A Theory of Wage Dynamics with Assignment and Pareto Learning

Jin Li*
Kellogg School of Management
Northwestern University
July 18, 2008

Abstract

This paper develops a tractable model of assignment and Pareto learning that provides both qualitative and quantitative implication on wage dynamics. It sheds light on a puzzle concerning an important feature of wage dynamics: serial correlations of wage changes in a few years apart are found to be zero in large representative datasets but are found to be positive in subpopulations of the workforce.

In addition, the model is consistent with three basic facts on wage dynamics. First, the expected wage change increases with experience but the rate of wage increase declines with it. Second, more educated workers have higher rates of wage growth. Third, the percentage of workers with a wage gain in any cohort decreases with the cohort experience but is independent of the schooling levels.

JEL Classifications: C78, D83, J31
Keywords: Assignment, Pareto Learning, Wage Distribution, Wage Dynamics

*This paper is based on Chapter 2 of my MIT dissertation. I thank Robert Gibbons, Bengt Holmstrom and Daron Acemoglu for invaluable discussions, encouragements, and support. I also thank Gadi Barlevy, Leemore Dafny, Yuk-Fai Fong, Eric French, Ed Green, Ben Jones, Fabian Lange, Dale Mortensen, Derek Neal, Marco Ottaviani, Marek Pycia, Jose Tessada, Marko Tervio, Mike Waldman and the seminar participants at Chicago Fed, Chicago GSB, M.I.T., Kellogg Northwestern, and Penn State for helpful comments and discussions. All remaining errors are mine.
1 Introduction

In this paper, we develop a simple model of wage dynamics based on two assumptions: a) the abilities of workers are learnt publicly over time and b) there are assignments between firms and workers. Both learning and assignment have been recognized as important factors in determining wages.\(^1\) The purpose of this paper is to explore the extent to which key features of wage dynamics can be explained by the combination of these two factors.

Concerning the relationship among experience, schooling, and wage levels, Rubinstein and Weiss (2006) summarize three key patterns for white U.S. males. First, average wages of workers increase with experience, but average rates of wage growth decline with it. This fact that the life-cycle earning profile is increasing and concave in experience is perhaps the most basic and robust fact about wage dynamics; see for example Mincer (1974), Murphy and Welch (1992) and Heckman, Lochner and Todd (2001).

Second, the average rate of wage growth increases with years of education. Although the magnitudes vary by data source, the pattern holds robustly in all three data source considered by Rubinstein and Weiss (2006): CPS-ORG, PSID, and NLSY. In fact, the differences in wage growth rates by educational attainment are sizeable. For example, using the CPS-ORG data, the authors find that workers with an advanced degree experience annual wage growth of 7.7% in the first ten years of their labor market experience, as compared to 3.9% for high school dropouts.

Third, the proportion of wage “gainers,” those with a wage increase between consecutive wage observations, decreases with labor market experience. However, this proportion is independent of the schooling level of the worker. CPS-ORG data shows that the proportion of wage gainers is 64.3% for college graduates with 0-10 years of experience, 60.2% for 11-15 years, 56.7% for 16-25 years, and 53.6% for 26-40 years. Corresponding numbers are similar across education groups and within the PSID and NLSY.

Rubinstein and Weiss (2006) also discuss two leading theories that help explain these facts: human capital and search theory. According to human capital theory,

workers invest in their human capital in earlier stages of their career and the amount of investment decreases with worker’s age. This gives a natural explanation for why the wage profile is increasing and concave. To explain why the rate of wage growth increases with schooling, one can introduce worker heterogeneity into the human capital models: more able workers will both acquire more schooling and invest more on-the-job training and thus experience faster wage growth. Finally, if one adds a stochastic shock in wage determination, then it is possible to generate that the proportion of wage gainers decreases with age and is independent of schooling.\(^2\)

According to search theory, workers receive random offers and leave the current firm when better outside offers are received. In standard search models, this implies that the worker’s wage is increasing over time. In addition, the wage level is concave in experience because it is harder to find better outside offers the higher the current wage is. To explain why the rate of wage growth increases with schooling, one can assume that the wage offer distribution is more dispersed for workers with more schooling, so that it is easier to find better outside matches. Finally, adding stochastic shocks in search models may help explain the relationship between the proportion of wage gainers, schooling, and experience levels.

Despite the tremendous success of human capital and search theory, there are important aspects of wage dynamics left to be explained. For example, on human capital theory Rubinstein and Weiss note that

“Although the investment interpretation is consistent with important features of the data on wage levels, it cannot explain some important features of wage changes.\(^3\) In particular, it was noted by MaCurdy (1982) and Abowd and Card (1989) that, after accounting for the common wage growth, the growth rates of individual wages are not correlated for periods that are more than a few years apart.”\(^4\)

Serial correlation of changes in wage residuals is also found to be zero by several other studies using large representative datasets including Topel (1991), Topel and Ward (1992), Lillard and Reville (1999), and Meghir and Pistaferri (2004).

\(^2\)The key assumption to explain this fact is that the rate of human capital accumulation decreases with experience at a faster rate than that of the shocks.

\(^3\)Italics in original.

\(^4\)The correlations of wage change for adjacent periods are typically found to be negative. This is consistent with the search theory but is also often attributed to measurement errors.
The zero serial correlation of changes in wage residuals is a puzzle to both human capital and search theory. Human capital theory predicts a positive serial correlation between early and late wage growth due to the heterogeneity in the rate of human capital accumulation. On the other hand, search theory predicts a negative correlation because workers who experience faster wage growth in the past have higher current wages, so they are less likely to find higher future wages. Perhaps a combination of human capital theory and search theory will generate zero correlations, but this will require strong assumptions on the relative importance of the two theories in explaining wage formation for the entire life-cycle of workers because the positive serial correlations from the human capital models needs to exactly cancel out with the negative ones from search models for all years.

Adding to the puzzle, several studies on subpopulations of the workforce including Baker, Gibbs, Holmstrom (1994) (BGH 94a hereafter) using data on managerial workers, Lillard and Weiss (1979) using data on American scientists, and Hause (1980) using data on young Swedish males, have all found positive serial correlations of changes in wage residuals despite the zero serial correlations found in the large representative datasets. As Gibbons (1996) notes,

"This welter of findings deserves further attention. One possible explanation is that only certain small groups of workers (such as the managerial and professional workers in Baker-Gibbs-Holmstrom and Lillard-Weiss) exhibit such a personal effect. If most groups of workers do not exhibit this effect then the representative cross-sections in Abowd-Card, Topel, and Topel-Ward would not either."

What Gibbons leaves open is the why some group of workers have a personal effect (positive serial correlations) and others don’t.

The learning and assignment model developed in this paper helps explain why positive serial correlation is likely to be found in some group of workers, but not in the representative workforce. In addition, the model fits the basic facts on the relationship among experience, schooling, and wage levels as described above. One key advantage of this model is its tractability: we introduce an attractive learning process called Pareto learning to obtain several explicit formulas about wage changes. The formulas provide both qualitative and quantitative predictions on wage dynamics.
Our numerical calculations suggest that the effects from learning and assignment on wage growth can be substantial.

In this model, there is a continuum of workers and a continuum of firms. Workers differ by their unknown abilities. Firms differ by their known technologies. Production takes place when a worker is matched with a firm. The production is stochastic, and the expected output is complementary in the worker’s ability and the firm’s technology. The output of the worker is observed publicly, so firms learn about the worker’s ability over time.

Learning takes place in the form of Pareto learning, as follows. The ability of a worker is drawn from a Pareto distribution. The exact value of the worker’s ability is unknown, but it can be learned from observing a sequence of signals, which are drawn uniformly between zero and the worker’s ability. Pareto learning implies that, based on the signals observed, the posterior distribution of the worker’s ability is also Pareto. Moreover, the posterior distribution is completely determined by the number of signals and the maximum value of the signals observed.

Pareto learning permits us to obtain explicit formulas for the rate of expected wage growth for workers of all experience levels. The formulas imply that, first, the expected wage growth is positive but declines with experience. The expected wage grows with experience because new information allows for better matching between firms and workers, so the expected output (and thus expected wage) increases over time. The expected rate of wage growth decreases with experience because the marginal value of information and of better assignment decreases over time.

More interestingly, the formula predicts a life-cycle earning profile similar to that of Mincerian regressions: in both cases, log wage is a function of experience. However, the Mincerian regression specifies that log wage is a quadratic function of experience and it is known that this specification underestimates the wage growth of workers in the earlier stage of their career. In the Pareto model, in contrast, the effect of learning and assignment is concentrated in the early stages and tapers off quickly. For example, our numerical calculations show that the wages grow by 5.5% to 18.5% in the first period and drops to 1.3% to 3% in the third period. This suggests that learning and assignment can help alleviate the problem that wage growth is underestimated by the Mincerian regression.
Second, the formulas imply that the expected wage growth is larger when assignment is more important, characterized by the firm’s technology distribution shifting to the right. When assignment is more important, the gain in productivity through better matching becomes bigger and this results in larger wage growth. If we believe assignment is more important for better-educated workers because they face a wider set of choices of jobs and the productivities on these jobs are more sensitive to ability, then this model implies that the rate of wage growth increases with schooling.

Third, the formulas imply that the percentage of wage gainers (those who experience positive wage increase) decreases with experience and the percentage is independent of the firm’s technology distribution. This is because wage increases with experience, and in Pareto learning, workers become less likely to experience increase in expected ability over time. Moreover, since a worker’s wage increases if and only if his expected ability increases, the distribution of firm’s technology is irrelevant for determining the proportion of wage gainers.

In addition to providing formulas that help explain basic patterns of wage levels, the model also give rise to explicit formulas that shed light on the patterns of wage changes. The formulas imply that the serial correlation of wage changes is a) positive, b) increases with the importance of assignment, and c) decreases with experience. The positive serial correlation results from the interaction of learning and assignment. Fast wage growth in the past is either because the worker is of high ability (ability effect) or the worker is lucky and has had good draws (luck effect). These two effects have opposite implications on future wage growth. In a pure learning model without assignment, these two effects exactly cancel each other out so that the future wage growth is independent of past wage growth and thus the serial correlation is zero.\(^5\)

With assignment, higher-ability workers receive higher wages not only because they are more productive on each job but also because they are assigned to more productive jobs.\(^6\) In other words, assignment can magnify the ability effect relative to the luck effect and thus generate positive serial correlation. The key insight here is that if the importance of assignment increases with wage level\(^7\), then workers with...

---

\(^5\) Mathematically, this follows because the wage of a worker is a martingale: see Baker, Gibbs, and Holmstrom (1994b) and Farber and Gibbons (1996) for excellent discussions.

\(^6\) Mathematically, this means that wage is a convex function of expected ability (which is a martingale), so that wage is a submartingale.

\(^7\) Mathematically, this means that wage is increasingly convex in expected ability.
higher wages will benefit more from the positive effect of assignment and experience larger wage gains in the future. Since workers with fast past wage growth are more likely to have higher current wage, this generates positive serial correlation of wage changes. Moreover, since the channel of wage changes operates through learning and assignment, it is easy to see why serial correlation is larger when learning is more important and when assignment is important.

We also compute the numerical values of the serial correlations using the same parameters for computing wage levels. We find that the magnitudes for serial correlations are typically very small: for most parameters, the serial correlations are less than or equal to 0.05. This helps to explain why large, representative datasets often fail to find positive serial correlations. On the other hand, when assignment is important and the workers have few years of experience, the value of the correlation can be as high as 0.16. This helps explain, for example, why BGH(1994a), Hause (1980), and Lillard and Weiss (1979) found positive serial correlations in managerial workers, young workers, and professional workers.

While it seems intuitive that assignment is more important for some subpopulation, it maybe difficult to empirically measure the importance of assignment. (Footnote on Hubbard).

The rest of the paper is organized as follows. We set up the model and analyze the equilibrium in Section 2. Section 3 applies the equilibrium outcome to study patterns of wage dynamics. Section 4 discusses the related literature. Section 5 concludes and discusses extensions of the model.

2 Model

In this section, we present a symmetric learning model with matching. Subsection 2.1 describes the model ingredients, including the workers, the firms, and the production technology. Subsection 2.2 introduces the equilibrium concept. Subsection 2.3 discusses the Pareto learning process. Subsection 2.4 solves for the unique equilibrium of the model and gives explicit formulas for the equilibrium matching and wage function.
2.1 Workers and Firms

The economy has an infinite number of periods. There is a unit mass of workers. In each period, a measure \(1 - \rho\) of new workers enter the economy and each of the existing workers exits with probability \(1 - \rho\). Workers who have been in the economy for \(n\) periods are in cohort \(n\), and in particular, new workers are in cohort 0. We examine an economy where the size distribution of workers is stationary: the measure of workers in cohort \(n\) is \((1 - \rho)\rho^n\).

Workers differ in their unknown abilities, which are denoted by \(a\). We assume that the ability of each worker is drawn independently from a Pareto distribution with parameter \((1, \alpha)\), where the first parameter determines the lower bound of the support of the distribution and the second parameter determines the rate the density of the distribution decreases:

\[
\begin{align*}
\Pr(A \geq a) &= a^{-\alpha} \quad \text{for } a > 1; \\
\Pr(A \geq a) &= 1 \quad \text{for } a \leq 1. 
\end{align*}
\]

In Pareto distribution, the proportion of workers with ability greater than a threshold decreases at a constant rate with the threshold level. We also assume \(\alpha > 1\), which guarantees that the ability distribution has a finite mean.

The utility of each worker is separably additive across periods with a discount factor \(\delta \in (0, 1)\). In each period, if a worker is hired by a firm that offers \(w\), his period utility is \(w\). If the worker is not hired by any firm, he receives his outside option, which gives a utility of 1,\(^9\) i.e.,

\[
U(w_0, w_1, ...) = \sum_{n=0}^{\infty} \delta^n (1_{\{\text{hired}\}} w_n + (1 - 1_{\{\text{hired}\}})),
\]

where \(1_{\{\text{hired}\}}\) is an indicator function that is equal to 1 if the worker is hired and 0 otherwise. Workers in this model are risk neutral and have no disutility of effort. We abstract away from issues in insurance and incentive here to focus on the effect of learning and assignment. These issues are no doubt important in wage determination and we hope to explore them in future work.

\(^8\)We use the upper-case to represent a random variable and the lower-case to represent its realization.

\(^9\)We normalize the outside option of the workers to be 1 for simplicity. When the outside option isn’t 1, the same analysis applies, but the wage formula will differ by a constant.
There is unit mass of infinitesimal firms. Each firm lives forever and hires at most one worker per period. The firms differ by their observable technological levels, which we denote as $s$. The distribution of $s$ across firms is a generalized Pareto distribution with parameter $(1, \gamma, 1 - \sigma^\gamma)$:

\[
F(s) = 1 - \left( \frac{s}{\sigma} \right)^{-\gamma} \quad \text{for } s > 1;
\]

\[
F(s) = 0 \quad \text{for } s < 1,
\]

where $F$ denotes the CDF of the technology distribution. In other words, this distribution is a linear combination of a point mass at 1 (of size $1 - \sigma^\gamma$) with a Pareto distribution with parameter $(1, \gamma)$.\(^{10}\) We assume $\gamma > 1$ for the distribution to have a finite mean.

Production takes place in a given period when one firm is matched with one worker. The output $y$ is linear in the firm’s technology and depends stochastically on the worker’s ability. If a firm with technology $s$ hires a worker of ability $a$, the output is given by

\[
Y(s, a, Z) = 2sZ,
\]

where $Z$ is a random variable drawn uniformly from $[0, a]$, and 2 is a normalizing term. The formula implies that for a firm-worker pair with technology $s$ and ability $a$, the expected output equals

\[
E[Y(s, a, Z)] = \int_0^a \frac{2sz}{a} \, dz = sa.
\]

Therefore, technology and ability are complements in the production.

Moreover, for a given firm, the expected output is linear in worker’s ability. This implies that the expected output is completely determined by the expected ability of the worker (and not on higher moments of the posterior ability distribution). We use $\Gamma$ to denote the expected ability of a worker and $G$ to denote the CDF of the expected ability distribution of the entire economy.

Learning takes place symmetrically in this model. Since the abilities of workers are unknown, firms form their beliefs about the abilities by observing the past history

\(^{10}\)The assumption of a point mass at the bottom helps simplify the expressions for the matching formula and wage formula. This assumption does not affect qualitatively any of our results on wage dynamics.
of outputs of the workers and the technologies they use. This implies that for any worker in cohort $n$, all firms know of $\{(y_m, s_m)\}_{m=0}^{n-1}$, where $y_m$ and $s_m$ are the worker’s output and the technology he used when he was in cohort $m$. Consequently, firms know $\{z_m\}_{m=0}^{n-1}$, where $z_m = \frac{y_m}{2s_m}$ is the stochastic factor in the production. This implies that the speed of learning does not depend on which firm the worker is employed with.\footnote{For an interesting paper where the speed of learning will depend on which firm the worker is employed with, see for example Felli and Harris (1996).}

Finally, each firm maximizes its expected payoff, which equals the expected output minus the wage cost if it hires the worker. If the firm does not hire a worker, its payoff is zero. More formally, the expected payoff of a firm with technology $s$ that hires a worker of ability $a$ with wage $w$ is

$$\Pi(s, w, a) = E[Y(s, a, Z)] - w = sa - w.$$ 

### 2.2 Equilibrium Concept

In this model, an allocation involves a) wage function that assigns wages to workers and b) a matching function that assigns workers to firms. An allocation is an equilibrium if

1. There exists a function $W(\eta)$ that maps expected abilities to wages, so that workers of expected ability $\eta$ are paid wage $W(\eta)$.

2. There exists a matching function $\eta^*(s)$ that maps technologies to expected abilities such that for each firm with technology $s$,

$$\eta^*(s)s - W(\eta^*(s)) \geq \eta s - W(\eta) \quad \text{for all } \eta.$$ 

3. The expected ability distribution of the economy satisfies

$$G(\eta) = \int_0^\infty 1_{\{\eta^*(s)\leq \eta\}} dF(s),$$

where recall that $G(\eta)$ is the CDF of the expected ability distribution and $F(s)$ is the CDF of the technology distribution.

The first condition of the equilibrium states that workers of the same expected ability are paid the same wage. This condition reflects the fact that the expected
output is linear in worker’s ability (equation (4)), so each worker’s value to a firm depends solely on his expected ability. The second condition is the usual profit maximization condition of firms in static assignment models. We abstract away from long-term contracts and focus on this spot market equilibrium because the workers are risk-neutral and the speed of learning of the worker’s ability is the same across all firms. The third condition states that the demand for workers equals the supply for every level of expected ability. It is written in an integral form because almost surely the measure of workers at any expected ability level is zero, so we require instead that the demand for workers below any expected ability level equals the labor supply.

2.3 Pareto Learning

In this subsection, we describe the Pareto learning process (of ability) and list its properties in the context of this model. First, we give a formula of a worker’s expected ability given his observed history. Next, we give explicit formulas of the expected ability distributions of any cohort and of the entire economy. These formulas will help us solve the wage formula and calculate the wage distributions. Finally, we provide a formula of how expected ability evolves over time. This formula will be used extensively in our discussions of the wage dynamics. For interested readers, we derive the formulas in appendix.

2.3.1 Basic Properties of Pareto Learning

Pareto learning (of ability) is a sequential learning process where a) the prior distribution of ability $A$ is Pareto distributed, and b) in each period $t$, a new signal (of ability) $Z_t$ is drawn uniformly from $[0, a]$ and is observed. By observing the signals, the firms can calculate the posterior distribution of ability, which takes a very simple form.

In particular, recall in this model that the prior distribution of ability is $Pareto(1, \alpha)$, so $\Pr(A = x) = \alpha x^{-\alpha - 1} = \frac{\alpha}{x} \left(\frac{x}{1}\right)^{-\alpha}$ for $x \geq 1$. Now assume the period 0 signal is $z_0$. Then for $a_1 \geq z_0$, 

\[\text{Pr}(A = x) = \alpha x^{-\alpha - 1} = \frac{\alpha}{x} \left(\frac{x}{1}\right)^{-\alpha} \text{ for } x \geq 1.\]
Pr(A = a_1 | z_0) = \frac{Pr(z_0 | A = a_1) Pr(A = a_1)}{\int_1^\infty Pr(z_0 | A = x) Pr(A = x) dx} = \frac{(\alpha + 1)}{a_1} \left( \frac{a_1}{\max\{z_0, 1\}} \right)^{-(\alpha + 1)}.

This implies that the posterior distribution of ability is \textit{Pareto}(\max\{z_0, 1\}, (\alpha + 1)). Applying the same calculation \(n\) times, we can show that the posterior distribution of ability of a worker in cohort \(n\) is \textit{Pareto}(m, \alpha + n), where \(m(z_0, .., z_{n-1}) = \max\{z_0, .., z_{n-1}, 1\}\) can be thought of as the maximal draw of ability. More formally, for a worker in cohort \(n\) with maximal draw \(m\),

\[
Pr(A \geq a_n | m, n) = \left( \frac{a_n}{m} \right)^{-(\alpha + n)} \quad \text{for } a_n \geq m; \\
= 1 \quad \text{for } a_n < m. \tag{5}
\]

Pareto learning also leads to a simple expression for the expected ability, which determines the wage of the worker. The expected ability of a worker in cohort \(n\) with maximum draw \(m\), denoted as \(\eta_n(m)\), is given by

\[
\eta_n(m) = \frac{\alpha + n}{\alpha + n - 1} m. \tag{6}
\]

This indicates that the expected ability of a worker is completely determined by a) his maximal draw of ability and b) how long he has been in the workforce.

The term \(\frac{\alpha + n}{\alpha + n - 1} > 1\) decreases with \(n\) and reflects the option value of worker’s ability. In other words, if young and old workers have the same maximal draw of ability, then the young worker has higher expected ability.

### 2.3.2 Distributions of Expected Ability

To help calculate the wage formula (as a function of expected ability), we give in this subsection formulas for the distributions of the expected abilities for each cohort of workers. The distribution of expected ability of workers in cohort \(n\) is a generalized
Pareto distribution with parameters \( \left( \frac{a+n}{a+n-1}, \frac{a}{a+n} \right) \), i.e.

\[
\begin{align*}
\Pr(\Gamma_n \geq \eta_n) &= \frac{n}{a+n} (\frac{a+n-1}{a+n})^{\frac{a+n}{a+n-1}} \eta_n^{-\alpha} & \text{for } \eta_n \geq \frac{a+n}{a+n-1}; \\
\Pr(\Gamma_n \geq \eta_n) &= 1 & \text{for } \eta_n < \frac{a+n}{a+n-1}.
\end{align*}
\]  

(7)

Similar to the ability distribution, the upper tail of the expected ability distribution is Pareto with parameter \( \alpha \). Different from the ability distribution, the expected ability distribution has a mass of \( \frac{a}{a+n} \) at the bottom, which locates at \( \frac{a+n}{a+n-1} \). This mass of expected ability corresponds to workers whose maximum draw is less than or equal to 1. Note as \( n \) goes to infinity, the size of this bottom mass goes to 0 and the location of the mass goes to 1, so the expected ability distribution converges to the ability distribution.

While the size and location of the bottom mass differs for each expected ability distribution, the upper tail of this rate \( \alpha \) is independent of cohort age \( n \). This makes it easy to aggregate the expected ability distribution across all cohorts. Let \( G \) denotes the expected ability distribution of the entire economy. Then for an expected ability level \( \eta > \frac{\alpha}{\alpha-1} \), \( G \) satisfies

\[
1 - G(\eta) = \left( \frac{\eta}{\lambda} \right)^{-\alpha},
\]

where \( \lambda = \left[ \sum_{\rho=0}^{\infty} (1 - \rho) \rho^n \frac{n}{a+n} (\frac{a+n-1}{a+n})^{-\alpha} \right]^{\frac{1}{\beta}} \). In other words, the upper-part of the expected ability distribution is like a Pareto with power parameter \( \alpha \).

### 2.3.3 Dynamics of Expected Ability

Finally, we give a formula for the dynamics of expected ability, which is the keystone in analyzing wage dynamics. For a worker in cohort \( n \) with expected ability \( \eta_n \), his conditional distribution of expected ability in \( t \) periods in the future is a generalized Pareto distribution with parameter \( \left( \frac{(a+n-1)(a+n+t)}{(a+n)(a+n+t-1)} \eta_n, \alpha + n, \frac{a+n}{a+n+t} \right) \). More formally,

\[
\begin{align*}
\Pr(\Gamma_{n+t} > \eta_{n+t}|\Gamma_n = \eta_n) &= \frac{t}{\alpha+n+t} \left( \frac{(a+n-1)(a+n+t)}{(a+n)(a+n+t-1)} \eta_n \right)^{\alpha+n} \eta_{n+t} \quad \text{for } \eta_{n+t} \geq \frac{(a+n-1)(a+n+t)}{(a+n)(a+n+t-1)} \eta_n, \\
\Pr(\Gamma_{n+t} > \eta_{n+t}|\Gamma_n = \eta_n) &= \frac{1}{\alpha+n+t} \quad \text{for } \eta_{n+t} = \frac{(a+n-1)(a+n+t)}{(a+n)(a+n+t-1)} \eta_n.
\end{align*}
\]

(9)

Just like the expected ability distribution, the bottom mass of the conditional
distribution represents the sample path where no new maximum draw has occurred in the $t$ periods. This happens with probability \( \frac{a+n}{a+n+t} \), which increases with the cohort age and decreases with $t$. When this happens, the expected ability of the worker is adjusted downwards to \( \frac{(a+n-1)(a+n+t)}{(a+n)(a+n+t-1)} \eta_n \). The proportion of downward adjustment is bounded by \( \frac{1}{a+n} \) (corresponding to no new maximum draw in all future periods), which goes to zero as $n$ increases to infinity. This suggests that the Pareto learning process is downward rigid.

### 2.4 Solving the Equilibrium

In this subsection, we prove the existence and uniqueness of the equilibrium and show that this equilibrium is efficient. In addition, we give explicit formulas of the equilibrium matching function and wage function, and we analyze their properties. To simplify the exposition, we first introduce the following definition.

**Definition 1:** Let $\eta(1+) = \lambda \sigma^{-\frac{n}{2}}$ be the unique expected ability level such that $G(\eta(1+)) = F(1)$.

As will be seen in Theorem 1, the equilibrium assignment is positive assortative, so any worker with expected ability $\eta > \eta(1+)$ will be assigned to firms with technology $s > 1$.

**Theorem 1:** There exists a unique equilibrium with matching function $\eta^*$ and wage function $W$ such that

\[
\eta^*(s) = G^{-1}(F(s)) \quad \text{for } s > 1;
\]
\[
W(\eta) = \eta \quad \text{for } \eta \leq \eta(1+);
\]
\[
W(\eta^*(s)) = s\eta^*(s) - \int_1^s \eta^*(x)dx \quad \text{for } \eta^*(s) > \eta(1+). \tag{10}
\]

The market equilibrium is efficient in the sense that it maximizes the total expected outputs of the economy given the available information.

**Proof.** See Appendix. ■

The problem of characterizing the market equilibrium in this model is similar to a mechanism design problem. To see this, we rewrite the firm’s profit maximization
condition (ii) in the equilibrium definition) as:

$$s\eta^*(s) - W(\eta^*(s)) \geq s\eta^*(s') - W(\eta^*(s')) \quad \text{for all } s, s'.$$

In this representation, we can think of $\eta^*(s)$ as the assignment rule and $W(\eta)$ as the transfer rule in a mechanism. If a firm of technology $s$ announces type $s'$, it receives a payoff of $s\eta^*(s') - W(\eta^*(s'))$ by being assigned a worker of expected ability $\eta^*(s')$ and paying a transfer of $W(\eta^*(s'))$. The condition above is then exactly the condition that each firm finds it optimal to truthfully announce its type. This suggests that we can use the standard technique in mechanism design to solve for this problem. However, because the technology distribution $F$ has an atom in the bottom so that firms with technology $s = 1$ will be matched with workers of different expected abilities, we need to solve for the equilibrium separately for the atom before applying the standard mechanism design technique.

The statement (and proof) of Theorem 1 is true for arbitrary $F$ and $G$ (subject to the atoms of $F$ and $G$ at the bottom of the distribution). With Pareto learning, we know the functional form of $F$ and $G$, and this allows us to give explicit formulas for the equilibrium. To simplify the formulas, we assume that $\sigma^{-\gamma} < (\frac{\lambda(\alpha-1)}{\alpha})^{-\alpha}$, so all the atoms in the expected ability distribution $G$ will be matched with firms of technology $s = 1$ in equilibrium. We also define $w(s)$ to be the equilibrium wage as a function of technology.

Finally, to describe the importance of assignment, we define $k = 1 + \frac{\alpha}{\gamma}$. For a fixed $\alpha$, $k$ becomes larger when $\gamma$ is smaller, corresponding to a right-shift of the technology distribution. Therefore, a bigger $k$ means that assignment is more important.

**Corollary 1:** The equilibrium satisfies the following conditions:

(i): The equilibrium matching function is equal to

$$\eta^*(s) = \lambda(\frac{S}{\sigma})^{\frac{2}{\gamma}} \quad \text{for } s > 1. \quad (11)$$

(ii): The equilibrium wage as a function of expected ability is equal to

$$W(\eta) = \eta \quad \text{for } \eta < \eta(1_+);$$

$$W(\eta) = \frac{\sigma \lambda}{k} (\frac{\eta}{\lambda})^k + \frac{\alpha}{\alpha + \gamma} \eta(1_+) \quad \text{for } \eta \geq \eta(1_+). \quad (12)$$
(iii): The equilibrium wage as a function of technology is equal to

\[ w(s) = \frac{\lambda}{k} \sigma^{-\frac{2}{\alpha}} (s_0^{\frac{2}{\alpha} + 1} + 1), \quad \text{for } s > 1^{13}. \]

Statement (i) in Corollary 1 shows that the equilibrium assignment is positive assortative: workers of higher expected abilities are matched with firms of better technologies. This is because expected ability and technology are complements in production so firms with better technology benefits more from a high ability worker. The matching formula makes it clear that, first, as the ability distribution shifts to the right (corresponding to an decrease in \( a \)), firms of any technology level \( s > 1 \) will be matched with workers of higher expected ability. On the other hand, as the technology distribution shifts to the right (corresponding to an decrease in \( \gamma \)), firms of a given technology level \( s > 1 \) will be matched with workers of lower expected ability because there are more firms with technology above \( s \).

Statement (ii) in Corollary 1 shows that the equilibrium wage (as a function of expected ability) can be rewritten as \( W(\eta) = A\eta^k + B \), where \( A \) and \( B \) are constants that are independent of \( \eta \). This implies that wage is roughly exponential in expected ability. This exponential property, combined with the fact that expected abilities (of any cohort of workers and of the entire economy) have generalized Pareto distributions (see equation (7) and (8)), implies that the right tail of the wage distribution (of any cohort and of the entire economy) is almost Pareto.\(^{14}\)

The wage formula also implies that wage is convex in expected ability (because \( k \) is strictly greater than 1). The source of wage convexity results from positive assortative matching. Higher expected ability leads to matching with better technologies, which make workers even more productive and thus creates a convex wage profile. This convexity suggests that the wage distribution is skewed to the right of ability distribution, and a small difference in ability at the top of distribution can lead to a large difference in wages, which seems to conform to casual observations; see for example Rosen (1981) and Waldman (1984) for more discussions. Moreover, the wage formula implies that the wage difference becomes larger when assignment is more important (corresponding to an decrease in \( \gamma \)), so the model also implies that skill-

\(^{13}\)Firms with \( s = 1 \) are matched with workers of expected abilities ranging from 1 to \( \eta(1_+) \).
\(^{14}\)For a more precise statement and more discussions on wage distribution, see Li (2007).
biased technology change, measured as a right-shift of technology distribution, can increase wage inequality.

Statement (iii) in Corollary 1 gives the equilibrium wage as a function of technology level. From the formula, we see that the equilibrium wage increases with and is convex in technology. If we think of better technology corresponds to higher ranks within a firm, this fits the observations that wage is often convex in job ranks inside a firm, see for example Treble et al (2001).

3 Wage Dynamics

In this section, we apply the Pareto learning model to study wage dynamics. Because workers differ in this model only in their experience level and their unobservable abilities, the predictions of the model are about wage changes after controlling for observables such as education, race, and sex. In subsection 3.1, we explore the relationship among experiences, schooling, and wage levels. In subsection 3.2, we focus on patterns of wage changes that distinguish this model from other models of wage formation.

3.1 Wage Levels

In this subsection, we show that the model is consistent with important features of wage dynamics on wage levels. We derive an explicit formula that gives both qualitative and quantitative implications on how wage changes with experience over the life-cycle of a worker.

**Proposition 1** For a worker in cohort $n$ with expected ability $\eta > \frac{\alpha + n}{\alpha + n - 1} \eta(1_+)$, the expected wage level after $t$ periods satisfies

$$E[W(\Gamma_{n+t}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+)|\Gamma_n = \eta] = B(t, n, k)(W(\eta) - \frac{\alpha}{\alpha + \gamma} \eta(1_+)),$$

(14)

where $B(t, n, k) = (\frac{\alpha + n}{\alpha + n + t})(\frac{\alpha + n - k + t}{\alpha + n - k}) \left(\frac{(\alpha + n - 1)(\alpha + n + t)}{(\alpha + n)(\alpha + n + t - 1)}\right)^k > 1$. This implies that

---

15The model can be applied separately for each observationally equivalent group of workers. This allows for different wage growth pattern for each subgroup.
(i): Wage increases with experience:

\[
\frac{dE[W(\Gamma_{n+t})|\Gamma_n = \eta]}{dt} > 0.
\]

(ii): The rate of wage growth decreases with experience:

\[
d(E\left[\frac{W(\Gamma_{n+t}) - W(\Gamma_n)}{W(\Gamma_n)}|\Gamma_n = \eta\right])/dn < 0.
\]

(iii): Wage growth is larger when assignment is more important:

\[
d(E\left[\frac{W(\Gamma_{n+t}) - W(\Gamma_n)}{W(\Gamma_n)}|\Gamma_n = \eta\right])/dk > 0.\]

**Proof.** See Appendix.

Recall that \(\eta(1_+)\) is an upper bound of the expected ability of workers who are assigned to firms of the lowest technology level \((s = 1)\). The condition \(\eta > \frac{\alpha+n}{\alpha+n-1}\eta(1_+)\) guarantees that the worker’s expected ability is sufficiently high so that he will not be assigned to firms of the lowest technology level in all future periods. Since we assume \(\eta > \frac{\alpha+n}{\alpha+n-1}\eta(1_+)\), the model describes the wage dynamics of workers in the right tail of the wage distribution.

Proposition 1 shows that the expected wage increase of a worker is proportional to the current wage minus a small constant.\(^{17}\) When the current wage is sufficiently high, the expected wage increase becomes almost proportional to the current wage. Therefore, the empirical content of this model is similar to the Mincerian regression in the sense that log-wage is a function of experience in both cases. Different the Mincerian regression, this model does not restrict the effect of experience on log-wage to be quadratic.

Table I calculates the numerical value of wage growth for some parameters. The numerical results suggest that learning and assignment can explain a substantial proportion of wage growth, especially for young workers. For example, when assignment is important and the abilities of workers are diverse \((\gamma = 2, \alpha = 2)\), learning and assignment contributes to a wage growth of 18.5% in the first period and 28.0% in

\(^{16}\)Here, as in subsequent derivates with respect to \(k\), we hold \(\alpha\) constant and assume that the change of \(k\) results from changes in \(\gamma\).

\(^{17}\)Note that the constant is a fraction of the wage paid to workers with expected ability \(\eta(1_+)\).
the first three periods. The numerical results also show that the rate of wage growth drops quickly. This suggests that the model can complement the human capital model to better explain patterns of wage growth because the Mincerian regression underestimates the wage growth of workers in the earlier stage of their career.

[Insert Table I]

In addition to enabling numerical calculations of wage growth, Proposition 1 gives three qualitative implications on the life-cycle pattern of wage dynamics. All three of the implications have natural economic interpretations. The first implication shows that the expected wage of a worker increases with experience. Since no human capital accumulation occurs in this model, wage increases solely because of learning and assignment. Note that under learning alone, the expected ability is a martingale, so its expected change in expected ability is zero. When there is also assignment, the equilibrium wage is a convex function of the expected ability, so Jensen’s Inequality implies that wage is a submartingale and thus the expected change in expected wage increases is positive. In economic terms, extra information acquired through learning about the worker’s ability enable the workers to be better assigned to firms over time. Therefore, the average outputs (not the expected ability) of the workers and thus their average wages increase over time.

The second implication of Proposition 1 is that the rate of wage growth decreases with experience. This is because the marginal value of information decreases over time in two ways. First, earlier signals have larger impacts on the expected ability of workers and lead to larger changes to wages than later ones. Second, marginal return from better assignment becomes smaller over time because the existing assignment has become better.

The third implication of Proposition 1 is that the rate of wage growth is larger when assignment is more important (corresponding to a smaller \( \gamma \)). This follows because when assignment is more important, the increase in productivity (and thus wage) from better assignment is also larger. Since workers in this model differ only in their experiences and unobservable abilities, we can apply the model separately to workers with different observables. If we believe assignment is more important for better-educated workers because they face a wider set of choices of jobs and the productivities on these jobs are more sensitive to ability, then this model implies that
the rate of wage growth increases with schooling, which fits the second basic fact of
dynamics of wage levels in Rubinstein and Weiss (2006).

The next proposition helps explain the third fact of dynamics on wage levels.

**Proposition 2** For a worker in cohort $n$ with expected ability $\eta > \frac{n+n}{\alpha+n-1} \eta(1_+)$, the
probability that the worker is a wage gainer next period decreases with $n$ and is inde-
dependent of $k$:

\[
\begin{align*}
\frac{d}{dn} \left[ \Pr(W(\Gamma_{n+1}) > W(\Gamma_n)|\Gamma_n = \eta) \right] &< 0, \\
\frac{d}{dk} \left[ \Pr(W(\Gamma_{n+1}) > W(\Gamma_n)|\Gamma_n = \eta) \right] & = 0
\end{align*}
\]

**Proof.** See Appendix. ■

The first part of Proposition 2 indicates that the proportion of wage gainers (those
who experience positive wage changes) decreases with experience. The intuition is
that, for the wage of a worker to increase, his next period draw of productivity must
be higher than the current maximum. Since the maximum weakly increases with
experience, it is harder for workers with more experience to have a new draw that
exceeds the maximum, and therefore the proportion of wage gainers decreases. Note
that this result relies heavily on the property of Pareto learning and is not true in
general learning models. In a normal learning model (with or without matching), for
example, the proportion of wage gainers is always one half.

The second part of Proposition 2 shows that the proportion of wage gainers is
independent of the importance of assignment. This result is true in a general learn-
ing model with assortative matching because wage increases whenever the expected
ability increases. And the change in the expected ability of the worker is independent
of the matching.

Rubinstein and Weiss (2006) finds that, among college graduates, the percent-
age wage gainers is 64.3% for workers with less than 10 years of experience, 60.2%
for workers with experience between 10-15 years, 56.7% for workers with experience
between 16-25 years, and 53.6% for workers with experience between 26-40 years.
The corresponding number of high school graduates are 60.2% for less than 10 years,
58.8% for 10-15 years, 56.2% for 16-25 years and 54.6% for 26-40 years.
Therefore, Proposition 2 fits the qualitative pattern of the data. However, it does not fit the pattern of wage changes well quantitatively. In fact, the proportion of workers who experience real wage gain in cohort $n$ is only roughly $1/n$ for large $n$, so this model underestimates the proportion of wage gainers for large $n$. One way to better fit the data quantitatively is to allow for exogenous human capital accumulation. We discuss how the model can integrate human capital accumulation in Section 5.

### 3.2 Wage Changes

This subsection explores the implication of the model on features of wage dynamics concerning wage changes. We focus on results that distinguish this model from human capital and search models.\(^{18}\) The main result, given in Proposition 3, relates serial correlation of wage changes with experience and assignment.

**Proposition 3** For a worker in cohort $n$ with expected ability $\eta > \frac{\alpha+n-1}{\alpha+n} \eta(1_+)$, the correlation of wage changes satisfies

$$
Corr(W(\Gamma_{n+1}) - W(\Gamma_{n+1}), W(\Gamma_{n+1}) - W(\Gamma_n)|\Gamma_n = \eta) = \frac{1}{\sqrt{\left(\frac{r(t+1,n)-1}{r(1,n)-1}\right)^2 + 1}}
$$

where $r(t,n) = \frac{(\alpha+n+t)(1+\frac{1}{\alpha+n-2k})}{(\alpha+n)(1+\frac{1}{\alpha+n-k})^2}$. In particular,

(i): The correlation decreases with experience:

$$\frac{\partial \text{Corr}}{\partial n} < 0.$$

(ii): The correlation is larger when assignment is more important:

$$\frac{\partial \text{Corr}}{\partial k} > 0.$$

**Proof.** See Appendix. \(\blacksquare\)

\(^{18}\)For example, this model can also show that variances of individual wage changes decrease with experience. Since this result can also be obtained by human capital and search models, we omit it here.
The formula in Proposition 3 makes it clear that the serial correlation of wage changes is positive. This results from the interaction of learning and assignment. In a pure learning model, it is well known that wage is a martingale, i.e., the expected wage change is zero conditional on the current wage; see for example (BGH 94b) and Farber and Gibbons (1996) for excellent discussions. It follows that, without assignment, the serial correlation of wage changes is zero.

Intuitively, there are two effects in determining the wage change of a worker in a learning model. Consider a worker with fast past wage growth. This maybe because a) this worker is of high ability (ability effect) or b) this worker has had lucky draws (luck effect). The ability effect implies the wage of this worker will continue to grow because the current wage is a reflection of the true ability of the worker and the prior average worker ability. The luck effect implies the current wage reflects an overestimate of ability so regression-to-the-mean implies negative (or slow) wage growth in the future. In a pure learning model, these two effects exactly cancel out with each other, and the expected wage change is zero.

In a learning model with assignment, BGH 94b mentions that wage is a (strict) submartingale: i.e., the expected wage change is positive conditional on the current wage.\footnote{Mathematically, this is because assignment makes wage a convex function of expected ability. Convex function on a martingale is a submartingale.} Intuitively, this is because learning improves assignment, and this leads to higher average productivities and higher wages over time. Once wage becomes a submartingale, serial correlation is no longer zero in general, but its sign is in general uncertain. As BGH 94b notes, “[O]nce a contingent decision is introduced into the pure learning model, wages will be serially correlated, but it seems impossible to say anything general about the direction of dependence.”
Serial correlation is positive in this model. And this result is more general. Recall that the expected wage change is zero (conditional on current wage) in a pure learning model. Expected wage changes become positive (for all wage levels) when there is assignment. The key insight is that if assignment is more important for more productive jobs (which pay higher wages), then the expected wage gains are larger for workers with higher wages. Since workers with fast wage growth in the past have higher wages, the condition above implies that they are also more likely to experience larger future wage gains. For interested readers, sufficient conditions for positive serial correlations in general learning process with assignment can be found in Li (2007).  

To explore the empirical implications of Proposition 3, we first calculate the values of serial correlations using the same parameters that predict wage growth. We report the values in Table I, and it can be seen that more than two thirds of the values has a serial correlation less than or equal to 0.05. Moreover, a larger proportion of the serial correlations will be less than 0.05 if we include workers with more years of experience. This sheds light on why research using large representative data sets, such as MaCurdy (1982), Abowd and Card (1989), Topel (1991), Topel and Ward (1992), Lillard and Reville (1999), and Meghir and Pistaferri (2004) typically find small but statistically insignificant positive serial correlations.

Second, Proposition 3 proves that the size of serial correlation decreases with cohort age. This is because the positive serial correlation is due to the common component of ability that is reflected in the wage changes. As the cohort age increases, less amount of information about ability is reflected in the wage change, so the serial correlation becomes less positive. This helps explain, for example, why Hause (1980) found positive serial correlations in young Swedish males.

Third, Proposition 3 also shows that serial correlation increases when assignment is more important (smaller $\gamma$). This is intuitive because assignment magnifies the personal effect and gives rise to the positive serial correlation. In particular, when there is no assignment, the personal effect always equals the regression to the mean effect and the serial correlation is zero. Table 1 shows that a decrease of $\gamma$ from 3 to 2 increases the size of the serial correlation by 50% in general. Interestingly, when

\footnote{For example, if a) the wage is increasingly convex in expected ability (positive third derivative) and b) dispersion of the conditional distribution of expected ability is nondecreasing (in the sense of Second Order Stochastic Dominance) with expected ability, then this property is satisfied. The second condition is satisfied in a normal learning model.}
\( \alpha = \gamma = 2, n = 1, t = 3 \) the serial correlation is 0.16, which is very close to the value (0.17) found in BGH (1994). If assignment is more important for managerial and professional workers, this model gives an explanation for why a positive serial correlation is more likely to be found in these workers.

The empirical implications above help distinguish this model from standard search and human capital theories. Search theory predicts that serial correlations are negative because workers with fast wage growth in the past have found good jobs and are thus less likely to find higher wages. This is clearly inconsistent with the data. The implication from human capital model on serial correlations is more nuanced. Human capital model with identical workers predicts zero serial correlations and is inconsistent with the positive serial correlations found in certain subpopulations. When heterogeneity in the speed of human capital accumulation is introduced, this implies positive and persistent serial correlations and is thus inconsistent with that serial correlations are close to zero for a few years apart, as noted by Rubenstein and Weiss (2006).

It might be possible, however, that heterogeneity in human capital accumulation is important only in some subpopulations but not in the representative workforce. Then human capital theory may explain why serial correlations are close to zero for the representative workforce but are positive for some subpopulations. This explanation, however, require strong assumptions on the distribution of heterogeneities. In particular, since heterogeneities are used in human capital models to explain that the speed of wage growth differs with schooling, it thus requires that the heterogeneities being important across school levels, unimportant within schools levels for the representative workforce, and again important for some workforce! Moreover, even if heterogeneities introduced in this way can fit the data, they do not help identify characteristics of the subpopulations in which positive serial correlations arise.

A better way to empirically distinguish this model from human capital models is to directly examine how changes in the importance of assignment affect serial correlations. However, it can be difficult to measure the importance of assignment, so instead we provide joint implications on the change of assignments. In particular, it can be shown that as assignment becomes more important, this model generates a) an increase in serial correlation (Proposition 3), b) an increase in the average speed
of wage increase (Proposition 1), and c) an increase in inequality.\footnote{See Li (2007) for details.} In recent years, it has been argued that assignment has become more important,\footnote{See for example, Costrell and Loury (2003), Gabaix and Landier (2008), and Tervio (2008).} and it is well-known that inequality has increased. It will be interesting to examine whether serial correlations and the average speed of wage growth has also increased.

For the final empirical implication, we derive an explicit formula of regressing future wage changes on past wage changes.

**Corollary 2:** For a worker in cohort $n$ with expected ability $\eta > \frac{\alpha + n - 1}{\alpha + n} \eta(1_+)$, the regression coefficient of future wage change $(W(\Gamma_{n+t}) - W(\Gamma_n))$ on past wage change $(W(\Gamma_n) - W(\Gamma_{n-1}))$ satisfies

$$\frac{\text{Cov}(W(\Gamma_{n+t}) - W(\Gamma_n), W(\Gamma_n) - W(\Gamma_{n-1})|\Gamma_{n-1} = \eta)}{\text{Var}(W(\Gamma_n) - W(\Gamma_{n-1})|\Gamma_{n-1} = \eta)} = B(t, n, k) - 1. \quad (16)$$

In particular,

(i): The regression coefficient increases with time horizon of wage changes.

(ii): The regression coefficient increases with the importance of assignment.

(iii): The regression coefficient decreases with the cohort’s experience.

**Proof.** Contained in the proof of Proposition 3. $\blacksquare$

As in Proposition 3, the formula above implies that the regression coefficient is larger if a) assignment is more important and b) the worker has fewer years of experience. The intuitions for these two results are identical to the corresponding results in Proposition 3.

The more interesting implication of Corollary 2 is on how serial correlation changes when the wage difference comes from a longer horizon. Note that Proposition 3 does not have a result on how the serial correlation changes with the time difference. This because there is no general result. On the one hand, a bigger $t$ means that more information about the worker’s ability will be captured in the future wage change so that the common component between $W(\Gamma_{n+t+1}) - W(\Gamma_{n+1})$ and $W(\Gamma_{n+1}) - W(\Gamma_n)$ becomes larger, and this increases the correlation. On the other hand, a bigger $t$ implies that more noises are brought into $(W(\Gamma_{n+t+1}) - W(\Gamma_{n+1}))$, (so its variance increases with $t$), and this reduces the correlation. In general, it is unclear which
of the two effects dominates. Since the regression coefficient captures the common component of ability: the regression coefficient increases with the time difference because more information about ability is captured when the time difference is larger.

4 Conclusion

This paper develops a simple model of learning and assignment to explain a number of basic facts of wage dynamics. On the patterns of wage changes, we show that, first, wage increases with experience and the rate of wage growth decreases with experience. Second, more educated workers have higher rates of wage growth. Third, the percentage of workers with a wage gain in any cohort decreases with the cohort experience but this percentage is independent of the schooling levels.

The model also sheds light on the puzzle of serial correlation of changes in wage residuals. This model predicts that serial correlations are positive but small. On the other hand, the serial correlations increase with the importance of job assignment and decrease with experience. This offers a potential explanation for why positive serial correlations have only been found in a subpopulation of workforce such as young, professional, and managerial workers and not in large, representative data sets.

In addition to studying wage dynamics, the model can be used to study the wage distribution. In fact, the model predicts that the right tail of the income distribution is very similar to Pareto, which fits the real distribution reasonably well. Moreover, the model gives an explicit formula for the power parameter of this Pareto distribution, and this enables us to study how wage inequality is affected by policies and events that affect the underlying ability and technology distributions. For example, if the technology distribution shifts to the right, possibly due to a skill-biased technology change, then this model implies that the wage distribution will become more unequal, both in terms of variance and in terms of 90/10 ratio. More interestingly, this model predicts that such technological change will also increase the rate of wage growth, especially for young workers.

In this model, we focus on the role of learning and assignment in wage determination and intentionally leave out many elements that are important for wage formation. As such there are clear limitations to this model. For example, the expected wage of

\[23\text{See Li (2007) for details.}\]
a worker, if he stays with the same job, does not change over time. In other words, there is no wage growth within jobs in this model. In addition, the wage distribution derived in the model fits the upper part of the wage distribution but not quite the lower parts. Our assessment of this model is that it can complement the existing models such as human capital and search to better capture the wage formation process of some subpopulation of the workforce, especially those for whom assignment is important (and inequality is high). Given this role, the model’s tractability becomes more attractive because it makes it easier to add other elements into the model.

One way to capture wage growth within jobs is to incorporate human capital accumulation into the model. Pareto learning makes this extension remains tractable. In particular, if all workers accumulate their human capital at a constant rate, then we again have explicit formula of the right tails of the wage distributions of each cohort and of the economy: they are again Pareto. The explicit formulas can be used to calibrate the model to match the empirical wage distributions across cohorts across years. The calibration can help illustrate the relative importance of human capital compared to learning and assignment in the formation of wages. In addition, it can help understand the extent the skill-biased technological change affects wage inequality, especially in the upper tail of the wage distribution.

Finally, we can extend the model to allow firms to have multiple jobs. In doing so, we can study wage dynamics both between firms and within firms. Similar logic in this paper is likely to explain several empirical regularities on promotion dynamics. For example, the (almost) downward rigid property of Pareto learning may help explain why demotions are rare. In addition, the logic of the positive serial correlation of changes in wage residuals can help explain why promotion is positively serially correlated, or why there is a fast track phenomenon.

References


\(^{24}\) In this extension, each job is modeled as an interval of technologies.

\(^{25}\) See Gibbons (1997) and Gibbons and Waldman (1999b) for reviews of wage dynamics within firms.


5 Appendix

5.1 Properties of Pareto Learning

We take as given that, for a worker in cohort \( n \) with maximum draw \( m \),

\[
\Pr(A \geq a_n|m, n) = \left(\frac{a_n}{m}\right)^{-(\alpha+n)} \quad \text{for } a_n \geq m; \\
= 1 \quad \text{for } a_n < m.
\]

(A): Formula of expected ability

The expected ability of a worker in cohort \( n \) with maximum draw \( m \), denoted as \( \eta_n(m) \), satisfies

\[
\eta_n(m) = \int_m^\infty x \Pr(A = x|m, n) dx \\
= \int_m^\infty x (\alpha + n)m^{\alpha+n} \frac{x^{\alpha+n+1}}{\alpha+n+1} dx \\
= \frac{\alpha + n}{\alpha + n - 1} m.
\]

(B): Formula of expected ability distribution

To calculate the expected ability distribution, first note that for workers in cohort \( n \), the proportion of workers with maximum draw less than \( m \geq 1 \) satisfies

\[
\Pr(M_n \leq m) = \int_1^\infty \min\{1, \left(\frac{m}{x}\right)^n\} \frac{\alpha}{x^{\alpha+1}} dx \\
= 1 - \frac{n}{\alpha + n} m^{-\alpha},
\]

where \( \min\{1, \left(\frac{m}{x}\right)^n\} \) is the probability that the maximum draw of a worker with ability \( x \) is less than \( m \). Therefore, the distribution of maximum draw in cohort \( n \) is a generalized Pareto distribution with parameter \((1, \alpha, \frac{\alpha}{\alpha+n})\). Since equation (6)) links the maximum draw with the expected ability, we have that the distribution of expected ability of workers in cohort \( n \) satisfies

\[
\Pr(\Gamma_n \geq \eta_n) = \frac{n}{\alpha+n} \left(\frac{\alpha+n-1}{\alpha+n} \eta_n\right)^{-\alpha} \quad \text{for } \eta_n \geq \frac{\alpha+n}{\alpha+n-1}; \\
\Pr(\Gamma_n \geq \eta_n) = 1 \quad \text{for } \eta_n < \frac{\alpha+n}{\alpha+n-1}.
\]
In other words, the distribution of the expected ability is a generalized Pareto distribution with parameters \( \left( \frac{\alpha + n}{\alpha + n - 1}, \alpha, \frac{\alpha}{\alpha + n} \right) \).

(C): Formula of the conditional expected ability distribution

For a worker in cohort \( n \) with expected ability \( \eta_n \), we calculate the conditional distribution of his expected ability in \( n + t \). This calculation is facilitated by the one-to-one relationship between expected ability and maximum draw in equation (6). For a worker with expected ability \( \eta_n \), equation (6) implies that his maximal draw \( m_n = \frac{\alpha + n - 1}{\alpha + n} \eta_n \). Pareto learning (equation (5)) implies that the worker’s conditional ability distribution (given expected ability \( \eta_n \)) is therefore Pareto \( (m_n, \alpha + n) \). It follows that the conditional distribution of the maximum draw in \( n + t \) satisfies, for \( m_{n+t} \geq m_n \),

\[
\Pr(M_{n+t} \leq m_{n+t} | M_n = m_n) = \int_{m_n}^{\infty} \max \{ 1, \left( \frac{m_{n+t}}{x} \right)^t \} \frac{(\alpha + n)m_{\alpha + n}}{x^{\alpha + n + 1}} dx
\]

\[
= 1 - \frac{t}{\alpha + n + t} \left( \frac{m_n}{m_{n+t}} \right)^{\alpha + n},
\]

where \( \max \{ 1, \left( \frac{m_{n+t}}{x} \right)^t \} \) is the probability that, for a worker with ability \( x \), his new draws in the \( t \) periods are all smaller than or equal to \( m_{n+t} \). Therefore, the conditional distribution of \( M_{n+t} \) given \( M_n = m_n \) is generalized Pareto \( (m_n, \alpha + n, \frac{\alpha + n}{\alpha + n + t}) \).

The conditional distribution of the maximum draw \( M_{n+t} \) implies that for \( \eta_{n+t} \geq \frac{(\alpha + n - 1)(\alpha + n + t)}{(\alpha + n)(\alpha + n + t - 1)} \eta_n \),

\[
\Pr(\Gamma_{n+t} > \eta_{n+t} | \Gamma_n = \eta_n) = \Pr(M_{n+t} > \frac{\alpha + n + t - 1}{\alpha + n + t} \eta_{n+t} | M_n = \frac{\alpha + n - 1}{\alpha + n} \eta_n)
\]

\[
= \frac{t}{\alpha + n + t} \left( \frac{(\alpha + n - 1)(\alpha + n + t)}{(\alpha + n)(\alpha + n + t - 1)} \eta_n \right)^{\alpha + n}.
\]

And \( \Pr(\Gamma_{n+t} = \frac{(\alpha + n - 1)(\alpha + n + t)}{(\alpha + n)(\alpha + n + t - 1)} \eta_n | \Gamma_n = \eta_n) = \frac{\alpha + n}{\alpha + n + t} \). Therefore, for a worker in cohort \( n \) with expected ability \( \eta_n \), his conditional distribution of expected ability in the \( t \) periods in the future is a generalized Pareto distribution with parameter \( \left( \frac{(\alpha + n - 1)(\alpha + n + t)}{(\alpha + n)(\alpha + n + t - 1)} \eta_n, \alpha + n, \frac{\alpha + n}{\alpha + n + t} \right) \).
5.2 Proofs of Results

**Theorem 1:** There exists a unique equilibrium with matching function \( \eta^* \) and wage function \( W \) such that

\[
\eta^*(s) = G^{-1}(F(s)); \quad \text{for } s > 1;
\]
\[
W(\eta) = \eta \quad \text{for } \eta \leq \eta(1_+);
\]
\[
W(\eta^*(s)) = s\eta^*(s) - \int_1^s \eta^*(x)dx \quad \text{for } \eta^*(s) > \eta(1_+).
\]

The market equilibrium is efficient in the sense that it maximizes the total expected outputs of the economy.

**Proof.** We first show that, in any equilibrium, firms of higher technology must be matched with workers of higher expected ability. If not, then there exist two firms (firm 1 and 2) with technology \( s_1 > s_2 \) and two workers (worker 1 and 2) with expected ability \( \eta_1 < \eta_2 \) such that firm 1 hires worker 1 and firm 2 hires worker 2. In this case, the sum of equilibrium payoff of firm 1 and 2 is

\[
s_1\eta_1 - W(\eta_1) + s_2\eta_2 - W(\eta_2) < s_1\eta_2 - W(\eta_2) + s_2\eta_1 - W(\eta_1),
\]

where the second expression is the sum of payoff if firm 1 hires worker 2 and firm 2 hires worker 1. This implies that either firm 1 or firm 2 can find a profitable deviation, which leads to a contradiction. Therefore, the equilibrium allocation must be positive assortative, i.e.

\[
G(\eta^*(s)) = F(s).
\]

Since the expected output is complementary in expected output and technology, the resulting allocation is efficient.

Under positive assortative matching, any worker with expected ability \( \eta \leq \eta(1_+) \) is matched with a firm of \( s = 1 \). Condition (ii) of the equilibrium then implies that

\[
W(\eta) = \eta \quad \text{for } \eta \leq \eta(1_+).
\]

Now consider the remaining firms with technology greater than 1 and the remaining workers with expected ability greater than \( \eta(1_+) \). Since the remaining technology and expected ability distribution do not have atoms, we can characterize the equilibrium using the standard mechanism design technique. Let’s index a firm’s type
by its technology level $s$. Let $\eta^*(s)$ and $W(\eta)$ be the allocation and transfer rule: if a firm of type $s$ announces that it is of type $s'$, its payoff is $s\eta(s') - W(\eta(s'))$. Let $\Pi(s) = s\eta^*(s) - W(\eta^*(s))$ be the equilibrium payoff of a firm of type $s$. Then a standard condition for the firms to announce their types truthfully is that

$$\Pi(s) = \int_1^s \eta^*(x)dx.$$ 

Since $\Pi(s) = s\eta^*(s) - W(\eta^*(s))$, the above implies that

$$W(\eta^*(s)) = s\eta^*(s) - \int_1^s \eta(x)dx \quad \text{for} \quad \eta(s) > \eta(1_+)$$

This expression and $W(\eta) = \eta$ for $\eta \leq \eta(1_+)$ completely pins down the equilibrium wage schedule. Together with the positive assortative matching, this wage schedule shows that if an equilibrium exists, it must be unique. It is also easy to check that the above prescribed allocation and transfer rule is an equilibrium. ■

**Proposition 1:** For a worker in cohort $n \geq 1$ with expected ability $\eta > \frac{\alpha+n}{\alpha+n-1}\eta_{1_+}$, the expected wage change after $t$ periods satisfies

$$E[W(\Gamma_{n+t}) - \frac{\alpha}{\alpha+\gamma}\eta(1_+)|\Gamma_n = \eta] = B(t, n, k)(W(\eta) - \frac{\alpha}{\alpha + \gamma}\eta(1_+)),$$

where $B(t, n, k) = (\frac{\alpha+n}{\alpha+n+t})(\frac{\alpha+n-1}{\alpha+n-k})^k > 1$. In particular,

(i): Wage increases with experience:

$$\frac{dE[W(\Gamma_{n+t})|\Gamma_n = \eta]}{dt} > 0.$$ 

(ii): Wage growth decreases with experience:

$$d(E[W(\Gamma_{n+t}) - W(\Gamma_n)|\Gamma_n = \eta]/dn < 0.$$ 

(iii): Wage growth is larger when matching is more important:

$$d(E[W(\Gamma_{n+t}) - W(\Gamma_n)|\Gamma_n = \eta]/dk > 0.$$

\[26\] Here, as in subsequent derivates with respect to $k$, we hold $\alpha$ constant and assume that the change of $k$ results from changes in $\gamma$. 

34
Proof. The condition $\eta > \frac{\alpha+n}{\alpha+n-t} \eta(1+)$ guarantees that the expected ability of the worker in any future periods will be greater $\eta(1+)$. Therefore, we have $W(\Gamma) - \frac{\alpha}{\alpha+T} \eta(1+)$. Define $C(t, n) = \frac{(\alpha+n-1)(\alpha+n+t)}{(\alpha+n)(\alpha+n+t-1)}$. From the discussion on the conditional distribution of expected ability, we see that $\Gamma_{n+t} = \text{Pareto}(C(t, n) \eta, \alpha + n, \frac{\alpha+n}{\alpha+n+t})$, we have

$$E[W(\Gamma_{n+t}) - \frac{\alpha}{\alpha+T} \eta(1+)|\Gamma_n = \eta] = \frac{\alpha+n}{\alpha+n+t} \left[ \frac{\lambda}{k} \frac{\eta}{\lambda} \right] C(t, n)^k + \int_{C(t, n)\eta}^\infty \frac{\lambda}{\alpha+n+t} \left[ (\alpha+n+t) \frac{t(\alpha+n)}{\alpha+n+k} C(t, n)^k \right] \frac{d\eta}{\alpha+n+k}. $$

To show that $B(t, n, k) = \frac{(\alpha+n-1)(\alpha+n+t)}{(\alpha+n)(\alpha+n+t-1)} > 1$ for all $k > 1$, we first note that $B(t, n, 1) = 1$. Therefore, it suffices to show that $\frac{d \log(B(t, n, k))}{dk} > 0$. We can write

$$\frac{d \log(B(t, n, k))}{dk} = \frac{d \log(B(0, n, k))}{dk} + \int_0^t \frac{d^2 \log(B(x, n, k))}{dkdx} dx.$$

It is easy to check that

$$\frac{d^2 \log(B(t, n, k))}{dk^2} = \frac{-1}{(\alpha+n+t)(\alpha+n+t-1)} + \frac{1}{(\alpha+n+k+t)^2} > 0,$$

where the inequality uses $k = 1 + \frac{\alpha}{\gamma} > 1$. Therefore, $\frac{d \log(B(t, n, k))}{dk} > 0$ and thus $B(t, n, k) > 1$.

Note that $E\left[ \frac{W(\Gamma_{n+t}) - W(\Gamma_n)}{W(\Gamma_n)} | \Gamma_n = \eta \right] = (B(t, n, k) - 1)\left[ \frac{W(\eta) - W(\eta(1+))}{W(\eta)} \right]$. Therefore,

$$\text{sgn}(d(E\left[ \frac{W(\Gamma_{n+t}) - W(\Gamma_n)}{W(\Gamma_n)} | \Gamma_n = \eta \right])/dk) = \text{sgn}(d(\log(E\left[ \frac{W(\Gamma_{n+t}) - W(\Gamma_n)}{W(\Gamma_n)} | \Gamma_n = \eta \right])/dk)) = \text{sgn}(d(\log(B(t, n, k) - 1))/dk),$$

so we have proved (iii).
Next, note that
\[
\frac{d \log(B(t, n, k))}{dt} = -\frac{1}{\alpha + n + t} + k\left(\frac{1}{\alpha + n + t} - \frac{1}{\alpha + n + t - 1}\right) + \frac{1}{\alpha + n - k + t},
\]
which implies that \(\frac{d \log(B(t, n, 1))}{dt} = 0\). Therefore,
\[
\frac{d \log(B(t, n, k))}{dt} = \frac{d \log(B(t, n, 1))}{dt} + \int_1^k \frac{d^2 \log(B(t, n, x))}{dtdx} dx > 0.
\]
This implies that \(dE[W(\Gamma_{n+1})|\Gamma_n = \eta] > 0\) and we have proved (ii).

Finally, note that \(\frac{d \log(B(t, n, k))}{dn} = 0\), \(\frac{d^2 \log(B(0, n, x))}{dn dx} = 0\), and
\[
\frac{d^3 \log(B(t, n, k))}{dk dt dn} = \frac{d^3 \log(B(t, n, k))}{dt^2 dk} = \frac{-1}{(\alpha + n + t)^2} + \frac{1}{(\alpha + n + t - 1)^2} - \frac{2}{(\alpha + n - k + t)^3} < 0
\]
because \(\frac{d^3 \log(B(t, n, 1))}{dt^2 dk} = \frac{1-3(\alpha+n+t)}{(\alpha+n+t)^2(\alpha+n+t-1)^3} < 0\) and \(\frac{d^4 \log(B(t, n, k))}{dt^2 dk^2} = \frac{-6}{(\alpha+n-k+t)^4} < 0\).

Therefore, we have
\[
\frac{d \log(B(t, n, k))}{dn} = \frac{d \log(B(t, n, 1))}{dn} + \int_1^k \frac{d^2 \log(B(t, n, x))}{dn dx} dx
\]
\[= \frac{d \log(B(t, n, 1))}{dn} + \int_1^k \left(\frac{d^2 \log(B(0, n, x))}{dn dx} + \int_0^t \frac{d^3 \log(B(y, n, x))}{dy dn dx} dy\right) dx
\]
\[< 0.
\]
And this implies (ii).

**Proposition 2**: For a worker in cohort \(n \geq 1\) with expected ability \(\eta > \frac{\alpha + n}{\alpha + n - 1} \eta(1+)\), the probability that the worker is a wage gainer next period decreases with \(n\):
\[d(\Pr(W(\Gamma_{n+1}) > W(\Gamma_n)|\Gamma_n = \eta))/dn < 0.\]
Proof. Since the matching is monotone,

\[ \text{sgn}(d(\Pr(W(\Gamma_{n+1}) > W(\Gamma_n)|\Gamma_n = \eta))/dn) = \text{sgn}(d(\Pr(\Gamma_{n+1}) > \Gamma_n|\Gamma_n = \eta))/dn) = \text{sgn}(d(\frac{1}{\alpha + n + 1} \left( \frac{(\alpha + n)^2 - 1}{(\alpha + n)^2} \right)^{\alpha+n})/dn) = \text{sgn}(d(\log(\frac{1}{\alpha + n + 1} \left( \frac{(\alpha + n)^2 - 1}{(\alpha + n)^2} \right)^{\alpha+n})/dn) = \text{sgn}(3 - (\alpha + n) \left( \frac{(\alpha + n)^2 - 1}{(\alpha + n)^2} \right)) \),

where the second equality uses the conditional distribution of the expected ability. Now it is clear that \( \log \left( \frac{(\alpha+n)^2-1}{(\alpha+n)^2} \right) < 0 \). In addition, we have \( 3 - (\alpha + n) < 0 \) because \( \alpha \geq 2 \) and \( n \geq 1 \).

Proposition 3: For a worker in cohort \( n \) with expected ability \( \eta > \frac{\alpha+n-1}{\alpha+n} \eta(1+) \), the correlation of wage changes satisfies

\[
\text{Corr}(W(\Gamma_{n+t+1}) - W(\Gamma_{n+1}), W(\Gamma_{n+1}) - W(\Gamma_n)|\Gamma_n = \eta) = \sqrt{\frac{1}{\left( \frac{r(t+1,n)-1}{r(1,n)-1} - 1 \right) \left( \frac{B(t,n+1)}{B(t,n+1)-1} \right)^2 + 1}}
\]

where \( r(t,n) = \frac{(\alpha+n+t)(1+\frac{t}{\alpha+n-t})}{(\alpha+n)(1+\frac{t}{\alpha+n-t})^2} \). In particular,

(i): The correlation decreases with experience:

\[
\frac{\partial \text{Corr}}{\partial n} < 0.
\]

(ii): The correlation is larger when matching is more important

\[
\frac{\partial \text{Corr}}{\partial k} > 0.
\]

(iii): The regression coefficient of future wage change \( (W(\Gamma_{n+t+1}) - W(\Gamma_{n+1})) \) on current wage change \( (W(\Gamma_{n+1}) - W(\Gamma_n)) \) satisfies

\[
\frac{\text{Cov}(W(\Gamma_{n+t+1}) - W(\Gamma_{n+1}), W(\Gamma_{n+1}) - W(\Gamma_n)|\Gamma_n = \eta)}{\text{Var}(W(\Gamma_{n+1}) - W(\eta_{1+}|\Gamma_n = \eta))} = B(t, n + 1, k) - 1.
\]

This coefficient increases with \( t \) and \( k \) and decreases with \( n \).
**Proof.** First, we note that

\[
E[(W(\Gamma_{n+t}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+))^2 | \Gamma_n = \eta]
\]

\[
= \frac{\alpha + n}{\alpha + n + t} C(t, n)^{2k} (1 + \frac{t}{\alpha + n - 2k})(W(\Gamma_n) - \frac{\alpha}{\alpha + \gamma} \eta(1_+))^2
\]

\[
= Z(t, n)(W(\Gamma_n) - \frac{\alpha}{\alpha + \gamma} \eta(1_+))^2.
\]

Therefore,

\[
Var(W(\Gamma_{n+t}) | \Gamma_n = \eta) = Var(W(\Gamma_{n+t}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+) | \Gamma_n = \eta)
\]

\[
= (Z(t, n) - B(t, n)^2)(W(\Gamma_n) - \frac{\alpha}{\alpha + \gamma} \eta(1_+))^2.
\]

Now we can calculate that

\[
E[(W(\Gamma_{n+t+1}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+))(W(\Gamma_{n+1}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+)) | \Gamma_n = \eta]
\]

\[
= E[(W(\Gamma_{n+1}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+))E[W(\Gamma_{n+t+1}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+)] | \Gamma_n = \eta, \Gamma_{n+1} = \eta_{n+1}]
\]

\[
= E[B(t, n + 1)(W(\Gamma_{n+1}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+))^2 | \Gamma_n = \eta]
\]

\[
= B(t, n + 1)Z(1, n)(W(\Gamma_{n+1}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+))^2.
\]

This implies that

\[
Cov(W(\Gamma_{n+t+1}), W(\Gamma_{n+1}) | \Gamma_n = \eta)
\]

\[
= Cov(W(\Gamma_{n+t+1}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+), W(\Gamma_{n+1}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+)) | \Gamma_n = \eta)
\]

\[
= (B(t, n + 1)Z(1, n) - B(t + 1, n)B(1, n))(W(\Gamma_n) - \frac{\alpha}{\alpha + \gamma} \eta(1_+))^2
\]

\[
= B(t, n + 1)(Z(1, n) - B^2(1, n))(W(\Gamma_n) - \frac{\alpha}{\alpha + \gamma} \eta(1_+))^2
\]

\[
= B(t, n + 1)Var(W(\Gamma_{n+1}) | \Gamma_n = \eta),
\]

where the second equality uses \(E[W(\Gamma_{n+t}) - \frac{\alpha}{\alpha + \gamma} \eta(1_+) | \Gamma_n = \eta] = B(t, n)(W(\Gamma_n) - \frac{\alpha}{\alpha + \gamma} \eta(1_+))\), and the third equality uses \(B(t + 1, n) = B(t, n + 1)B(1, n)\).
Therefore,

\[
Cov(W(\Gamma_{n+1}) - W(\Gamma_n), W(\Gamma_{n+1}) - W(\Gamma_n)|\Gamma_n = \eta) \\
= Cov(W(\Gamma_{n+1}), W(\Gamma_n)|\Gamma_n = \eta) - Var(W(\Gamma_{n+1}) - W(\Gamma_n)|\Gamma_n = \eta) \\
= (B(t, n + 1) - 1)Var(W(\Gamma_{n+1})|\Gamma_n = \eta).
\]

Since Proposition 5 shows that \(B(t, n + 1) > 0\), the covariance above is positive. The covariance formula also implies that

\[
\frac{Corr^2(W(\Gamma_{n+1}) - W(\eta_{n+1}), W(\Gamma_{n+1}) - W(\Gamma_n)|\Gamma_n = \eta)}{(B(t, n + 1) - 1)^2Var(W(\Gamma_{n+1}) - \frac{\alpha}{\alpha + \gamma}\eta(1)|\Gamma_n = \eta)} \\
= \frac{Var(W(\Gamma_{n+1}) - W(\Gamma_n)|\Gamma_n = \eta)}{(B(t, n + 1) - 1)^2Var(W(\Gamma_{n+1}) - \frac{\alpha}{\alpha + \gamma}\eta(1)|\Gamma_n = \eta)} \\
= \frac{Var(W(\Gamma_{n+1}) - 2Cov(W(\Gamma_{n+1}), W(\Gamma_n)) + Var(W(\Gamma_n))}{(B(t, n + 1) - 1)^2(r(1, n) - 1)} \\
= B(t, n + 1)^2(r(t + 1, n) - r(1, n)) + (B(t, n + 1) - 1)^2(r(1, n) - 1) \\
= \frac{1}{\left(\frac{r(t+1,n)-1}{r(1,n)-1} - 1\right)^2 + 1}
\]

where \(r(t, n) = \frac{Z(t,n)}{B^2(t,n)} = \frac{(\alpha+n+t)(1+\frac{t}{\alpha+n-k})}{(\alpha+n)(1+\frac{t}{\alpha+n-k})^2}\).

It is clear that \(\frac{B(t,n+1)}{B(t,n+1)-1}\) decreases with \(k\) and increases with \(n\). On the other hand,

\[
\frac{r(t+1, n) - 1}{r(1, n) - 1} = \frac{((\alpha + n - k + t + 1)^2 - (\alpha + n - k)^2)}{(\alpha + n - k + t + 1)^2} - 1
\]

Therefore, it is clear that \(d(\frac{r(t+1,n)-1}{r(1,n)-1})/dt < 0\). In addition, let \(x = \alpha + n - k\) and then we have

\[
\frac{d\log(\frac{r(t+1,n)-1}{r(1,n)-1})}{dx} = 2\left(\frac{1}{2x + t + 1} - \frac{1}{x + t + 1} + \frac{1}{x + 1} - \frac{1}{2x + 1}\right) \\
= 2\left(\frac{-x}{(x + t + 1)(2x + t + 1)} + \frac{1}{x + 1} - \frac{1}{2x + 1}\right) \\
> 0
\]

39
because this expression increases with $t$ and it is 0 when $t = 0$. This implies that

$$\frac{d \log \left( \frac{r(t+1,n)}{r(1,n)} \right)}{dn} > 0, \quad \text{and} \quad \frac{d \log \left( \frac{r(t+1,n)}{r(1,n)} \right)}{dk} < 0.$$  

The discussion above implies that both $\frac{B(t,n+1)}{B(t,n+1)}$ and $\frac{1}{\frac{B(t,n+1)}{B(t,n+1)}}$ decreases with $n$, increases with $t$ and $k$. Therefore, the correlation increases with $t$ and $k$ and decreases with $k$. ■