_t) - u(z_t + w_{13,t}^*, 1; X_t)}{u'(z_t + w_{13,t}^*, 1; X_t)}.$$

For determining the sign of γ_{13} (respectively, γ_{23}), we need to know whether $w_{13,t}^*$ (respectively., $w_{23,t}^*$) is lower or higher than $w_{03,t}^*$ (respectively., $w_{13,t}^*$).

A.3 Proofs of Conditions 1 and 2

As in Section 3, we assume that: (a) $c_M > \gamma_1$; and (b) $u(\cdot)$ is concave. We then consider four possible cases.

A. $w_{02,t}^* < w_{01,t}^*$ and the marginal utility of consumption is higher when working: The concavity of $u(\cdot)$ with respect to consumption implies that

$$\begin{aligned} u'(z_t + w_{01,t}^*, 1; X_t) &< u'(z_t + w_{02,t}^*, 1; X_t), \quad \text{and hence} \\ \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)} &< \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)}. \end{aligned} \quad (44)$$

A.1. If $w_{02,t}^* < w_{22,t}^*$: The concavity of the utility function implies that

$$\begin{aligned} u'(z_t + w_{22,t}^*, 1; X_t) &< u'(z_t + w_{02,t}^*, 1; X_t), \quad \text{and hence} \\ \frac{u'(z_t + w_{22,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)} &< 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)}. \end{aligned} \quad (45)$$

1. If $u'(z_t, 0; X_t) > u'(z_t + w_{02,t}^*, 1; X_t)$ then $u'(z_t, 0; X_t) > u'(z_t + w_{01,t}^*, 1; X_t)$ and (44) implies that $\gamma_{11} < \gamma_{12} < 0$. Consequently, (40) implies that $w_{02,t}^* < w_{12,t}^* < w_{11,t}^*$ and $w_{02,t}^* < w_{01,t}^* < w_{11,t}^*$. Moreover, since $w_{02,t}^* < w_{22,t}^*$, we have $u'(z_t, 0; X_t) > u'(z_t + w_{22,t}^*, 1; X_t)$. Hence, from (45), it follows that $\gamma_{22} < \gamma_{12} < 0$. Consequently (43) implies then that $w_{22,t}^* < w_{02,t}^*$, which contradicts the initial assumption. Thus, $w_{22,t}^*$ cannot be greater than $w_{02,t}^*$ when $u'(z_t, 0; X_t) > u'(z_t + w_{02,t}^*, 1; X_t)$.

2. If $u'(z_t, 0; X_t) < u'(z_t + w_{01,t}^*, 1; X_t)$ then $u'(z_t, 0; X_t) < u'(z_t + w_{02,t}^*, 1; X_t)$ and $0 < \gamma_{11} < \gamma_{12}$. Thus, (40) implies that $w_{12,t}^* < w_{11,t}^* < w_{01,t}^*$ and $w_{12,t}^* < w_{02,t}^* < w_{01,t}^*$. Two cases must be considered then.

Case 1: If $u'(z_t + w_{22,t}^*, 1; X_t) < u'(z_t, 0; X_t)$, then $\gamma_{22} < 0$ and $\gamma_{12} > 0$. Thus $\gamma_{22} - \gamma_{12} < 0$, and (43) implies then that $w_{22,t}^* < w_{02,t}^*$, which contradicts the initial assumption.

Case 2: If $u'(z_t, 0; X_t) < u'(z_t + w_{22,t}^*, 1; X_t)$, then

$$0 < \frac{u'(z_t + w_{22,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)} < 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)}.$$

Thus, $w_{22,t}^* > w_{02,t}^*$ if and only if $\gamma_{22} - \gamma_{12} > 0$, or alternatively, if and only if

$$c_M > \gamma_1 \left[\frac{u'(z_t + w_{02,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{22,t}^*, 1; X_t) - u'(z_t, 0; X_t)} \right]. \quad (46)$$

The expression in square brackets is greater than 1, and hence it is a stronger assumption than $c_M > \gamma_1$. Nevertheless, it is a sufficient condition to have $w_{22,t}^* > w_{02,t}^*$. Moreover, (42) implies that $w_{22,t}^* > w_{02,t}^* > w_{12,t}^*$. If $w_{21,t}^* > w_{22,t}^*$, which is satisfied if $\gamma_{21} > \gamma_{22}$, a mover becomes a non-participant at the end of period t if he/she is offered a wage less than $w_{22,t}^*$. A stayer becomes a non-participant if he or she is offered a wage less than $w_{12,t}^*$, because $w_{12,t}^* < w_{11,t}^*$. Thus, the participation decision at period t can be characterized by the equation

$$\begin{aligned} y_t &= \mathbf{1} [w_t > w_{02,t}^* - \gamma_{12} y_{t-1} + \gamma_{22} y_{t-1} m_{t-1}] \\ &= \mathbf{1} [w_t - w_{02,t}^* + \gamma_{12} y_{t-1} - \gamma_{22} y_{t-1} m_{t-1} > 0], \end{aligned} \quad (47)$$

with $\gamma_{22} > \gamma_{12} > 0$. If we assume that, in each period t ,

$$\frac{d}{dw_t} V_t^2(1, k; X_t) \leq \frac{d}{dw_t} V_t^1(1, k; X_t), \quad k = 0, 1, \quad (48)$$

then, for employed workers who were stayers and movers in period $t - 1$, there exists two wage values, denoted $w_{13,t}^*$ and $w_{23,t}^*$ respectively, that equate the value function of participation without interfirm mobility in period t with that of participation with interfirm mobility in period t . These wage thresholds are defined by

$$V_t^1(1, 0; X_t | w_{13,t}^*) = V_t^2(1, 0; X_t | w_{13,t}^*)$$

and

$$V_t^1(1, 1; X_t | w_{23,t}^*) = V_t^2(1, 1; X_t | w_{23,t}^*).$$

Consequently, a stayer (at $t - 1$) may decide to move to another firm at the end of period t if he/she is offered a wage w_t^* such that $w_{12,t}^* < w_t^* < w_{13,t}^*$, but he/she will stay in the current firm for at least one more period if $w_t^* > w_{13,t}^*$. On the other hand, a mover (at $t - 1$) will decide to move again at t only if the wage offer is such that $w_{22,t}^* < w_t^* < w_{23,t}^*$ and he/she will stay in the firm for at least one more period if $w_t^* > w_{23,t}^*$. Consequently, the decision to stay in the current firm at end of period t can be characterized by the dummy variable s_t defined as:

$$\begin{aligned} s_t &= 1 - m_t = \mathbf{1} [w_t > w_{13,t}^* + \gamma_{23} y_{t-1} m_{t-1}] \\ &= \mathbf{1} [w_t - w_{13,t}^* - \gamma_{23} y_{t-1} m_{t-1} > 0], \end{aligned} \quad (49)$$

where m_t is the dummy variable indicating whether the worker i decides to move ($m_t = 1$) or not ($m_t = 0$) at the end of period t .

3. If $u'(z_t + w_{01,t}^*, 1; X_t) < u'(z_t, 0; X_t) < u'(z_t + w_{02,t}^*, 1; X_t)$, then (44) implies that $\gamma_{11} < 0 < \gamma_{12}$. Thus, (40) implies that $w_{12,t}^* < w_{02,t}^* < w_{01,t}^* < w_{11,t}^*$. Note that two cases are possible.

Case 1: If $u'(z_t + w_{22,t}^*, 1; X_t) < u'(z_t, 0; X_t)$, then $\gamma_{22} < 0$. Thus $\gamma_{22} - \gamma_{12} < 0$ and (43) implies that $w_{22,t}^* < w_{02,t}^*$, which contradicts the initial assumption.

Case 2: If $u'(z_t, 0; X_t) < u'(z_t + w_{22,t}^*, 1; X_t)$, then

$$0 < \frac{u'(z_t + w_{22,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)} < 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)}.$$

Thus, $w_{22,t}^* > w_{02,t}^*$ if and only if $\gamma_{22} - \gamma_{12} > 0$, or alternatively, if and only if, as in (46):

$$c_M > \gamma_1 \left[\frac{u'(z_t + w_{02,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{22,t}^*, 1; X_t) - u'(z_t, 0; X_t)} \right].$$

In turn, (42) implies then that $w_{22,t}^* > w_{02,t}^* > w_{12,t}^*$, and if $\gamma_{21} > \gamma_{22}$, then $w_{21,t}^* > w_{22,t}^*$. Consequently, if condition (48) is verified, (47) and (49) are still valid.

A.2. If $w_{02,t}^* > w_{22,t}^*$: The concavity of the utility function implies that $u'(z_t + w_{22,t}^*, 1; X_t) > u'(z_t + w_{02,t}^*, 1; X_t)$. Thus,

$$\frac{u'(z_t + w_{22,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)} > 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)}. \quad (50)$$

Again, two cases are possible.

Case 1: If $u'(z_t + w_{22,t}^*, 1; X_t) < u'(z_t, 0; X_t)$, then (50) implies that $0 > \gamma_{22} > \gamma_{12}$. Then, (43) implies that $w_{22,t}^* > w_{02,t}^*$, which contradicts the assumption.

Case 2: If $u'(z_t, 0; X_t) < u'(z_t + w_{22,t}^*, 1; X_t)$, then (50) implies that $\gamma_{22} > \gamma_{12} > 0$. Then, (43) implies that $w_{22,t}^* > w_{02,t}^*$, which contradicts the above assumption.

Consequently, the assumption that $w_{02,t}^* > w_{22,t}^*$ cannot be true if the mobility cost c_M is strictly greater than the search cost γ_1 .

B. $w_{02,t}^* < w_{01,t}^*$ and the marginal utility of consumption is lower when working (or equal in both states): If the the marginal utility is either lower when working, or equal in both states, then

$$u'(z_t + w_{jl,t}^*, 1; X_t) < u'(z_t, 1; X_t) \leq u'(z_t, 0; X_t), \quad j = 0, 1, 2; \quad l = 1, 2.$$

Moreover, if $w_{02,t}^* < w_{01,t}^*$, then $u'(z_t + w_{01,t}^*, 1; X_t) < u'(z_t + w_{02,t}^*, 1; X_t) < u'(z_t, 0; X_t)$, which implies that $\gamma_{11} < \gamma_{12} < 0$. Then, (40) implies that $w_{02,t}^* < w_{01,t}^* < w_{11,t}^*$ and $w_{02,t}^* < w_{12,t}^* < w_{11,t}^*$.

B.1. If $w_{02,t}^* < w_{22,t}^*$, then $u'(z_t + w_{22,t}^*, 1; X_t) < u'(z_t + w_{02,t}^*, 1; X_t)$ and

$$\frac{u'(z_t + w_{22,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)} < 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)} < 0. \quad (51)$$

Thus, (51) implies that $\gamma_{22} < \gamma_{12} < 0$. From (42), we infer that $w_{22,t}^* < w_{02,t}^*$, which contradicts the assumption above.

B.2. If $w_{02,t}^* > w_{22,t}^*$, then $u'(z_t + w_{22,t}^*, 1; X_t) > u'(z_t + w_{02,t}^*, 1; X_t)$ and

$$0 > \frac{u'(z_t + w_{22,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)} > 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)}.$$

Thus, (43) implies that $w_{02,t}^* > w_{22,t}^*$ if and only if

$$c_M > \gamma_1 \left[\frac{u'(z_t + w_{02,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{22,t}^*, 1; X_t) - u'(z_t, 0; X_t)} \right].$$

Under this condition, (42) implies that $w_{12,t}^* > w_{02,t}^* > w_{22,t}^*$ if $0 > \gamma_{21} > \gamma_{22}$. Then, $w_{21,t}^* > w_{22,t}^*$ and, hence, if condition (48) is verified, (47) and (49) are still valid.

C. $w_{01,t}^* < w_{02,t}^*$ and the marginal utility of consumption is higher when working: This condition implies that

$$\frac{u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)} < \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{02,t}^*, 1; X_t)}. \quad (52)$$

C.1. If $w_{01,t}^* < w_{21,t}^*$, then the concavity of $u(\cdot)$ implies that $u'(z_t + w_{21,t}^*, 1; X_t) < u'(z_t + w_{01,t}^*, 1; X_t)$. Thus,

$$\frac{u'(z_t + w_{21,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)} < 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)}. \quad (53)$$

1. If $u'(z_t, 0; X_t) > u'(z_t + w_{01,t}^*, 1; X_t)$, then $u'(z_t, 0; X_t) > u'(z_t + w_{02,t}^*, 1; X_t)$, and (52) implies that $\gamma_{12} < \gamma_{11} < 0$. Then, (40) implies that $w_{01,t}^* < w_{11,t}^* < w_{12,t}^*$ and $w_{01,t}^* < w_{02,t}^* < w_{12,t}^*$. Moreover, since $w_{01,t}^* < w_{21,t}^*$, it follows that

$$u'(z_t, 0; X_t) > u'(z_t + w_{01,t}^*, 1; X_t) > u'(z_t + w_{21,t}^*, 1; X_t).$$

By (53), this inequality implies that $\gamma_{21} < \gamma_{11} < 0$. Thus, (43) implies then that $w_{21,t}^* < w_{01,t}^*$, which contradicts the initial assumption. Hence, $w_{21,t}^*$ cannot be greater than $w_{01,t}^*$ when $u'(z_t, 0; X_t) > u'(z_t + w_{01,t}^*, 1; X_t)$.

2. If $u'(z_t, 0; X_t) < u'(z_t + w_{02,t}^*, 1; X_t)$, then $u'(z_t, 0; X_t) < u'(z_t + w_{01,t}^*, 1; X_t)$ and $0 < \gamma_{12} < \gamma_{11}$. Thus, (40) implies that $w_{11,t}^* < w_{12,t}^* < w_{02,t}^*$ and $w_{11,t}^* < w_{01,t}^* < w_{02,t}^*$. Two cases must be considered.

Case 1: If $u'(z_t + w_{21,t}^*, 1; X_t) < u'(z_t, 0; X_t)$, then $\gamma_{21} < 0$ and $\gamma_{11} > 0$. Thus $\gamma_{21} - \gamma_{11} < 0$, and (43) implies that $w_{21,t}^* < w_{01,t}^*$, which contradicts the initial assumption.

Case 2: If $u'(z_t, 0; X_t) < u'(z_t + w_{21,t}^*, 1; X_t)$, then

$$0 < \frac{u'(z_t + w_{21,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)} < 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)}.$$

Thus, $w_{21,t}^* > w_{01,t}^*$ if and only if $\gamma_{21} - \gamma_{11} > 0$, or alternatively, if and only if

$$c_M > \gamma_1 \left[\frac{u'(z_t + w_{01,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{21,t}^*, 1; X_t) - u'(z_t, 0; X_t)} \right]. \quad (54)$$

As before the expression in square brackets is greater than 1, which is a stronger assumption than $c_M > \gamma_1$. Nevertheless, it is a sufficient condition to have $w_{21,t}^* > w_{01,t}^*$. Moreover, (42) implies that $w_{21,t}^* > w_{01,t}^* > w_{11,t}^*$. Now, assume that $\gamma_{22} > \gamma_{21}$, then $w_{22,t}^* > w_{21,t}^*$. In that case, there is

no inter-firm mobility at t . That is, a non-participant becomes employed (respectively, stays in the non-participation state) at the end of period t if he/she is offered a wage greater (respectively, lower) than $w_{01,t}^*$. A participant becomes a non-participant (respectively, remains employed) if he/she is offered a wage less than $w_{11,t}^*$. Thus, the participation decision at period t can be characterized by

$$\begin{aligned} y_t &= \mathbf{1} [w_t > w_{01,t}^* - \gamma_{11} y_{t-1}] \\ &= \mathbf{1} [w_t - w_{01,t}^* + \gamma_{11} y_{t-1} > 0], \end{aligned} \quad (55)$$

with $\gamma_{11} > 0$.

3. If $u'(z_t + w_{02,t}^*, 1; X_t) < u'(z_t, 0; X_t) < u'(z_t + w_{01,t}^*, 1; X_t)$, then (52) implies that $\gamma_{12} < 0 < \gamma_{11}$, and (40) implies that $w_{11,t}^* < w_{01,t}^* < w_{02,t}^* < w_{12,t}^*$. Again, there are two possible cases.

Case 1: If $u'(z_t + w_{21,t}^*, 1; X_t) < u'(z_t, 0; X_t)$, then $\gamma_{21} < 0$. Thus, $\gamma_{21} - \gamma_{11} < 0$, and (43) implies that $w_{21,t}^* < w_{01,t}^*$, which contradicts the initial assumption.

Case 2: If $u'(z_t, 0; X_t) < u'(z_t + w_{21,t}^*, 1; X_t)$, then

$$0 < \frac{u'(z_t + w_{21,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)} < 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)}.$$

Thus, $w_{21,t}^* > w_{01,t}^*$ if and only if $\gamma_{21} - \gamma_{11} > 0$, or alternatively, if and only if

$$c_M > \gamma_1 \left[\frac{u'(z_t + w_{01,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{21,t}^*, 1; X_t) - u'(z_t, 0; X_t)} \right].$$

This is the same condition as in (54). Together with (42) it implies that $w_{21,t}^* > w_{01,t}^* > w_{11,t}^*$, and if $\gamma_{22} > \gamma_{21}$, then $w_{22,t}^* > w_{21,t}^*$, and (55) is still valid.

C.2. If $w_{01,t}^* > w_{21,t}^*$: The concavity of $u(\cdot)$ implies that $u'(z_t + w_{21,t}^*, 1; X_t) > u'(z_t + w_{01,t}^*, 1; X_t)$. Thus,

$$\frac{u'(z_t + w_{21,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)} > 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)}. \quad (56)$$

Note that there are two possible cases.

Case 1: If $u'(z_t + w_{21,t}^*, 1; X_t) < u'(z_t, 0; X_t)$, then (56) implies that $0 > \gamma_{21} > \gamma_{11}$. Then, (43) implies that $w_{21,t}^* > w_{01,t}^*$, which contradicts the assumption above.

Case 2: If $u'(z_t, 0; X_t) < u'(z_t + w_{21,t}^*, 1; X_t)$, then (56) implies that $\gamma_{21} > \gamma_{11} > 0$. Then, (43) implies that $w_{21,t}^* > w_{01,t}^*$, which still contradicts the assumption above.

Consequently, the assumption $w_{01,t}^*$ cannot be greater than $w_{21,t}^*$.

D. $w_{01,t}^* < w_{02,t}^*$ and the marginal utility of consumption is lower when working (or equal in both states): If the marginal utility is either lower when working, or equal in both states, then $u'(z_t + w_{j,l,t}^*, 1; X_t) < u'(z_t, 1; X_t) \leq u'(z_t, 0; X_t)$, for $j = 0, 1, 2; l = 1, 2$. Moreover, if $w_{01,t}^* < w_{02,t}^*$, then $u'(z_t + w_{02,t}^*, 1; X_t) < u'(z_t + w_{01,t}^*, 1; X_t) < u'(z_t, 0; X_t)$. This implies that $\gamma_{12} < \gamma_{11} < 0$. Then, (40) implies that $w_{01,t}^* < w_{02,t}^* < w_{12,t}^*$ and $w_{01,t}^* < w_{11,t}^* < w_{12,t}^*$.

D.1. If $w_{01,t}^* < w_{21,t}^*$, then $u'(z_t + w_{21,t}^*, 1; X_t) < u'(z_t + w_{01,t}^*, 1; X_t)$ and

$$\frac{u'(z_t + w_{21,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)} < 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)} < 0. \quad (57)$$

Thus, (57) implies that $\gamma_{21} < \gamma_{11} < 0$. In turn, (42) implies that $w_{21,t}^* < w_{01,t}^*$, which contradicts the assumption above.

D.2. If $w_{01,t}^* > w_{21,t}^*$, then $u'(z_t + w_{21,t}^*, 1; X_t) > u'(z_t + w_{01,t}^*, 1; X_t)$ and

$$0 > \frac{u'(z_t + w_{21,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)} > 1 - \frac{u'(z_t, 0; X_t)}{u'(z_t + w_{01,t}^*, 1; X_t)}.$$

Thus, (43) implies that $w_{01,t}^* > w_{21,t}^*$ if and only if

$$c_M > \gamma_1 \left[\frac{u'(z_t + w_{01,t}^*, 1; X_t) - u'(z_t, 0; X_t)}{u'(z_t + w_{21,t}^*, 1; X_t) - u'(z_t, 0; X_t)} \right].$$

Under this condition, (42) implies then that $w_{11,t}^* > w_{01,t}^* > w_{21,t}^*$. If $0 > \gamma_{22} > \gamma_{21}$, then $w_{22,t}^* > w_{21,t}^*$. Once again, the ranking of the reservation wages implies that there is no interfirm mobility, and the participation decision at period t is simply characterized by (55).

Figure A.1: Value Function, Alternative 1

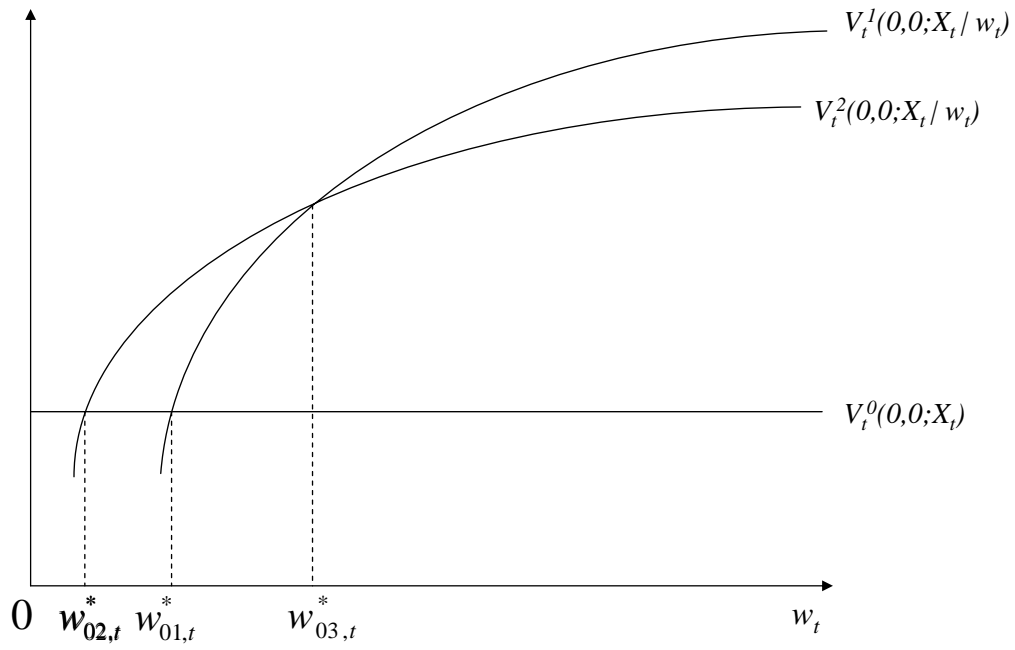
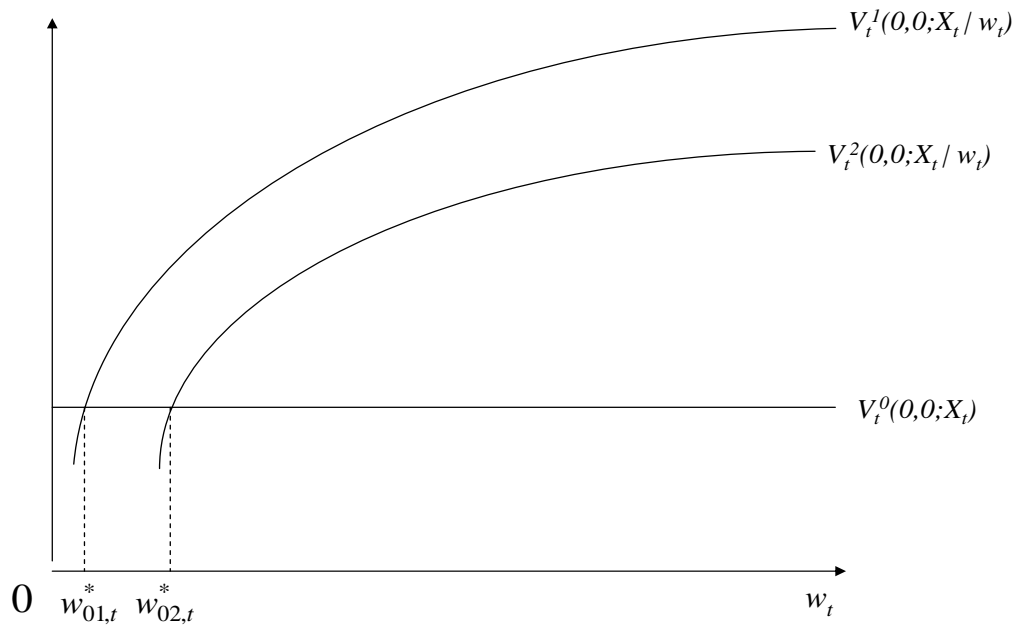


Figure A.2: Value Function, Alternative 2



Appendix B—Drawing from the Posterior Distribution

Following the notation of Section (4.2), note that in matrix form we can write the model in (29), (30), and (31) as

$$z_{it}^* = \tilde{x}_{it}\beta + L_t\alpha_i + \tau_{it}, \quad (58)$$

for $t = 1, \dots, T$, where $\alpha_i \sim N(f(x_{i1}, \dots, x_{iT}), \Gamma_i)$, as is defined in (34), $\tau_{it} \sim N(0, \Sigma)$, as defined in (37).

For $t = 1$ we define

$$\tilde{x}_{i1} = \begin{pmatrix} x_{yi1} & 0 & 0 & 0 & 0 \\ 0 & x_{mi1} & 0 & 0 & 0 \\ 0 & 0 & x_{wi1} & 0 & 0 \end{pmatrix} \quad \text{and} \quad L_1 = \begin{pmatrix} \delta_1 & 0 & 0 \\ 0 & \delta_2 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

while for $t > 1$ we define

$$\tilde{x}_{it} = \begin{pmatrix} 0 & 0 & 0 & x_{yit} & 0 \\ 0 & 0 & 0 & 0 & x_{mit} \\ 0 & 0 & x_{wit} & 0 & 0 \end{pmatrix} \quad \text{and} \quad L_t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

For clarity of presentation we define a few other quantities as follows. The parameter vector β consists of the regression coefficients in (29), (30), and (31), including the parameters from the function J_{it}^W defined in (5), and the parameter vectors from the initial condition equations in (33). The parameter vector γ consists of the coefficients in (35). Note that the covariance matrix for α_i , Γ_i , is constructed from γ and Δ_ρ as defined in (34). Let the vector α contain all the individuals specific random effects, that is, $\alpha' = (\alpha'_1, \dots, \alpha'_N)$. For convenience we use the notation $\Pr(t | \theta_{-t})$ to denote the distribution of t , conditional on all the elements in θ , not including t . Below we explain the sampling of each of the parts in θ (augmented by z^*), conditional on all the other parts and the data.

Sampling the Latent Variables z^ :*

There are three latent dependent variables: y_{it}^* , m_{it}^* , and w_{it}^* . While y_{it}^* and m_{it}^* are never directly observed, w_{it}^* is observed if the i th individual worked in year t . Conditional on θ , the distribution of the latent dependent variables is

$$z_{it}^* | \theta \sim N(\tilde{x}_{it}^*\beta + L_t\alpha_i, \Sigma).$$

From this joint distribution we can infer the conditional univariate distributions of interest, that is $\Pr(y_{it}^* | m_{it}^*, w_{it}^*, \theta)$ and $\Pr(m_{it}^* | y_{it}^*, w_{it}^*, \theta)$, which are truncated univariate normals, with truncation regions that depend on the values of y_{it} and m_{it} , respectively. Note that m_{it} and w_{it} are observed only if $y_{it} = 1$. Therefore, when $y_{it} = 1$ we sample m_{it}^* from the appropriate truncated distribution. In contrast, when $y_{it} = 0$, the distribution of m_{it}^* is not truncated. Similarly, we can infer the distribution of the unobserved (hypothetical) wages, $\Pr(w_{it}^* | y_{it}^*, m_{it}^*, \theta)$.

Sampling the Regression Coefficients β :

It can be easily shown (see Chib and Greenberg (1998) for details) that if the prior distribution of β is given by $\beta \sim N(\tilde{\beta}_0, \tilde{B}_0)$, then the posterior distribution of β , conditional on all other parameters is

$$\begin{aligned}\beta|\theta_{-\beta} &\sim N(\hat{\beta}, B), \quad \text{where} \\ \hat{\beta} &= B \left(\tilde{B}_0^{-1}\beta_0 + \sum_{i=1}^N \sum_{t=1}^T \tilde{x}'_{it} \Sigma^{-1}(z_{it}^* - L_t \alpha_i) \right), \\ B &= \left(\tilde{B}_0^{-1} + \sum_{i=1}^N \sum_{t=1}^T \tilde{x}'_{it} \Sigma^{-1} \tilde{x}_{it} \right)^{-1}, \\ \tilde{\beta} &= (\beta', \delta_1, \delta_2)', \\ x_{it}^* &= (\tilde{x}'_{it}, D_{y1}, D_{m1}),\end{aligned}$$

$D_{y1} = 1$ and $D_{m1} = 1$ if the person is in his/her first year in the sample, and $D_{y1} = 0$ and $D_{m1} = 0$, otherwise.

Sampling the Individuals' Random Effects α_i :

The conditional likelihood of the random effects for individual i is

$$l(\alpha_i) \propto \Sigma^{-T/2} \exp \left\{ -0.5 \sum_{t=1}^T (z_{it}^* - \tilde{x}_{it}\beta - L_t \alpha_i)' \Sigma^{-1} (z_{it}^* - \tilde{x}_{it}\beta - L_t \alpha_i) \right\}.$$

The prior distribution for the random effects is $N(0, \Gamma_i)$, so that the posterior distribution of α_i is

$$\begin{aligned}\alpha_i &\sim N(A_i, V_{\alpha_i}), \quad \text{where} \\ V_{\alpha_i} &= \left(\Gamma_i^{-1} + \sum_{t=1}^T L_t' \Sigma^{-1} L_t \right)^{-1}, \quad \text{and} \\ A_i &= V_{\alpha_i} \sum_{t=1}^T L_t' \Sigma^{-1} (z_t^* - \tilde{x}_{it}\beta).\end{aligned}$$

Sampling the Covariance Matrix Σ :

Recall that the covariance matrix of the idiosyncratic error terms, τ_{it} , is given in (38). Since the conditional distribution of Σ is not a standard, known distribution, it is impossible to sample from it directly. Instead, we sample the elements of Σ using the Metropolis-Hastings (M-H) algorithm (see Chib and Greenberg (1995)). The target distribution here is the conditional posterior of Σ , that is,

$$p(\Sigma|\theta_{-\Sigma}) \propto l(\Sigma|\theta_{-\Sigma}, \alpha_i, z_{it}^*) p(\sigma_\xi^2) p(\rho).$$

The likelihood component is given by

$$l(\Sigma|\theta_{-\Sigma}, \alpha_i, z_{it}^*) = |\Sigma|^{-NT/2} \exp \left\{ \sum_{i=1}^N \sum_{t=1}^T A_{it}' \Sigma^{-1} A_{it} \right\},$$

where, $A_{it} = z_{it}^* - \tilde{x}_{it}\beta - L_t \alpha_i$.

The prior distributions for $\rho = (\rho_{uv}, \rho_{u\xi}, \rho_{v\xi})'$ and σ_ξ^2 are chosen to be the conjugate distributions, truncated over the relevant regions. For ρ we have $p(\rho) = N_{[-1,1]}(0, V_\rho)$, a truncated normal distribution between -1 and 1. For σ_ξ^2 we have $p(\sigma_\xi^2) = N_{(0,\infty)}(\mu_{\sigma_\xi}, V_{\sigma_\xi})$, a left truncated normal distribution truncated at 0. The candidate generating function is chosen to be of the autoregressive form, $q(x', x^*) = x^* + v_i$, where v_i is a random normal disturbance. The tuning parameter for ρ and σ_ξ^2 is the variance of v_i 's.

Sampling Γ_i , Δ_ρ and γ :

Recall that the covariance matrix Γ_i has the form given by

$$\begin{aligned} \Gamma_i &= \text{diag}(\sigma_{i1}, \sigma_{i2}, \sigma_{i3}) \Delta_\rho \text{diag}(\sigma_{i1}, \sigma_{i2}, \sigma_{i3})', \quad \text{where} \\ \sigma_j &= (\exp(\bar{x}'_{ij} \gamma_j))^{1/2} \end{aligned} \tag{59}$$

and Δ_ρ is the correlation matrix given in (34). As in the sampling of Σ , we have to use the M-H algorithm. The sampling mechanism is similar to the sampling of Σ . The only difference is that now we sample elements of γ and Δ_ρ , conditional on each other, and the rest of the elements of θ .

The part of the conditional likelihood that involves Γ_i is

$$l(\Gamma_i | \alpha_i) \propto \prod_{i=1}^n |\Gamma_i|^{-\frac{1}{2}} \exp \left\{ \sum_{i=1}^N \alpha_i \Gamma_i^{-1} \alpha_i' \right\},$$

and the prior distributions of γ and elements of Δ_ρ are taken to be $N(0, V_\gamma)$ and $N_{[-1,1]}(0, V_\delta)$, respectively.

Appendix C—Why do we Need the J^W Function?

For illustration purposes, assume that one considers a simple wage regression in which there are only returns to experience and returns to seniority. Note that new workers entering new jobs gain one year of experience, but they also lose the returns to seniority that they have accumulated in the previous job they have just left. Hence, two workers endowed with similar experience and leaving firms after spells of similar lengths should also receive on average similar wages at the entry level in their new firms. Moreover, if the returns to experience were linear, then two workers with distinct levels of experience who are leaving their respective firms after spells of similar lengths, should also see on average their wage change by similar amounts. That is, interfirm mobility that takes place at different points in one's career (e.g., after one, two, or three jobs) should not matter much for one's wage change in his/her new firm. To see whether or not this is evident in the data, we present in Table C.1 the average (log) wages of individuals who move right before, and immediately after, the move, for all three educational groups. Each row of the table provide these statistics for a given level of experience and seniority at the time of the move. The table clearly shows that wage changes depend crucially on the level of experience and seniority of the worker at the time of a move.

For example, we compare the individuals with the same low level of seniority (less than two years) in line 1, i.e., those with less than 5 years of experience with those in line 3, i.e., those with 5 to 9 year of experience, and with those in line 6, i.e., those with at least 10 years of experience. For the high school dropouts we see huge differences in the average growth rate of 2.2, 3.3, and 4.5, respectively. This seems to indicate that interfirm mobility benefits low skilled workers later in their career. This pattern seems quite similar for the college graduates but very different for the high school graduates. Note that there are generally large differences in the patterns of changes at different levels of seniority-experience combinations.

To check whether or not these raw statistics are indicative of some real phenomena, we regressed the (log) wage at entry level of a job on experience at entry, seniority in the previous job, as well as a set of dummy variable indicators for the number of previous jobs, for the three education groups.³⁵ The results clearly show that seniority in the previous job has a large, positive, and significant association with the wage in the new job for all education groups. Furthermore we find that for the high school graduates and college graduates experience also has positive and significant association with the new wage. For these two education groups the number of previous job changes has a negative and significant effect on the entry level wage. In contrast, these factors seem to have no effect on the group of high school dropouts.

While these results are largely descriptive, they seem to strongly indicate that the entire career path of an individual has to be taken into account. Moreover, they give rise to the particular specification of the J^W function adopted in this paper. Namely, we allow for distinct jumps in wages which are functions of the level of experience and seniority at the time of the jump.

³⁵For brevity, we do not report the results here, but they are available upon request from the authors.

Table C.1: Wages at Time of Job Change

No.	Experience (X)	Seniority (S)	Less than High School			High School Graduate			College Graduate					
			Obs	Wage at $t-1$	Wage at t	Wage Growth	Obs	Wage at $t-1$	Wage at t	Wage Growth	Obs	Wage at $t-1$	Wage at t	Wage Growth
1.	$X < 5$	$S < 2$	80	9.192 (.818)	9.328 (.605)	2.184 (.186)	686	9.332 (.806)	9.561 (.586)	3.127 (10.21)	271	9.757 (1.01)	9.817 (.583)	1.648 (12.10)
2.	$X < 5$	$S \geq 2$	16	9.073 (.749)	9.679 (.352)	7.304 (9.76)	100	9.618 (.533)	9.585 (.542)	-0.233 (4.75)	19	9.335 (1.29)	9.890 (.621)	8.905 (25.28)
3.	$5 \leq X < 10$	$S < 2$	116	9.068 (.773)	9.305 (.568)	3.258 (9.92)	598	9.638 (.605)	9.710 (.598)	0.978 (6.80)	252	9.710 (.685)	9.996 (.549)	3.362 (8.63)
4.	$5 \leq X < 10$	$2 \leq S < 5$	109	9.345 (.601)	9.384 (.676)	0.595 (7.22)	499	9.654 (.591)	9.759 (.532)	1.294 (5.89)	236	9.874 (.643)	10.075 (.505)	2.364 (6.98)
5.	$5 \leq X < 10$	$S \geq 5$	21	9.266 (1.04)	9.488 (.610)	4.581 (22.4)	177	9.782 (.577)	9.782 (.552)	0.170 (5.59)	44	9.795 (.790)	10.017 (.645)	2.625 (7.04)
6.	$X \geq 10$	$S < 2$	302	9.362 (.750)	9.501 (.660)	1.883 (8.31)	805	9.679 (.759)	9.759 (.711)	1.257 (9.93)	311	10.002 (.815)	10.145 (.739)	1.800 (7.81)
7.	$X \geq 10$	$2 \leq S < 5$	243	9.465 (.737)	9.533 (.690)	1.000 (6.94)	724	9.831 (.691)	9.851 (.676)	0.444 (6.78)	364	10.166 (.706)	10.283 (.570)	1.432 (6.53)
8.	$X \geq 10$	$5 \leq S < 10$	178	9.592 (.753)	9.554 (.712)	-0.111 (7.00)	479	9.868 (.665)	9.916 (.610)	0.713 (6.24)	227	10.324 (.685)	10.388 (.557)	0.845 (5.64)
9.	$X \geq 10$	$S \geq 10$	139	9.625 (.674)	9.563 (.772)	-0.495 (6.95)	269	10.031 (.650)	9.976 (.684)	-0.342 (7.61)	122	10.361 (.784)	10.287 (.909)	-0.473 (8.58)

Table 1: Summary Statistics for the PSID Extract for Selected Years, 1975–1992

Variable	Year									
	1975	1978	1980	1982	1984	1986	1988	1990	1992	
Individual and Family Characteristics:										
Observations	3,385	4,019	4,310	4,423	4,451	4,569	4,668	5,728	5,397	
1. Education	11.589 (3.317)	11.716 (3.156)	11.887 (3.037)	12.024 (2.967)	12.155 (2.929)	12.406 (2.964)	12.482 (2.887)	12.100 (3.289)	12.108 (3.308)	
2. Experience	20.808 (14.364)	20.728 (14.478)	20.632 (14.442)	21.071 (14.343)	21.330 (14.148)	21.571 (13.930)	21.921 (13.718)	22.513 (13.266)	24.169 (12.877)	
3. Seniority	5.152 (7.686)	5.526 (7.544)	5.434 (7.377)	5.600 (7.312)	5.993 (7.142)	6.258 (7.231)	6.588 (7.434)	6.699 (7.384)	7.285 (7.563)	
4. Participation	0.942 (0.233)	0.921 (0.270)	0.915 (0.279)	0.897 (0.305)	0.881 (0.324)	0.880 (0.325)	0.871 (0.336)	0.879 (0.327)	0.862 (0.345)	
5. Mobility	0.236 (0.424)	0.141 (0.348)	0.141 (0.348)	0.092 (0.288)	0.087 (0.281)	0.090 (0.286)	0.085 (0.279)	0.083 (0.275)	0.065 (0.247)	
6. Log Wage	9.869 (0.789)	9.899 (0.819)	9.887 (0.814)	9.889 (0.891)	9.902 (0.877)	9.981 (0.810)	9.962 (0.805)	9.918 (0.788)	9.924 (0.842)	
7. Black	0.309 (0.462)	0.327 (0.469)	0.334 (0.472)	0.324 (0.468)	0.320 (0.466)	0.319 (0.466)	0.319 (0.466)	0.263 (0.440)	0.254 (0.435)	
8. Hispanic	0.036 (0.186)	0.035 (0.184)	0.032 (0.177)	0.034 (0.180)	0.034 (0.180)	0.032 (0.176)	0.031 (0.174)	0.069 (0.253)	0.071 (0.257)	
9. Family other income	0.671 (1.926)	0.923 (2.610)	1.219 (3.797)	1.734 (6.315)	2.042 (7.093)	2.312 (6.600)	2.591 (7.239)	2.649 (10.853)	2.960 (13.438)	
10. No. of children	1.350 (1.225)	1.207 (1.157)	1.132 (1.116)	1.076 (1.079)	1.042 (1.043)	1.023 (1.049)	1.002 (1.046)	1.092 (1.189)	1.107 (1.166)	
11. Children 1 to 2	0.216 (0.325)	0.220 (0.348)	0.231 (0.376)	0.229 (0.377)	0.210 (0.364)	0.200 (0.362)	0.188 (0.354)	0.197 (0.401)	0.173 (0.370)	
12. Children 3 to 5	0.225 (0.333)	0.215 (0.360)	0.212 (0.361)	0.204 (0.357)	0.213 (0.368)	0.202 (0.367)	0.198 (0.366)	0.190 (0.387)	0.196 (0.386)	
13. Married	0.854 (0.482)	0.807 (0.492)	0.795 (0.496)	0.775 (0.496)	0.789 (0.497)	0.778 (0.498)	0.765 (0.498)	0.770 (0.496)	0.777 (0.499)	

Table 1: (Continued)

Variable	Year									
	1975	1978	1980	1982	1984	1986	1988	1990	1992	
Cohort Effects:										
14. Age 15 or less in 1975	0 (0)	0.093 (0.290)	0.156 (0.363)	0.204 (0.403)	0.256 (0.436)	0.317 (0.465)	0.368 (0.482)	0.512 (0.500)	0.516 (0.500)	
15. Age 16 to 25 in 1975	0.218 (0.413)	0.261 (0.439)	0.274 (0.446)	0.275 (0.446)	0.269 (0.443)	0.252 (0.434)	0.240 (0.427)	0.186 (0.389)	0.186 (0.390)	
16. Age 26 to 35 in 1975	0.296 (0.457)	0.248 (0.432)	0.220 (0.415)	0.203 (0.402)	0.187 (0.390)	0.173 (0.378)	0.159 (0.366)	0.123 (0.329)	0.124 (0.329)	
17. Age 36 to 45 in 1975	0.173 (0.378)	0.142 (0.349)	0.125 (0.330)	0.114 (0.318)	0.105 (0.307)	0.097 (0.297)	0.090 (0.286)	0.070 (0.255)	0.070 (0.255)	

Table 2: Participation by Education Groups

Variable	High School Dropouts				High School Graduates				College Graduates			
	Mean	Std	95% CI		Mean	Std	95% CI		Mean	Std	95% CI	
			Low	Up			Low	Up			Low	Up
1. Constant	-0.507	0.347	-1.192	0.176	0.185	0.296	-0.411	0.746	-1.245	0.650	-2.514	-0.017
2. Education	-0.017	0.015	-0.047	0.012	0.046	0.014	0.018	0.073	0.130	0.027	0.077	0.184
3. Lagged Experience	-0.029	0.043	-0.111	0.055	-0.089	0.027	-0.143	-0.035	-0.068	0.061	-0.187	0.052
4. Lagged Exp. ² /100	0.180	0.264	-0.346	0.691	0.794	0.191	0.414	1.173	0.807	0.413	-0.022	1.615
5. Lagged Exp. ³ /1,000	-0.033	0.063	-0.154	0.091	-0.251	0.050	-0.350	-0.154	-0.272	0.105	-0.480	-0.064
6. Lagged Exp. ⁴ /10,000	-0.000	0.005	-0.010	0.010	0.022	0.004	0.013	0.030	0.023	0.009	0.006	0.041
7. Lagged Participation	1.723	0.061	1.608	1.846	1.677	0.051	1.581	1.781	1.890	0.104	1.685	2.090
8. Lagged Mobility	0.654	0.132	0.407	0.921	0.420	0.076	0.275	0.569	0.372	0.156	0.071	0.692
Family Variables:												
9. Family other income	0.008	0.006	-0.004	0.021	-0.004	0.003	-0.009	0.002	-0.003	0.002	-0.006	0.001
10. No. of Children	0.070	0.025	0.022	0.118	-0.019	0.022	-0.062	0.026	0.193	0.065	0.067	0.320
11. Children 1 to 2	-0.054	0.065	-0.179	0.075	-0.119	0.048	-0.214	-0.024	-0.158	0.127	-0.403	0.097
12. Children 3 to 5	-0.122	0.064	-0.248	0.004	-0.008	0.047	-0.101	0.083	-0.062	0.142	-0.340	0.215
13. Married	0.184	0.064	0.059	0.308	0.378	0.058	0.264	0.492	0.173	0.120	-0.062	0.406
Location:												
14. Northeast	-0.238	0.084	-0.397	-0.069	0.037	0.053	-0.064	0.143	-0.034	0.088	-0.209	0.141
15. North Central	0.056	0.069	-0.084	0.187	-0.105	0.02	-0.191	-0.024	0.084	0.087	-0.082	0.259
16. South	0.202	0.060	0.089	0.320	0.078	0.041	-0.004	0.158	-0.009	0.081	-0.163	0.151
17. Living in SMSA	-0.074	0.068	-0.212	0.061	-0.070	0.050	-0.168	0.027	-0.032	0.113	-0.250	0.190
18. County unemp. rate	-0.026	0.009	-0.044	-0.009	-0.031	0.007	-0.045	-0.017	-0.000	0.019	-0.038	0.036
Race:												
19. Black	-0.464	0.094	-0.646	-0.278	-0.434	0.076	-0.585	-0.285	-0.224	0.173	-0.550	0.120
20. Hispanic	-0.026	0.128	-0.273	0.228	-0.051	0.146	-0.330	0.238	-0.738	0.313	-1.359	-0.140
Cohort Effects (as of 1975):												
21. Age 15 or less	1.090	0.173	0.768	1.452	0.717	0.160	0.412	1.041	0.866	0.342	0.211	1.527
22. Age 16 to 25	1.059	0.190	0.682	1.428	0.497	0.150	0.202	0.801	0.326	0.359	-0.362	1.041
23. Age 26 to 35	1.130	0.188	0.755	1.490	0.485	0.143	0.204	0.779	0.673	0.323	0.033	1.322
24. Age 36 to 45	0.782	0.151	0.487	1.075	0.461	0.122	0.217	0.701	0.543	0.258	0.048	1.063

Note: Omitted from the table are the coefficients on the year dummy variables.

Table 3: Mobility by Education Groups

Variable	High School Dropouts				High School Graduates				College Graduates			
	Mean	Std	95% CI		Mean	Std	95% CI		Mean	Std	95% CI	
			Low	Up			Low	Up			Low	Up
1. Constant	-1.201	0.273	-1.743	-0.659	-1.293	0.165	-1.621	-0.970	-1.248	0.266	-1.768	-0.727
2. Education	-0.000	0.011	-0.021	0.020	0.001	0.008	-0.014	0.017	-0.006	0.010	-0.027	0.015
3. Lagged Experience	0.002	0.030	-0.057	0.062	-0.012	0.014	-0.040	0.016	-0.014	0.027	-0.065	0.0385
4. Lagged Exp. ² /100	-0.186	0.207	-0.598	0.222	-0.106	0.115	-0.333	0.119	-0.232	0.211	-0.651	0.178
5. Lagged Exp. ³ /1,000	0.061	0.055	-0.047	0.170	0.048	0.035	-0.019	0.117	0.103	0.064	-0.021	0.229
6. Lagged Exp. ⁴ /10,000	-0.006	0.005	-0.016	0.004	-0.005	0.003	-0.012	0.001	-0.011	0.007	-0.024	0.001
7. Lagged Seniority	-0.113	0.023	-0.158	-0.067	-0.161	0.012	-0.185	-0.137	-0.149	0.023	-0.194	-0.105
8. Lagged Sen. ² /100	0.729	0.306	0.093	1.304	1.035	0.155	0.728	1.335	1.133	0.319	0.502	1.763
9. Lagged Sen. ³ /1,000	-0.215	0.138	-0.467	0.074	-0.291	0.064	-0.414	-0.162	-0.393	0.153	-0.692	-0.089
10. Lagged Sen. ⁴ /10,000	0.022	0.020	-0.020	0.057	0.029	0.008	0.013	0.044	0.048	0.023	0.003	0.092
11. Lagged Mobility	-0.679	0.069	-0.817	-0.548	-0.854	0.037	-0.926	-0.781	-1.017	0.062	-1.141	-0.898
Family Variables:												
12. Family other income	-0.034	0.009	-0.052	-0.018	-0.027	0.003	-0.034	-0.021	-0.010	0.002	-0.014	-0.006
13. No. of Children	-0.006	0.015	-0.036	0.023	0.009	0.011	-0.012	0.030	-0.057	0.020	-0.097	-0.019
14. Children 1 to 2	0.063	0.039	-0.016	0.139	0.036	0.021	-0.005	0.078	0.056	0.035	-0.013	0.125
15. Children 3 to 5	-0.017	0.040	-0.095	0.060	0.004	0.022	-0.038	0.047	-0.017	0.039	-0.094	0.059
16. Married	-0.006	0.046	-0.097	0.084	-0.091	0.027	-0.144	-0.038	-0.065	0.041	-0.144	0.016
Location:												
17. Northeast	-0.003	0.044	-0.089	0.084	-0.037	0.019	-0.075	0.001	-0.016	0.026	-0.067	0.035
18. North Central	-0.035	0.036	-0.105	0.035	-0.003	0.017	-0.036	0.031	-0.022	0.025	-0.073	0.027
19. South	0.034	0.030	-0.026	0.093	0.036	0.015	0.006	0.066	0.018	0.024	-0.030	0.064
20. Living in SMSA	-0.130	0.037	-0.202	-0.059	-0.001	0.021	-0.042	0.040	0.053	0.033	-0.011	0.118
21. County unemp. rate	0.002	0.007	-0.011	0.015	0.004	0.004	-0.003	0.011	-0.002	0.007	-0.015	0.011
Race:												
22. Black	-0.027	0.040	-0.104	0.052	-0.005	0.023	-0.049	0.040	0.106	0.046	0.016	0.195
23. Hispanic	0.002	0.070	-0.136	0.137	0.018	0.049	-0.079	0.114	-0.019	0.107	-0.231	0.188
Cohort Effects (as of 1975):												
24. Age 15 or less	0.045	0.125	-0.203	0.287	0.033	0.085	-0.131	0.202	0.033	0.128	-0.215	0.286
25. Age 16 to 25	-0.033	0.116	-0.264	0.188	0.020	0.082	-0.139	0.182	0.040	0.125	-0.201	0.285
26. Age 26 to 35	0.023	0.102	-0.180	0.218	0.035	0.077	-0.114	0.188	0.043	0.114	-0.180	0.270
27. Age 36 to 45	-0.015	0.080	-0.171	0.139	0.089	0.067	-0.040	0.224	-0.002	0.101	-0.197	0.195

Note: Omitted from the table are the coefficients on the year dummy variables and the coefficients on the industry dummy variables.

Table 4: Wage Equation by Education Group

Variable	High School Dropouts					High School Graduates					College Graduates				
	Mean	Std	95% CI			Mean	Std	95% CI			Mean	Std	95% CI		
			Low	Up	Low			Up	Low	Up			Low	Up	
1. Constant	7.662	0.139	7.392	7.933	7.842	0.082	7.683	8.006	8.143	0.137	7.874	8.412			
2. Education	0.025	0.006	0.014	0.036	0.048	0.004	0.041	0.055	0.040	0.005	0.029	0.050			
3. Experience	0.064	0.011	0.042	0.086	0.076	0.005	0.067	0.085	0.112	0.008	0.096	0.127			
4. Exp. ² /100	-0.359	0.068	-0.494	-0.224	-0.467	0.032	-0.530	-0.404	-0.577	0.054	-0.683	-0.472			
5. Exp. ³ /1,000	0.088	0.016	0.056	0.121	0.123	0.009	0.106	0.141	0.134	0.015	0.104	0.163			
6. Exp. ⁴ /10,000	-0.008	0.001	-0.011	-0.006	-0.012	0.001	-0.014	-0.011	-0.013	0.001	-0.016	-0.010			
7. Seniority	0.072	0.006	0.060	0.085	0.069	0.004	0.062	0.076	0.075	0.006	0.063	0.088			
8. Sen. ² /100	-0.301	0.075	-0.447	-0.154	-0.285	0.040	-0.362	-0.207	-0.395	0.079	-0.548	-0.241			
9. Sen. ³ /1,000	0.097	0.030	0.037	0.156	0.077	0.016	0.046	0.108	0.134	0.034	0.067	0.201			
10. Sen. ⁴ /10,000	-0.011	0.004	-0.019	-0.004	-0.007	0.002	-0.011	-0.004	-0.016	0.005	-0.025	-0.006			
Job switch in 1st sample year:															
11. Job change in 1 st year (ϕ_0^s)	-0.006	0.062	-0.125	0.117	0.067	0.027	0.014	0.120	0.161	0.048	0.068	0.259			
12. Lagged Experience (ϕ_0^e)	0.013	0.003	0.007	0.019	0.012	0.002	0.008	0.016	0.008	0.004	0.001	0.015			
Job switches after:															
13. Up to 1 year (ϕ_{10})	0.049	0.023	0.004	0.093	0.109	0.009	0.091	0.128	0.213	0.017	0.179	0.247			
14. 2 to 5 years (ϕ_{20})	0.137	0.028	0.081	0.192	0.107	0.013	0.082	0.133	0.157	0.019	0.119	0.194			
15. 6 to 10 years (ϕ_{30})	0.215	0.076	0.066	0.363	0.190	0.042	0.108	0.272	0.327	0.070	0.189	0.464			
16. Over 10 years (ϕ_{40})	0.057	0.106	-0.149	0.267	0.373	0.054	0.268	0.479	0.517	0.088	0.345	0.687			
Seniority at job that lasted:															
17. 2 to 5 years (ϕ_2^s)	0.026	0.012	0.004	0.049	0.028	0.005	0.018	0.037	0.054	0.007	0.040	0.068			
18. 6 to 10 years (ϕ_3^s)	0.011	0.011	-0.011	0.033	0.018	0.006	0.007	0.029	0.008	0.010	-0.012	0.027			
19. Over 10 years (ϕ_4^s)	0.026	0.005	0.016	0.036	0.035	0.003	0.028	0.041	0.002	0.006	-0.009	0.013			
Experience at job that lasted:															
20. Up to 1 year (ϕ_1^e)	0.003	0.001	0.000	0.006	-0.002	0.001	-0.004	0.000	-0.006	0.002	-0.009	-0.003			
21. 2 to 5 years (ϕ_2^e)	-0.002	0.002	-0.006	0.001	0.0001	0.001	-0.002	0.002	-0.005	0.002	-0.008	-0.002			
22. 6 to 10 years (ϕ_3^e)	0.001	0.002	-0.003	0.006	-0.003	0.002	-0.006	-0.000	-0.001	0.002	-0.006	0.004			
23. Over 10 years (ϕ_4^e)	0.002	0.003	-0.004	0.008	-0.017	0.002	-0.021	-0.012	-0.004	0.003	-0.010	0.002			

Table 4: (Continued)

Variable	High School Dropouts			High School Graduates			College Graduates					
	Mean	Std	95% CI		Mean	Std	95% CI		Mean	Std	95% CI	
			Low	Up			Low	Up			Low	Up
Location:												
24. Northeast	0.023	0.029	-0.034	0.078	0.046	0.012	0.022	0.069	0.050	0.015	0.020	0.079
25. North Central	0.083	0.024	0.035	0.130	-0.006	0.010	-0.026	0.014	-0.049	0.014	-0.077	-0.022
26. South	-0.101	0.020	-0.139	-0.062	-0.044	0.009	-0.062	-0.026	-0.012	0.013	-0.039	0.014
27. Living in SMSA	0.080	0.020	0.040	0.120	0.063	0.009	0.045	0.080	0.040	0.013	0.015	0.064
28. County unemp. rate	-0.002	0.002	-0.007	0.003	-0.005	0.001	-0.007	-0.002	-0.004	0.002	-0.008	0.000
Race:												
29. Black	-0.312	0.032	-0.376	-0.248	-0.264	0.017	-0.297	-0.230	-0.224	0.036	-0.297	-0.152
30. Hispanic	0.057	0.041	-0.024	0.137	-0.082	0.030	-0.140	-0.025	0.017	0.073	-0.125	0.158
Cohort effects (as of 1975):												
31. Age 15 or less	0.539	0.072	0.402	0.686	0.372	0.048	0.279	0.469	-0.010	0.080	-0.168	0.146
32. Age 16 to 25	0.395	0.074	0.253	0.545	0.250	0.048	0.158	0.346	-0.232	0.081	-0.387	-0.073
33. Age 26 to 35	0.372	0.065	0.246	0.502	0.235	0.044	0.151	0.324	-0.152	0.071	-0.291	-0.014
34. Age 36 to 45	0.275	0.060	0.156	0.398	0.255	0.042	0.172	0.337	0.071	0.069	-0.062	0.206

Note: Omitted from the table are the coefficients on the year dummy variables and the coefficients on the industry dummy variables. The coefficients in brackets in lines (11) through (23) are according to the definition of the J^W function defined in the text.

Table 5: Estimated Cumulative and Marginal Returns to Experience

Group	Cumulative Returns				Marginal Returns (in %)			
	Years of Experience				Years of Experience			
	5	10	15	20	5	10	15	20
Panel A: Quartic Model								
HS Dropouts	.241 (.042)	.362 (.062)	.410 (.071)	.420 (.073)	3.442 (.597)	1.550 (.307)	0.482 (.231)	-0.012 (.232)
HS Graduates	.277 (.017)	.402 (.024)	.440 (.028)	.438 (.031)	3.776 (.233)	1.447 (.142)	0.232 (.135)	-0.234 (.134)
College Graduates	.430 (.029)	.661 (.044)	.762 (.052)	.786 (.058)	6.339 (.421)	3.116 (.264)	1.117 (.244)	-0.045 (.244)
Panel B: Quadratic Model								
HS Dropouts	0.101 (.005)	0.246 (.012)	0.472 (.022)	0.678 (.028)	2.661 (.286)	1.959 (.248)	1.256 (.215)	0.554 (.191)
HS Graduates	0.146 (.008)	0.253 (.015)	0.320 (.022)	0.349 (.027)	2.526 (.154)	1.744 (.135)	0.961 (.122)	0.179 (.116)
College Graduates	0.256 (.015)	0.446 (.029)	0.567 (.040)	0.622 (.051)	4.455 (.285)	3.109 (.253)	1.763 (.231)	0.417 (.221)
Panel C: Quartic Model with no J^W Function								
HS Dropouts	0.312 (.040)	0.489 (.058)	0.583 (.065)	0.634 (.067)	4.684 (.557)	2.553 (.281)	1.344 (.217)	0.774 (.218)
HS Graduates	0.370 (.016)	0.572 (.023)	0.674 (.026)	0.724 (.028)	5.462 (.214)	2.853 (.128)	1.397 (.126)	0.713 (.125)
College Graduates	0.588 (0.028)	0.939 (.041)	1.129 (.047)	1.214 (.051)	9.105 (.384)	5.198 (.230)	2.600 (.219)	0.911 (.222)
Panel D: Quadratic Model with no J^W Function								
HS Dropouts	0.211 (.013)	0.383 (.025)	0.515 (.036)	0.607 (.044)	3.830 (.253)	3.034 (.221)	2.238 (.193)	1.442 (.174)
HS Graduates	0.237 (.007)	0.423 (.014)	0.558 (.020)	0.643 (.025)	4.231 (.140)	3.214 (.125)	2.197 (.114)	1.181 (.109)
College Graduates	0.401 (.013)	0.710 (.024)	0.928 (.035)	1.053 (.044)	7.105 (.244)	5.263 (.220)	3.420 (.203)	1.578 (.197)

Table 6: Estimated Cumulative and Marginal Returns to Seniority

Group	Cumulative Returns				Marginal Returns (in %)			
	Years of Seniority				Years of Seniority			
	2	5	10	15	2	5	10	15
Panel A: Quartic Model								
HS Dropouts	.133 (.010)	.297 (.019)	.507 (.024)	.677 (.029)	6.128 (.408)	4.882 (.236)	3.669 (.249)	3.242 (.238)
HS Graduates	.127 (.006)	.283 (.011)	.475 (.014)	.616 (.018)	5.850 (.227)	4.590 (.141)	3.215 (.150)	2.555 (.150)
College Graduates	.136 (.010)	.294 (.019)	.477 (.026)	.615 (.034)	6.109 (.412)	4.513 (.257)	3.035 (.279)	2.630 (.272)
Panel B: Quadratic Model								
HS Dropouts	0.151 (.015)	0.266 (.029)	0.347 (.040)	0.392 (.050)	4.966 (.253)	4.721 (.216)	4.314 (.169)	3.906 (.156)
HS Graduates	0.094 (.003)	0.228 (.008)	0.435 (.013)	0.619 (.018)	4.612 (.153)	4.347 (.132)	3.906 (.108)	3.464 (.105)
College Graduates	0.097 (.006)	0.236 (.014)	0.449 (.025)	0.637 (.034)	4.775 (.281)	4.485 (.247)	4.001 (.209)	3.517 (.205)
Panel C: Quartic Model with no J^W Function								
HS Dropouts	0.099 (.010)	0.204 (.018)	0.304 (.021)	0.370 (.022)	4.296 (.385)	2.774 (.195)	1.479 (.209)	1.302 (.200)
HS Graduates	0.092 (.006)	0.184 (.010)	0.256 (.012)	0.280 (.013)	3.918 (.214)	2.303 (.110)	0.778 (.122)	0.369 (.126)
College Graduates	0.097 (.010)	0.177 (.017)	0.212 (.020)	0.216 (.021)	3.838 (.381)	1.675 (.179)	0.095 (.196)	0.308 (.195)
Panel D: Quadratic Model with no J^W Function								
HS Dropouts	0.054 (.005)	0.131 (.011)	0.251 (.018)	0.359 (.022)	2.653 (.216)	2.510 (.177)	2.273 (.125)	2.035 (.114)
HS Graduates	0.042 (.003)	0.102 (.006)	0.192 (.010)	0.268 (.013)	2.070 (.121)	1.915 (.099)	1.656 (.074)	1.397 (.076)
College Graduates	0.027 (.004)	0.067 (.010)	0.133 (.016)	0.198 (.020)	1.342 (.201)	1.331 (.162)	1.313 (.119)	1.294 (.130)

Table 7: Comparison of Alternative Estimates of Cumulative Returns to Seniority

		Topel replication sample 1968-1983							
Tenure		2 years	5 years	10 years	15 years				
OLS		0.092 (0.013)	0.188 (0.022)	0.273 (0.025)	0.315 (0.026)				
IV1 (Altonji & Williams)		0.053 (0.010)	0.097 (0.019)	0.119 (0.028)	0.123 (0.037)				
BFKT sample 1975-1992									
		High School Dropouts			College Graduates				
Tenure		2 years	5 years	10 years	15 years	2 years	5 years	10 years	15 years
OLS		0.099 (0.009)	0.197 (0.015)	0.273 (0.016)	0.300 (0.017)	0.068 (0.007)	0.136 (0.012)	0.189 (0.013)	0.208 (0.015)
IV1 (Altonji & Williams)		0.062 (0.010)	0.112 (0.017)	0.131 (0.019)	0.131 (0.018)	0.043 (0.008)	0.078 (0.014)	0.092 (0.015)	0.090 (0.016)
Quadratic model, no J^W function		0.054 (0.005)	0.131 (0.011)	0.251 (0.018)	0.359 (0.022)	0.027 (0.004)	0.067 (0.010)	0.133 (0.016)	0.198 (0.020)
Quartic model, no J^W function		0.099 (0.010)	0.204 (0.018)	0.304 (0.021)	0.370 (0.022)	0.097 (0.010)	0.177 (0.017)	0.212 (0.020)	0.216 (0.021)
Quadratic model with J^W function		0.151 (0.015)	0.266 (0.029)	0.347 (0.040)	0.392 (0.050)	0.097 (0.006)	0.236 (0.014)	0.449 (0.025)	0.637 (0.035)
Quartic model with J^W function		0.133 (0.010)	0.297 (0.019)	0.507 (0.024)	0.677 (0.029)	0.136 (0.010)	0.294 (0.019)	0.477 (0.026)	0.615 (0.034)

Note: The first panel estimates are taken from AW, Table 2. The OLS and IV (based on Altonji and Williams methodology) estimates for high school dropouts and college graduates from our sample extract. The remaining rows come from Table 6.

Table 8: Estimates of the Stochastic Elements by Education Groups

Variable	High School Dropouts				High School Graduates				College Graduates			
	Mean	St.dev.	Range		Mean	St.dev.	Range		Mean	St.dev.	Range	
			Min	Max			Min	Max			Min	Max
Covariance Matrix of White Noises (elements of Σ)												
1. ρ_{uw}	-0.0012	0.0098	-0.0201	0.0180	-0.0019	0.0072	-0.0145	0.0134	-0.0021	0.0103	-0.0222	0.0178
2. $\rho_{u\xi}$	-0.0251	0.0069	-0.0390	-0.0124	-0.0323	0.0060	-0.0431	-0.0218	-0.0299	0.0135	-0.0538	-0.0032
3. $\rho_{v\xi}$	0.0100	0.0071	-0.0040	0.0249	0.0064	0.0049	-0.0042	0.0150	0.0210	0.0094	0.0027	0.0386
4. σ_{ξ}^2	0.2954	0.0038	0.2882	0.3029	0.2086	0.0016	0.2054	0.2117	0.2048	0.0023	0.2004	0.2093
Correlations of Individual Specific Effects (elements of Δ_{ρ})												
5. $\rho_{\alpha_y \alpha_m}$	-0.1835	0.1762	-0.5327	0.1487	-0.5024	0.1033	-0.6203	-0.2394	-0.5521	0.0746	-0.6723	-0.4115
6. $\rho_{\alpha_y \alpha_w}$	0.3572	0.0368	0.2827	0.4235	0.3574	0.0262	0.3017	0.4070	0.2007	0.0508	0.1002	0.3061
7. $\rho_{\alpha_m \alpha_w}$	-0.1846	0.3343	-0.8439	0.5321	-0.6980	0.1589	-0.9089	-0.4161	-0.7682	0.0739	-0.8729	-0.6083

Figure 1: High School Graduates—Wage Growth Due to Experience and Seniority

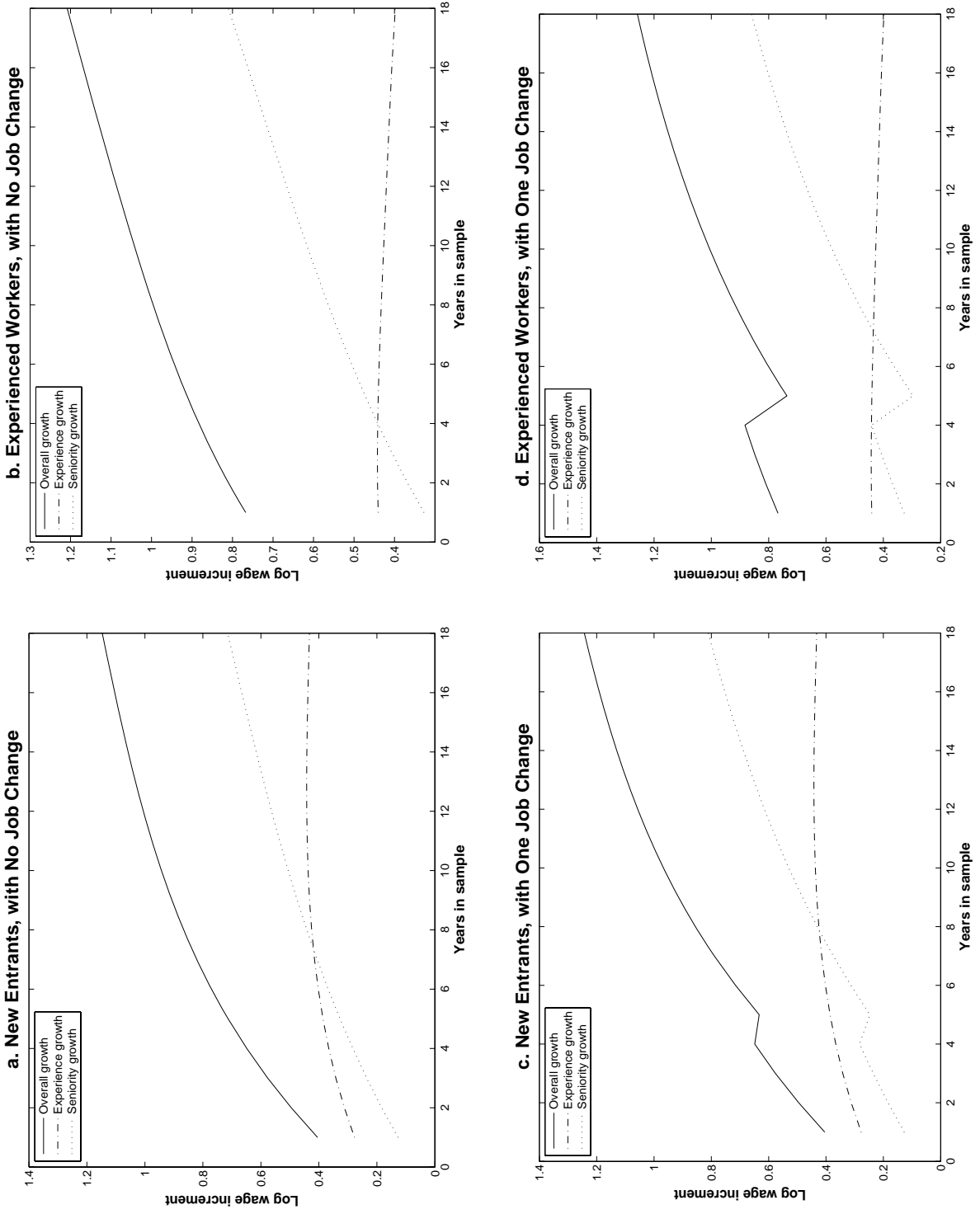


Figure 2: College Graduates—Wage Growth Due to Experience and Seniority

