A Theory of Wage Distribution and Wage Dynamics with Assignment and Pareto Learning

Jin Li*

Kellogg School of Management

Northwestern University

January 11, 2013

Abstract

This paper develops a model of wage distribution and wage dynamics based on assignment and Pareto learning. The model matches a large number of key facts about wage distribution and wage dynamics. The tractability of Pareto learning allows us to derive joint implications on the wage distribution and wage dynamics as assignment becomes more important. Our model also provides a natural framework for decomposing the earning variance into a permanent component and a transitory one, and it helps explain why the growing importance in assignment can lead to both higher wage inequality and higher wage instability.

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, Marko Tervio, Jose Tessada, Scott Stern, Mike Waldman, and the seminar participants at Chicago Fed, Chicago GSB, M.I.T., Kellogg Northwestern, and Penn State for helpful comments and discussions. I also thank Jun Yu for excellent research assistance. All remaining errors are mine.

1 Introduction

In this paper, we develop a simple model of wage formation based on two assumptions: a) information about workers' abilities is learned publicly over time and b) workers are assigned to firms of differing levels of productivity. Both learning and assignment have been recognized as important factors in the wage-determination process.[1] One objective of this paper is to show that the combination of the two helps explain a large number of important features of wage distribution and wage dynamics. To realize this objective, we introduce a convenient learning process called Pareto learning, which allows us to obtain explicit formulas about wage distribution and wage dynamics. In addition to matching the well-known patterns, the tractability of the model also points to new empirical implications and allows us to study the joint implication of assignment on wage dynamics, inequality, and instability.

Concerning wage distribution, Neal and Rosen (2000) summarize several important empirical regularities. First, wage distributions tend to be skewed to the right with the average wage exceeding the median wage and the top earners capturing a large proportion of the earnings. Moreover, wage distributions have a fat tail similar to distributions such as Pareto or log-normal. Third, in terms of its evolution, the wage distribution moves to the right and the wage dispersion increases as the cohort ages. These findings are robust across populations (Lydall (1968)) and the skewness property is known at least since Pareto (1897).

Concerning wage dynamics, Rubinstein and Weiss (2006) point to several key patterns for white U.S. males. First, the average wages of workers increase with experience. Second, the average rates of wage growth decline with experience. Third, the proportion of nominal wage "gainers," those with a wage increase between consecutive wage observations, decreases with labor market experience. Together, these patterns indicate that the life-cycle earning profile is increasing with and concave in experience, which is a basic finding of wage dynamics; see, for example, Mincer (1974), Murphy and Welch (1992), and Heckman, Lochner and Todd (2006).

One purpose of the paper is to develop a simple model that captures these well-known facts about the wage distribution and wage dynamics. In our model, there is

[1] Papers that use learning to explain wage changes include Jovanovic (1979), Harris and Holmstrom (1982), Holmstrom (1999), and Farber and Gibbons (1996). Sattinger (1993) and Neal and Rosen (2003) provide excellent reviews for assignment models.

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 assigned to a firm. Production is stochastic, and expected output is complementary in the worker's ability and firm's technology. The complementarity implies that in equilibrium, more-productive firms are assigned to workers with a higher expected ability.

A worker's output is observed publicly, allowing the worker's ability to be learned 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 enables us to obtain explicit formulas for the aggregate wage distributions that match the key empirical regularities. The formulas show that the wage distribution is skewed to the right, with the mean wage exceeding the median. Moreover, the right tail of the wage distribution approximates a Pareto distribution with a power parameter affected by the importance of assignment. Third, in terms of the evolution of wage distributions, both the median wage and wages at higher quantile levels increase with cohort age. And finally, wage distributions become more dispersed as the cohort ages in terms of the wage ratio between quantiles.

Pareto learning also allows us to explicitly calculate various aspects of wage dynamics at the individual level. In terms of the expected changes in wages, the formulas show that the expected wage grows with experience, and the rate of wage growth decreases with experience. Specifically, the formula shows that the model's implication on wage growth is similar to the well-known Mincerian regression so that the expected wage growth is proportional to the existing wage level. In terms of the distribution of the wage changes, the formulas show that younger workers are more likely to have wage gains, and wage decreases for older workers are more moderate.

Beyond matching these well-known findings on the wage distribution and wage dynamics separately, our model shows that there are important connections between wage distributions at an aggregate level and wage dynamics at an individual level.

A key application of the model is to study the impact of assignment on earning instability and inequality jointly. Pioneering work by Gottschalk and Moffitt (1994) shows that both earning instability and inequality have risen from the 1970s to 1980s and that one-third to one-half of the rise in the variance of earnings can be attributed to the rise in earning instability. While it is typically thought that the increases in earning inequality and instability result from separate sources, our model provides a straightforward mechanism for the joint increase. As assignment becomes more important, it amplifies both the wage distribution and wage changes resulting from changes in expected ability.

In linking earning instability with earning inequality, our model offers a natural framework for decomposing the variance of earning to a permanent component and a transitory one. Our decomposition is similar to those used in empirical analyses (for example in Gottschalk and Moffitt (1994)) and has the advantage of being easily implemented. More importantly, by being explicit about the wage-generating process, our model imposes structure on the error terms (disturbances) associated with the realized wages. For example, our model suggests that the distribution of the (unexpected) wage changes should have a positive skewness, which can be tested empirically. In addition, our model suggests that while wage dispersion is larger as a cohort ages, the transitory variance of earnings for each worker should decrease because less information is learned over time. This is consistent with the finding of Gottschalk and Moffitt (1994) that transitory variance decreases with experience across all education groups. While these features–positive skewness of wage changes and diminishing earning instability–have not received much attention, they are nevertheless important because decisions such as buying a house or having a child depend crucially upon one's expectation of how much and in which direction one's future earnings will change.

This paper contributes to two strands of literature. First, by relating the model to several empirical regularities, the paper contributes to the literature of models that explain broad patterns of wage distributions and wage dynamics. Closely related papers in this category include MacDonald (1982), Chiappori, Salanie, and Valentin (1999), Gibbons, Katz, Lemieux, and Parent (2005) (hereafter GKLP), and Gibbons and Waldman (1999).[2] MacDonald (1982) studies learning and matching in a general

[2] Other models in this category include Jovanovic (1979), Harris and Holmstrom (1982), MacLeod and Malcomson (1988), Demougin and Siow (1994), and Bernhardt (1995).

equilibrium model, where matching is at a horizontal level, rather than the vertical level here. Chiappori, Salanie, and Valentin (1999) consider models with learning and downward rigid wages, and here the downward rigidity is directly related to the learning process. GKLP explore the implications of learning and the comparative advantage (of ability across sectors) on wage dynamics using a normal learning model. Gibbons and Waldman (1999) explain a broad pattern of wage and promotion dynamics inside firms by combining human capital accumulation, job assignment, and learning. Relative to these papers, our model is more tractable analytically. Our explicit formula allows us to the study the joint effect of assignment on wage inequality, wage instability, and the rate of wage growth.

Second, this paper complements workhorse models of the earning process such as human capital and search models by better capturing some aspects of wage dynamics. In human capital models (Becker (1975), Mincer (1974), and Ben-Porath (1967)), for example, the standard Mincerian regression with a quadratic function of experience underestimates the wage growth of workers in the earlier stages of their careers (Murphy and Welch (1992)). Our model suggests that the effect of learning and assignment is concentrated in the early stages of a worker's career and tapers off quickly, pointing to the importance of these factors in wage formation for young workers. Relative to search models, a reinterpretation of our model as assigning workers to tasks within firms helps generate wage growth within firms. In addition, our model is perhaps better suited for describing the earnings of workers at the top of the distribution where a small difference in ability can generate an enormous wage difference. Our model illustrates the dynamic process in which these enormous wage differences occur.

The rest of the paper is organized as follows. We set up and solve for the model in Section 2. Section 3 explores the empirical implications of the model. Section 4 concludes, and the proofs are provided in the Appendix.

2 Model

In this section, we present a symmetric learning model with assignment. Subsection 2.1 describes the basic economic environment, and we discuss the special feature of the model, the Pareto learning process in Subsection 2.2. Subsection 2.3 specifies the equilibrium conditions. Subsection 2.4 characterizes the unique equilibrium and gives explicit formulas for the equilibrium assignment 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 in which 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 is the lower bound of the support of the distribution and the second parameter determines the rate the density of the distribution decreases:

$$\Pr(A \geq a) = \begin{cases} a^{-\alpha} & \text{for } a > 1; \\ 1 & \text{for } a \leq 1. \end{cases}$$

We use upper-cases to denote random variables and lower-cases to denote their realizations. In a Pareto distribution, the proportion of workers with an ability above a threshold level decreases with the threshold at a constant rate. To ensure that the ability distribution has a finite variance, we assume $\alpha > 2$.

The utility of each worker is separably additive across periods with a discount factor $\delta \in (0,1)$. Workers are risk neutral and have no disutility of effort. 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. The outside option is normalized to be 1 for simplicity. When the outside option is not 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 $(1,\gamma,1-\sigma^\gamma)$, where the last parameter denotes the size of the point mass at the lower-bound of the distribution:

$$\Pr(S \geq s) = \begin{cases} \left(\frac{s}{\sigma}\right)^{-\gamma} & \text{for } s > 1; \\ 1 & \text{for } s \leq 1. \end{cases}$$

In other words, this distribution is a linear combination of a point mass at the bottom (of size $1-\sigma^{\gamma}$) with a Pareto distribution with parameter $(1, \gamma)$. The point mass at the bottom helps simplify the expressions for the assignment formula and wage formula and does not affect qualitatively our results on wage dynamics and wage distributions. We assume $\gamma > 1$ for the distribution to have a finite mean, and let F denote the CDF of the technology distribution.

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 is equal to

$$E[Y(s, a, Z)] = \int_0^a \frac{2sz}{a} dz = sa. \tag{1}$$

This production function has two properties. First, technology and ability are complements in the production. This implies that there is gain from assigning more able workers to better technologies. Second, 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.

2.2 Pareto Learning

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 of outputs of the workers and the technologies of the firms they are employed with. Consequently, firms know $\{z_t\}_{t=0}^{n-1}$, where $z_t = \frac{y_t}{2s_t}$ is the stochastic factor in the production. Given that a) for a worker of ability a, Z_t is drawn uniformly between $[0, a]$, and b) the prior ability distribution of the worker is Pareto, this process of learning about a from $\{Z_t\}$ is called *Pareto learning.*[3]

One key advantage of Pareto learning is its simplicity in calculating the posterior distributions. Recall that workers share a prior ability distribution $Pareto(1, \alpha)$. Now consider a worker in cohort n. Let $m = \max\{z_0, ..., z_{n-1}, 1\}$ be the *de facto* maximal draw of ability. The posterior ability of the worker is then given by $Pareto(m, \alpha+n)$:

$$\Pr(A \geq a_n | m, n) = \begin{cases} (\frac{a_n}{m})^{-(\alpha+n)} & \text{for } a_n \geq m; \\ 1 & \text{for } a_n < m. \end{cases} \tag{2}$$

The formula implies that the worker's posterior distribution of ability is completely determined by the maximal draw of ability and his cohort age.

Pareto learning leads to a simple expression for the expected ability. For a worker in cohort n with maximal draw m, his expected ability $\eta_n(m)$ is given by

$$\eta_n(m) = \frac{\alpha + n}{\alpha + n - 1} m. \tag{3}$$

This formula shows that the expected ability is increasing in the maximal draw and is decreasing in the number of draws. Since $(\alpha + n)/(\alpha + n - 1) > 1$, the worker's expected ability always exceeds m. This reflects the option value of worker's ability since there is always possibility of a higher draw in the future. In particular, if the worker has a lucky draw–m being close to a–his expected ability then exceeds his true ability especially when n is small. Finally, note that $(\alpha + n)/(\alpha + n - 1)$ decreases with n. This implies that the worker's expected ability drops when the new draw does not exceed the previous maximum.

For our analysis of wage dynamics, it is useful to provide a formula for the dynamics of the expected ability. Note that for a worker in cohort n, his expected ability (η_n) is a sufficient statistic for his ability distribution since there is a one-to-one relationship between the expected ability and the maximal draw. In this case, the distribution of the worker's expected ability in t periods in the future is a generalized

[3] See Section 9.7 of DeGroot (1970) for a detailed exposition of Pareto learning. Despite the attractive properties of Pareto learning, Barbarino and Jovanovic (2007) is the only paper I am aware of that uses it.

Pareto distribution with parameter $(\frac{(\alpha+n-1)(\alpha+n+t)}{(\alpha+n)(\alpha+n+t-1)}\eta_n, \alpha+n, \frac{\alpha+n}{\alpha+n+t})$:

$$\Pr(\Gamma_{n+t} \geq \eta_{n+t}|\Gamma_n = \eta_n) = \begin{cases} \frac{t}{\alpha+n+t}\left(\frac{(\alpha+n-1)(\alpha+n+t)}{(\alpha+n)(\alpha+n+t-1)}\frac{\eta_n}{\eta_{n+t}}\right)^{\alpha+n} & \text{for } \eta_{n+t} > \frac{(\alpha+n-1)(\alpha+n+t)}{(\alpha+n)(\alpha+n+t-1)}\eta_n, \\ 1 & \text{for } \eta_{n+t} \leq \frac{(\alpha+n-1)(\alpha+n+t)}{(\alpha+n)(\alpha+n+t-1)}\eta_n. \end{cases}$$

The formula is formally derived in the Appendix, and we discuss its properties here. The formula shows that changes in the expected ability under Pareto learning have an "upward skewness" property so that the adjustment in expected ability is unbounded above and bounded below. For a worker in cohort n, the maximal drop in expected ability happens when all of the new signals (in the next t periods) below the existing maximal draw. This happens with probability $(\alpha+n)/(\alpha+n+t)$, which increases with the cohort age n and decreases with t. In this case, the expected ability of the worker adjusts downwards to $(\alpha+n-1)(\alpha+n+t)\eta_n/((\alpha+n)(\alpha+n+t-1))$, which is bounded by $(\alpha+n-1)\eta_n/((\alpha+n)$. This implies that the expected ability is relatively downward rigid in the sense that the maximum proportional downward adjustment is bounded by $1/(\alpha+n)$.

Finally, for our analysis of the wage distributions, we provide formula (again derived in the Appendix) for the distributions of the expected abilities. The distribution of expected ability of workers in cohort n is a generalized Pareto with parameter $(\frac{\alpha+n}{\alpha+n-1}, \alpha, \frac{\alpha}{\alpha+n})$:

$$\Pr(\Gamma_n \geq \eta_n) = \begin{cases} \frac{n}{\alpha+n}(\frac{\alpha+n-1}{\alpha+n}\eta_n)^{-\alpha} & \text{for } \eta_n \geq \frac{\alpha+n}{\alpha+n-1}; \\ 1 & \text{for } \eta_n < \frac{\alpha+n}{\alpha+n-1}. \end{cases}$$

The formula shows that same as the ability distribution, the upper tail of the expected ability distribution is Pareto with parameter α. Different from the ability distribution, the expected ability distribution has a mass of $\alpha/(\alpha+n)$ at the bottom, which locates at $(\alpha+n)/(\alpha+n-1)$. 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 differ for each expected ability distribution, the upper tail of this rate α is independent of cohort age n. This makes it easy to aggregate the expected ability distribution across all cohorts. Let G denote CDF of the expected ability distribution of the entire economy. Then for an expected ability level $\eta > \frac{\alpha}{\alpha-1}$, G satisfies

$$1 - G(\eta) = (\frac{\eta}{\lambda})^{-\alpha},$$

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

2.3 Equilibrium Concept

In this model, an allocation involves a) a wage function that assigns wages to workers of varying expected abilities and b) an assignment function that assigns workers to firms. Recall that G is the CDF of the expected ability distribution of the entire economy and F is the CDF of the technology distribution. An allocation is an equilibrium if the following conditions are satisfied:

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

(ii): *There exists an assignment function $\eta^*(s)$ that maps technologies to expected abilities such that for each firm with technology s,*

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

(iii): *The expected ability distribution of the economy satisfies*

$$G(\eta) = \int_0^{\infty} 1_{\{\eta^*(s)\leq\eta\}} dF(s) \qquad \text{for all } \eta.$$

Condition (i) 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 (1)), so each worker's value to a firm depends solely on his expected ability. Condition (ii) 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.[4] Condition

[4]For papers that explore the implications of long-term contracts on the labor market outcome;

(iii) states that the demand for workers equals the supply for every level of expected ability. It is written in an integral form so that we require the demand for workers below any expected ability level is equal to the supply.

While we have cast the model as one in which workers are assigned to different firms, the model can also be thought of as assigning workers to different tasks.[5] In particular, we can recast the model so that it has a continuum of identical firms, where the production of each firm requires a distribution of tasks with different productivities. There is complementarity between the ability of the worker and the productivity of the task. In a market equilibrium, there is wage function associated with the worker's expected abilities. Each firm chooses the assignment of workers to tasks to maximize its profit, and the demand for workers equals the supply. Under this reinterpretation, changes of jobs are thought of as changes in tasks. Therefore, wage changes in this model can be interpreted as arising from either changes in tasks within firms or changes in jobs across firms. This implies that while the model has clear predictions on the wage distributions and wage dynamics, it does not speak directly to the source of the wage growth being either within and between firms.

2.4 Characterizing 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 wage and assignment function and discuss their properties. To simplify the exposition, we first introduce the following definition.

Definition 1: Let $\eta_l = \lambda\sigma^{-\frac{\gamma}{\alpha}}$ be the unique expected ability level such that $G(\eta_l) = F(1)$.

As will be seen in Theorem 1, η_l is the cutoff expected ability level so that a worker will be assigned to a firm with technology $s > 1$ if and only if his expected ability is above η_l.

Theorem 1: *There exists a unique equilibrium with assignment function* η^* and wage

see, for example, Harris and Holmstrom (1982), Beaudry and DiNardo (1991). Felli and Harris (1996) derives the optimal job mobility rule when the speed of learning depends on the firm the worker is employed with.

[5] I thank an anoymous referee for this suggestion.

function W *such that*

$$\begin{array}{ll} \eta^*(s) = G^{-1}(F(s)) & \text{for } s > 1; \\ W(\eta) = \eta & \text{for } \eta \leq \eta_l; \\ W(\eta^*(s)) = s\eta^*(s) - \int_1^s \eta^*(x)dx & \text{for } \eta^*(s) > \eta_l. \end{array}$$

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

Characterizing the market equilibrium in this model is similar to solving mechanism design problem. To see this, we rewrite the firm's profit maximization condition ((ii) in the equilibrium definition) as:

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

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 assigned 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 formula, we assume that $\sigma^{-\gamma} < (\lambda(\alpha-1)/\alpha)^{-\alpha}$, so all the atoms in the expected ability distribution G will be assigned to firms with technology $s = 1$ in equilibrium. We also define $w(s)$ to be the equilibrium wage that a firm of technology s pays.

Finally, to simplify the exposition, we define the following parameter as a measure for the importance of assignment. Let

$$k = 1 + \frac{\alpha}{\gamma},$$

then a bigger k implies that assignment is more important. In fixed α, k becomes larger when γ is smaller, corresponding to a right-shift of the technology distribution.

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

(i): *The equilibrium assignment function is equal to*

$$\eta^*(s) = \lambda(\frac{s}{\sigma})^{\frac{\gamma}{\alpha}} \qquad \text{for } s > 1.$$

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

$$\begin{array}{ll} W(\eta) = \eta & \text{for } \eta < \eta_l; \\ W(\eta) = \frac{\sigma\lambda}{k}(\frac{\eta}{\lambda})^k + \frac{\alpha}{\alpha+\gamma}\eta_l & \text{for } \eta \geq \eta_l. \end{array}$$

(iii): *The equilibrium wage as a function of technology is equal to*

$$w(s) = \frac{\lambda}{k}\sigma^{-\frac{\gamma}{\alpha}}(s^{\frac{\gamma}{\alpha}+1} + 1), \qquad \text{for } s > 1.$$

Statement (i) in Corollary 1 is a restatement of the result that the equilibrium assignment is positive assortative: workers of higher expected abilities are matched with firms of better technologies.[6] This is because expected ability and technology are complements in production so firms with better technology benefit more from a high ability worker. The assignment 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 assigned with workers of higher expected ability. Second, as the technology distribution shifts to the right (corresponding to an decrease in γ), firms of a given technology level $s > 1$ will be assigned 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 η. This implies that wage is roughly exponential in expected ability. The wage formula also implies that wage is convex in expected ability. The source of wage convexity results from positive assortative matching. Higher expected

[6] While positive assortative matching is an intuitive result, empirical findings on this appear to be mixed; see Abowd, Kramarz, Perez-Duarte, and Schmutte for a recent paper on this topic. I think an anoymous referee for pointing this out.

ability leads to matching with better technologies, which makes workers even more productive and, thus, creates a convex wage profile.

3 The Empirical Implications

In this section, we explore the empirical implications of the model. In the first two subsections, we show that the model is consistent with important features of wage distributions and wage dynamics. In the last subsection, we study the joint implication of the importance of assignment on earning instability, earning inequality, the rate of wage growth, and the serial correlation of wage growth.

3.1 Wage Distributions

In this subsection, we show that our model helps explain many stylized facts about wage distributions. In terms of the shape of the distribution, the model matches the fact that the wage distribution is skewed to the right and has a fat tail. In terms of the patterns of the evolution of the wage distribution, the model matches the fact that the median wage of any given cohort increases and that the wage distribution becomes more dispersed over time. We derive these implications by obtaining below the formula that describe the right tail of the wage distributions.

Proposition 1 *For $w > W(\eta_l)$, the wage distribution of the whole workforce is given by*

$$\Pr(W \geq w) = (\frac{\lambda\sigma}{k})^{\frac{\alpha}{k}}(w - \frac{\alpha\lambda\sigma^{-\frac{\gamma}{\alpha}}}{\alpha+\gamma})^{-\frac{\alpha}{k}}.$$

The wage distribution of workers in cohort n is given by

$$\Pr(W_n \geq w) = s(n)(\frac{\lambda\sigma}{k})^{\frac{\alpha}{k}}\lambda^{-\alpha}(w - \frac{\alpha\lambda\sigma^{-\frac{\gamma}{\alpha}}}{\alpha+\gamma})^{-\frac{\alpha}{k}},$$

where $s(n) = \frac{n}{\alpha+n}(\frac{\alpha+n-1}{\alpha+n})^{-\alpha}$ and increases with n.

Proposition 1 reflects the convenient property of Pareto learning in aggregating wages across different workers. In particular, we calculate the aggregate wage distribution of each cohort by integrating wage (as a function of expected ability) over the expected ability distribution. In general, the wage function and the density function of the expected ability belong to different classes for functions, making the aggregation difficult. For example, in the commonly used normal learning models, expected abilities are normally distributed, and when wage functions are proportional to (or are powers of) the expected abilities, the resulting wage distributions are gamma functions, which do not allow for an explicit formula. In our case, both the density function and the wage function are proportional to the powers of the expected ability, making the aggregation possible. Below, we use the formula above to study the basic properties of the wage distributions.

Corollary 2: *Wage distributions satisfy the following.*

(i): *The wage distributions, both for each cohort n and for the whole workforce, are skewed to the right:*

$$\begin{aligned} E[W_n] &> W_n^{1/2} \quad \text{for all } n. \\ E[W] &> W^{1/2}, \end{aligned}$$

where $W_n^{1/2}$ is the median wage of cohort n and $W^{1/2}$ is the median wage of the whole workforce.

(ii): *The right tail of the wage distribution approximates a Pareto distribution with power α/k :*

$$\lim_{w\to\infty} \frac{\Pr(W > rw)}{\Pr(W > w)} = r^{-\frac{\alpha}{k}} \quad \text{for all } r > 0.$$

(iii): *For a cohort with age $n > \alpha$, the median wage of the cohort increases with n. Moreover, the wage at quantile q also increases if $q > \alpha/(\alpha+n)$:*

$$W(\eta_{n+1}^q) > W(\eta_n^q).$$

(iv): *Wage distributions become more dispersed over time as measured by quantile ratios. Let $0 < q_l < q_h < 1$. For all n with $q_h > \alpha/(\alpha+n)$,*

$$\frac{W(\eta_{n+1}^{q_h})}{W(\eta_{n+1}^{q_l})} > \frac{W(\eta_n^{q_h})}{W(\eta_n^{q_l})},$$

where η_n^q the qth quantile of expected ability in cohort n.

Parts (i) and (ii) match important empirical regularities on the shape of wage distributions. Part (i) shows that the wage distribution is skewed to the right both for the whole workforce and for each cohort. The skewness of the wage distribution is known at least since Pareto (1897), who also shows that the wage distribution has a fat tail. Part (ii) shows that the right tail of the wage distribution is similar to a generalized Pareto distribution, and, thus, has a fat tail.

The skewness in Part (i) is a common property of models with assignment. The wage distribution is skewed to the right of the ability distribution when the wages of higher-ability workers reflect not only their superior abilities but also the better technologies with which they are matched. Such positive assortative matching, i.e., higher-ability workers being assigned to better technologies, happens when a worker's ability and the technology are complements in production, which is the case here. The fat-tail property in Part (ii), in contrast, depends on the underlying assumption of the ability and technology distributions. The wage distribution has a fat tail as long as either the distribution of ability or technology has a fat tail. In this model, we assume that ability has a Pareto distribution and the technology is a generalized Pareto. These assumptions imply that wage distributions not only have fat tails but also are similar to Pareto distributions.

Parts (iii) and (iv) match empirical regularities on the evolution of wage distributions. Part (iii) shows that the median wage increases under fairly mild conditions. In general, the q^{th} quantile of the wage distribution in cohort n increases as long as $q > \alpha/(\alpha + n)$. Part (iv) shows that the dispersion of the wage distribution, as measured by the ratio of wages at different quantiles, such as the 90/50 ratio, is increasing over time. It is well known that the median wage increases over time and the wage distribution also becomes more dispersed over time; see Neal and Rosen (2000) for a review.

The increasing median wage as the cohort ages reflects the *upward-skewness* of Pareto learning. In this model, the increase in the median wage is tied to the increase in the median expected ability. Under Pareto learning, the downward change in the expected ability is limited and small, but the upward change in the expected ability is unlimited. The small downward change implies that most workers with current expected abilities above the median will remain above it in the next period, and the unlimited upwardness implies that many workers with current expected abilities below the median will see their wage exceed the current median. This process implies that the median of the expected ability increases over time. Note that in a more standard learning process such as normal learning, the median wage will not change. In contrast, the upward-skewness property of Pareto learning allows us to better fit the empirical pattern on median wages.

The widening of the wage distribution over time in Part (iv) is a general feature of models with learning. In learning models, all workers share the same expected ability (conditional on the observables) upon entering the labor market. As more information about each worker is revealed over time, the expected ability distribution approaches the true ability distribution and becomes more dispersed in terms of Second Order Stochastic Dominance. Since wage is increasing in the expected ability, the widening of the expected ability distribution suggests that the wage distribution also becomes more dispersed.

3.2 Wage Dynamics

In this subsection, 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. When applied separately for each observationally equivalent group of workers, the model generates a wage-growth pattern for each subgroup.

We start by giving an explicit formula that governs the change of the expected wage. To state the formula more succinctly, define the approximate wage function

$$\widetilde{W}(\eta) = W(\eta) - \frac{\alpha}{\alpha+\gamma}\eta_l = \frac{\sigma\lambda}{k}(\frac{\eta}{\lambda})^k, \tag{4}$$

where recall that η_l is the cutoff expected ability below which workers are assigned to firms with the lowest technology level ($s = 1$).

Proposition 2 *For a worker in cohort n with expected ability $\eta > \frac{\alpha+n}{\alpha+n-1}\eta_l$, the expected wage change after t periods satisfies*

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

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

Proposition 2 shows that the model's prediction on wage growth is similar to the Mincerian regression in the sense that *log-wage* is a function of experience. In particular, the formula shows that the expected wage increase for a worker is proportional to the current wage minus a constant. When the current wage is sufficiently high, the expected wage increase is essentially proportional to the current wage, as specified in the Mincerian regression.

Unlike the Mincerian regression, this model does not restrict the effect of experience on log-wage to be quadratic. Murphy and Welch (1990) note that the quadratic formulation understates the early career earnings growth by about 30%-50%. Our model suggests that a sizable fraction of the early career earnings growth can be explained by learning and assignment. In particular, quantitative estimates suggest that much of the wage growth in this model appears in the earlier stages of the worker's career. For example, when $\gamma = 2$ and $\alpha = 2$, the model predicts that the total life-time expected wage growth is about 30%, and much of the wage growth is concentrated in the initial periods: the expected wage growth in the first period is about 18%.

From Proposition 2, we can derive two well-known facts about wage dynamics.

Corollary 3: *For* $\eta > \frac{\alpha+n}{\alpha+n-1}\eta_l$*, the following is true.*

(i): *The expected wage increases with experience:*

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

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

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

While we obtain Corollary 3 by direct calculations, the intuition behind the results is straightforward. The reason that the expected wage of a worker increases with experience is due to the combination of learning and assignment. In learning models without assignment, the expected wage is a martingale, so the expected wage change is zero. With assignment, extra information acquired through learning about the worker's ability enables better assignment over time. Better assignment raises the average output of the workers and, thus, their average wage increases over time. In this case, the expected wage is a sub-martingale, so the expected wage change is positive.

The reason that the rate of wage growth decreases with experience is because the productivity gain from better assignment decreases over time. In particular, there are two sources for the decrease. First, the marginal value of each new signal becomes smaller over time. This is because when more information about the worker's ability is known, his expected ability changes less with respect to new signals. Second, the marginal return from better assignment becomes smaller over time because the existing assignment has become better over time. Together, these two channels help explain why expected wage gains decrease over time, and our model helps demonstrate this result analytically by using the explicit formula of wage growth.

In addition to the change in the expected wage, our model also has implications on the distribution of the realized wage changes.

Proposition 3 *For* $\eta > \frac{\alpha+n}{\alpha+n-1}\eta_l$, *the following is true.*

(i): *Wage increases are more common for younger workers:*

$$d(\Pr(W(\Gamma_{n+1}) < W(\Gamma_n)|\Gamma_n = \eta]/dn < 0$$

(ii): *Wages cuts (in percentages) are limited:*

$$\log(W(\Gamma_{n+1})) - \log W(\Gamma_n) > -\frac{k}{(\alpha+n)^2 - 1}.$$

Part (i) shows that the proportion of wage gainers decreases with experience. This results directly from Pareto learning: to experience a positive wage change, it is necessary for the new draw of ability to exceed the current maximum. This becomes less likely for older workers as the maximum increases over time. While we are not aware of direct evidence supporting Part (i), there is some suggestive evidence for it. Rubinstein and Weiss (2006) find that, for nominal wages, the percentage of wage gainers decreases with experience both for high school and college graduates.

While Pareto learning helps show that wage gains are more common for young workers, the model in its current form also predicts that wage gains are rare for old workers: for a worker in cohort n, the chance of having a wage gain is roughly $1/n$, which seems too small to be plausible. The reason for the small probability of wage increase is because of the upward skewness property of Pareto learning. Especially for older workers, Pareto learning predicts that most workers experience a small drop in expected ability, and a few experiences large gain.

One way to obtain a more reasonable prediction for the frequency of wage cuts is to add a small learning-by-doing component to ability, which helps offset the decrease in expected ability.[7] Part (ii) shows why the learning-by-doing component does not need to be big: for workers in cohort n, the percentage of the (real) wage decrease is bounded by $k/((\alpha+n)^2-1)$. In particular, for workers with more than 10 periods of experience, and with the parameter for gauging the rates of wage growth ($\gamma = 2$ and $\alpha = 2$), Part (ii) implies the percentage real wage drop from learning and assignment is bounded by 2%. Even without the learning-by-doing component, the small decrease in real wages in the model has two empirical implications. First, to the extent that inflation is not negligible, our model is consistent with the findings that nominal wages are downward rigid. Second, to the extent that most workers have small wage decrease and a few have large wage gains, the distribution of errors in wage regressions has a positive skewness.

3.3 Joint Implication of Assignment on Earning Inequality, Instability, and Growth Rate

In this subsection, we study how the importance of assignment jointly affects earning instability, inequality, and the growth rate. We also relate the model's predictions to the empirical findings.

Our model provides a framework for studying the effects of assignment on both earning instability and wage inequality. Our model leads to a natural decomposition of the earnings into a transitory and permanent component, and we show that both the transitory and permanent variance of the wage distribution increase as assignment becomes more important. This sheds light on the joint increase in earning inequality and earning instability in the U.S. in recent years.

[7] I thank an anoymous referee both for pointing out this issue and for offering the learning-by-doing suggestion.

In an influential paper, Gottschalk and Moffitt (1994) provide a method for decomposing the total earning inequalities into a permanent part and a transitory part. In their conceptual framework, the log-wage of worker i in period t is the sum of the two parts:

$$w_{it} = a_i + \varepsilon_{it},$$

where a_i reflects the permanent component of earning, and ε_{it} represents the period-to-period random shocks. The variance of a_i is a measure of the permanent earning inequality and the variance of ε_{it} (average over i and t) is a measure of earning instability.

Our model implies a similar decomposition as in Gottschalk and Moffitt (1994). The corresponding measure for the permanent component of the worker's earnings in our setting is $\log W(a)$, which is the worker's long-run log-wage. Therefore, we define the permanent variance of earnings as

$$PV = Var_a[\log W(a)].$$

The transitory variance of earnings in cohort n is defined as

$$TV_n = E_{\eta_n}[Var_{\eta_{n+1}}[\log W(\eta_{n+1}) - E[\log W(\eta_{n+1})|\eta_n]].$$

The difference between $\log W(\eta_{n+1})$ and $E[\log W(\eta_{n+1})|\eta_n]$ captures the deviation of the actual log-wage from the (best guess of the) long-run average wage, and it corresponds to the ε_{it} term in Gottschalk and Moffitt (1994).

Since our model is explicit about the source of wage changes, this leads to several empirical implications. First, the transitory variance of earnings decreases with the cohort's age. The wage change here reflects the new information about the worker's ability. As the worker ages, less information is learned about the worker's ability, suggesting that the transitory variance of earnings should also decrease with experience. This is consistent with the empirical findings: Gottschalk and Moffitt (1994) find that between 1970-1978, the transitory variance of the earnings of the workers aged 20-29 is 53% higher than of those aged 30-39, which is in turn 35% higher than that of workers aged 40-49. Similarly, between 1979-87, the transitory variance of the earnings of workers aged 20-29 is 46% higher than of those aged 30-39, which is in turn 5% higher than that of workers aged 40-49.

Second, our model suggests that the changes in wages are skewed upwards. In particular, most of the wage changes are small and moderate, but there are a few that are big and positive. This is a direct consequence of Pareto learning. Third, the variance of the changes in log-wage depends on the current wage level (as determined by the worker's expected ability). To the extent that assignment is more important for workers with higher wages, our model points to a potential reason for the (conditional) transitory variance of wages being larger for high-earning workers. These implications can be tested directly, and if relevant, they help us better understand the earning process of the workers, and in turn, the consumption and saving decisions they make.

To assess the role of assignment on earning inequalities, ideally one would like to calculate the variance of the earnings directly. But direct calculation of the earning variances is complicated here because of the atoms at the bottom of the wage distribution. As an approximation, we study the approximate wage function defined in Equation (4). Recall that the approximate wage function differs from the wage function by $\alpha\eta_l/(\alpha+\gamma)$, where η_l is the cutoff expected ability of the workers matched to the worst firms. To the extent that we are interested in the wage dispersion at the right tail of the distribution, our results are less affected by this approximation.

Proposition 4 *Permanent and transitory earning variances are reflected by the followings.*

(i): The approximate permanent variance of log-wage is given by

$$Var_a[\log \widetilde{W}(a)] = k^2 Var_a[\log(a)].$$

(ii): The approximate transitory variance of log-wage is given by

$$\begin{aligned} & E_{\eta_n}[Var_{\eta_{n+1}}[\log \widetilde{W}(\eta_{n+1}) - E[\log \widetilde{W}(\eta_{n+1})|\eta_n]] \\ = \; & k^2 E_{\eta_n}[Var_{\eta_{n+1}}[\log(\eta_{n+1}) - E[\log(\eta_{n+1})|\eta_n]]. \end{aligned}$$

Proposition 4 implies that both earning inequality and earning instability increase as the importance of assignment increases. Proposition 4 states that both the permanent variance and the transitory variance are proportional to k^2. Since k reflects the importance of assignment, this suggests that as assignment becomes more important, both the permanent and the transitory variance of log-wage increase.

Our model gives a natural mechanism for how permanent earning inequality and earning instability can jointly increase. The reason for the increase in the permanent variance of log-wages is the same as in a static model of assignment: wage inequality increases when productivity difference (resulting from difference in expected ability) is amplified by assigning better workers to better technologies. The reason the increase in transitory variance is similar is that assignment magnifies wage changes that result from changes in expected ability. As assignment becomes more important, it leads to a larger difference in expected productivity from the same amount of change in expected ability. The larger changes in expected productivity, therefore, also lead to larger changes in wages. Since both the size of the wage changes and the wage distributions are related to the importance of assignment, our model can account for the joint movement of earning instability and earning inequalities.

Gottschalk and Moffitt (1994, 2011) find that both earning instability and earning inequality in the U.S. have increased significantly since the late 1970s. The increase in instability and inequality has risen for all skill groups, reflecting a change in the broad economic environment. The reasons for the rising earning instability have typically been thought to be separate from those that account for the widening earning inequality. Gottschalk and Moffitt (2011) mention that "[t]he causal literature on cross-sectional earnings inequality is not very informative about the causes of rising earnings instability, because it has largely focused on factors affecting inequality of permanent earnings, such as skill-biased technological change and increase international trade with low-wage producers."[8]

In contrast, our model points to assignment as a potential source for the rises in both earning instability and earning inequality. There is some evidence that the importance of assignment has increased recently, especially at the very top of the talent distribution; see, for example, Gabaix and Landier (2008) and Tervio (2008). Moreover, the growing importance of assignment in our model can be thought of as a version of the skill-biased technological change. To see this, take two workers with expected ability $\eta_H > \eta_L$ respectively. The ratio of approximate wages for these two

[8] One exception is Violante (2002), who shows that a model of vintage-human capital can explain both the rise in residual inequality and the increase in wage changes associated with job changes.

workers is then given by

$$\frac{\widetilde{W}(\eta_H)}{\widetilde{W}(\eta_L)} = (\frac{\eta_H}{\eta_L})^k.$$

This ratio increases as assignment becomes more important (k increases), so the growing importance of assignment favors the high-ability workers more than the low-ability ones.

Finally, a key feature of the model is that the wage distribution and wage dynamics are connected. Our last result shows that the importance of assignment also increases the rate of wage growth.

Proposition 5 *Wage growth is larger when the dispersion of technology is larger (so that assignment is more important). For a cohort-n worker with expected ability* $\eta > \frac{\alpha+n}{\alpha+n-1}\eta_l$,

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

The reason for the result is the same as for why the expected wage increases over time. A higher rate of wage growth follows from a higher rate of productivity growth. As assignment becomes more important, the average productivity gain from better assignment is larger. Recall that when there is no assignment, the expected wage does not change.

To the extent that assignment is more important for more-educated workers, our model predicts that both the residual wage inequality and the average rate of wage growth are higher for more-educated workers. The empirical evidence for both appears to be supportive. It is well known that both the average rate of wage growth and the residual wage inequality increase with years of education. For example, Rubinstein and Weiss (2006) show that wage increases faster for more-educated workers using three data sources: CPS-ORG, PSID, and NLSY. In fact, the differences in wage growth rates by educational attainment are sizeable. For example, Rubinstein and Weiss (2006) find that workers with an advanced degree experience annual wage growth of 7.7% in their first ten years of their labor market experience, compared to 3.9% for high school dropouts. In addition, Lemieux (2006) shows that for all experience levels, the variance of wages increases with education level. Again, the difference is sizable. Lemieux (2006) finds that for workers with less than 10 years of experience, the variance of wages for postgraduates is almost four times as large as that for high school dropouts.

4 Conclusion

This paper develops a model that combines learning and assignment to study wage distribution and wage dynamics. The predictions from our model match with many of the empirical regularities. Moreover, our model shows that wage dynamics at individual levels is connected with the aggregate wage distributions. A key feature of our model is its tractability: we obtain explicit formula of the wage distributions of all cohorts and the distributions of wage changes at individual level. Another feature of our model is to offer a natural framework for decomposing the variance of earnings into a permanent part and a transitory part. These features allows us to study the joint implication of assignment on wage dynamics, inequality, and instability.

With the explicit formula on the wage distributions, a natural next step is to calibrate the model. In particular, it is possible to study an extension of our model by incorporating human capital accumulation. If all workers accumulate their human capital at a constant rate, we again have explicit formula of the right tails of the wage distributions of each cohort and of the economy: they are 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.

References

[1] Abowd, J., Kramarz, F., Pérez-Duarte, S, and I. Schmutte, "A Formal Test of Assortative Matching in the Labor Market," NBER Working Paper No. 15546, (2009).

[2] Barbarino, A., and B. Jovanovic, "Shakeouts and Market Crashes," *International Economic Review*, 48 (2007), 385-420.

[3] Beaudry, P., and R. DiNardo, "The Effect of Implicit Contracts on the Movement of Wages over the Business Cycle: Evidence from Micro Data," *Journal of Political Economy*, 99 (1991), 665-688.

[4] Becker, G., Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, 2nd Edition. Chicago and London: University of Chicago Press, (1975).

[5] Ben-Porath, Y. "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75 (1967), 352-365.

[6] Bernhardt, D. "Strategic Promotion and Compensation," *Review of Economic Studies* 62 (1995), 315-339.

[7] Chiappori, A, Salanie B, and J. Valentin "Early Starters Versus Late Beginners," *Journal of Political Economy*, 107 (1999), 731-760.

[8] Demougin, D., and A. Siow, "Careers in Ongoing Hierarchies," *American Economic Review*, 84 (1994), 1261-1277.

[9] DeGroot, M. Optimal Statistical Decisions, 1st Edition. McGraw-Hill, (1970).

[10] Farber, H., and R. Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 111 (1996), 1007-1047.

[11] Felli, L., and C. Harris, "Learning, Wage Dynamics and Firm-Specific Human Capital," *Journal of Political Economy*, 104 (1996), 838-868.

[12] Gabaix, X., and A. Landier, "Why has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, 123 (2008), 49-100.

[13] Gibbons, R., and M. Waldman, "A Theory of Wage and Promotion Dynamics inside Firms," *Quarterly Journal of Economics*, 114 (1999), 1321-58.

[14] Gibbons, R., and M. Waldman, "Careers in Organizations: Theory and Evidence," Chapter 36 in Volume 3B of O. Ashenfelter and D. Card (eds.), Handbook of Labor Economics, North Holland, (1999).

[15] Gibbons, R., L. Katz, T. Lemieux, and D. Parent, "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics,* 23, (2005), 681-723.

[16] Gottschalk, P., and R. Moffitt, "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity*, 25 (1994), 217-272.

[17] Gottschalk, P., and R. Moffitt, "Trends in the Transitory Variance of Male Earnings in the U.S., 1970-2004," (2011), Working Paper.

[18] Harris, M., and B. Holmstrom, "A Theory of Wage Dynamics," *Review of Economic Studies* 72 (1982) 315-333.

[19] Heckman, J., L. Lochner, and P. Todd, "Earnings Functions, Rates of Return and Treatments: The Mincer Equation and Beyond, " in Handbook of the Economics of Education, Vol. 1, North Holland (2006).

[20] Holmstrom, B., "Managerial Incentive Problem-A Dynamic Perspective," Republished in *Review of Economic Studies* 66 (1999) 169-182.

[21] Jovanovic, B., "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87 (1979) 972-990.

[22] Lemieux, T, "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill," *American Economic Review*, 96 (2006), 461-498.

[23] Li, J., "Learning in the Labor Market," PhD Thesis, M.I.T. 2007.

[24] Lydall, H., The Structure of Earning, London: Oxford Press, (1968).

[25] MacDonald, G., "A Market Equilibrium theory of Job Assignment and Sequential Accumulation of Information," *American Economic Review*, 72 (1982) 1038-1055.

[26] MacLeod, B., and J. Malcomson, "Reputation and Hierarchy in Dynamic Models of Employment," *Journal of Political Economy*, 96 (1988) 832-854.

[27] Mincer, J., Schooling, Experiences, and Earnings, New York: Columbia University for National Bureau of Economic Research, (1974).

[28] Murphy, K., and F. Welch, "Empirical Age-Earnings Profiles," *Journal of Labor Economics* 8 (1990) 202-229.

[29] Neal, D., and S. Rosen., "Theories of the Distribution of Earnings," in Handbook of Income Distributions, Vol. 1, North Holland (2000).

[30] Pareto, V., *Cours d'Economic Politique*, Vol. 2, Paris: F.Pichou (1897).

[31] Rubinstein, Y., and Y. Weiss, "Post Schooling Wage Growth: Investment, Search and Learning," in Handbook of the Economics of Education, Vol. 1, North Holland (2006).

[32] Sattinger, M., "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 31 (1993), 831-80.

[33] Tervio, M., "The Difference that CEOs Make: An Assignment Model Approach," *American Economic Review*, 98 (2008), 642-668.

[34] Violante, G., "Technological Acceleration, Skill Transferability and the Rise of Residual Inequality," *Quarterly Journal of Economics* 117 (2002) 297-338.

5 Appendix

5.1 Properties of Pareto Learning

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

$$\begin{aligned} \Pr(A \geq a_n | m, n) &= (\frac{a_n}{m})^{-(\alpha+n)} && \text{for } a_n \geq m; \\ &= 1 && \text{for } a_n < m. \end{aligned}$$

(A): Formula of expected ability

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

$$\begin{aligned} \eta_n(m) &= \int_m^\infty x \Pr(A = x | m, n) dx \\ &= \int_m^\infty x \frac{(\alpha+n) m^{\alpha+n}}{x^{\alpha+n+1}} dx \\ &= \frac{\alpha+n}{\alpha+n-1} m. \end{aligned}$$

(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

$$\begin{aligned} \Pr(M_n \leq m) &= \int_1^\infty \min\{1, (\frac{m}{x})^n\} \frac{\alpha}{x^{\alpha+1}} dx \\ &= 1 - \frac{n}{\alpha+n} m^{-\alpha}, \end{aligned}$$

where $\min\{1, (\frac{m}{x})^n\}$ 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 (3) links the maximum draw with the expected ability, we have that the distribution of expected ability of workers in cohort n satisfies

$$\begin{aligned} &\Pr(\Gamma_n \geq \eta_n) = \tfrac{n}{\alpha+n}(\tfrac{\alpha+n-1}{\alpha+n}\eta_n)^{-\alpha} && \text{for } \eta_n \geq \tfrac{\alpha+n}{\alpha+n-1}; \\ &\Pr(\Gamma_n \geq \eta_n) = 1 && \text{for } \eta_n < \tfrac{\alpha+n}{\alpha+n-1}. \end{aligned}$$

In other words, the distribution of the expected ability is a generalized Pareto distribution with parameters $(\frac{\alpha+n}{\alpha+n-1}, \alpha, \frac{\alpha}{\alpha+n})$.

(C): Formula of the conditional expected ability distribution

For a worker in cohort n with expected ability η_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 (3). For a worker with expected ability η_n, equation (3) implies that his maximal draw $m_n = \frac{\alpha+n-1}{\alpha+n}\eta_n$. Pareto learning (equation (2)) implies that the worker's conditional ability distribution (given expected ability η_n) is therefore $Pareto(m_n, \alpha+n)$. It follows that the conditional distribution of the maximum draw in $n+t$ satisfies, for $m_{n+t} \geq m_n$,

$$\begin{aligned}
& \Pr(M_{n+t} \leq m_{n+t} | M_n = m_n) \\
= & \int_{m_n}^{\infty} \max\{1, \left(\frac{m_{n+t}}{x}\right)^t\} \frac{(\alpha+n)m^{\alpha+n}}{x^{\alpha+n+1}} dx \\
= & 1 - \frac{t}{\alpha+n+t} (\frac{m_n}{m_{n+t}})^{\alpha+n},
\end{aligned}$$

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 $\eta_{n+t} > \frac{(\alpha+n-1)(\alpha+n+t)}{(\alpha+n)(\alpha+n+t-1)}\eta_n$,

$$\begin{aligned}
& \Pr(\Gamma_{n+t} \geq \eta_{n+t} | \Gamma_n = \eta_n) \\
= & \Pr(M_{n+t} \geq \frac{\alpha+n+t-1}{\alpha+n+t}\eta_{n+t} | M_n = \frac{\alpha+n-1}{\alpha+n}\eta_n) \\
= & \frac{t}{\alpha+n+t}\left(\frac{(\alpha+n-1)(\alpha+n+t)}{(\alpha+n)(\alpha+n+t-1)} \frac{\eta_n}{\eta_{n+t}}\right)^{\alpha+n}.
\end{aligned}$$

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 η_n, his conditional distribution of expected ability in the t periods in the future is a generalized Pareto distribution with parameter $(\frac{(\alpha+n-1)(\alpha+n+t)}{(\alpha+n)(\alpha+n+t-1)}\eta_n, \alpha+n, \frac{\alpha+n}{\alpha+n+t})$.

5.2 Proofs of Results

Proof of Theorem 1: 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

$$\begin{aligned} & s_1\eta_1 - W(\eta_1) + s_2\eta_2 - W(\eta_2) \\ < \quad & s_1\eta_2 - W(\eta_2) + s_2\eta_1 - W(\eta_1), \end{aligned}$$

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 $\eta \leq \eta_l$ is matched with a firm of $s = 1$. Condition (ii) of the equilibrium then implies that

$$W(\eta) = \eta \quad \text{ for } \eta \leq \eta_l.$$

Now consider the remaining firms with technology greater than 1 and the remaining workers with expected ability greater than η_l. 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

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

This expression and $W(\eta) = \eta$ for $\eta \leq \eta_l$ 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.

Proof of Proposition 1: The wage distributions follow from direct computation. To check $s(n)$ increases with n, note that

$$\begin{aligned}
\frac{d\log(s(n))}{dn} &= \frac{1}{n} - \frac{1}{\alpha+n} - \frac{\alpha}{\beta}\left(\frac{1}{\alpha-\beta+n} - \frac{1}{\alpha+n}\right) \\
&= \frac{\alpha(\alpha-\beta)}{n(\alpha+n)(\alpha-\beta+n)} \\
&> 0.
\end{aligned}$$

Proof of Corollary 2: For (i), recall that the expected ability distribution for each cohort satisfies

$$\begin{array}{ll}
\Pr(A \geq \eta_n) = \frac{n}{\alpha+n}\left(\frac{\alpha-\beta+n}{\alpha+n}\eta_n\right)^{-\frac{\alpha}{\beta}} & \text{for } \eta_n \geq \frac{\alpha+n}{\alpha-\beta+n}; \\
\Pr(A \geq \eta_n) = 1 & \text{for } \eta_n < \frac{\alpha+n}{\alpha-\beta+n}.
\end{array}$$

It can be checked that this generalized Pareto distribution is skewed to the right:

$$\eta_n^{1/2} < E[\eta_n].$$

Since the wage is increasing and convex in expected ability, we have

$$W(\eta_n^{1/2}) < W(E[\eta_n]) \leq E[W(\eta_n)].$$

(ii) follows from direct calculation.

For (iii), since $n > \alpha$, less than half of the workers are assigned to the bottom mass. Therefore, the median expected ability level $\eta_n^{1/2}$ satisfies

$$\frac{1}{2} = \Pr(x \geq \eta_n^{1/2}) = s(n)(\eta_n^{1/2})^{-\alpha},$$

where again

$$s(n) = \frac{n}{\alpha+n}(\frac{\alpha-1+n}{\alpha+n})^{-\alpha}.$$

Recall from Proposition 1 $s(n)$ increases with n. Therefore, the median expected ability level $\eta_n^{1/2} = (2s(n))^{\frac{1}{\alpha}}$ is increasing in n. Since the wage is monotone in the expected ability, the median wage is increasing with n. In addition, for any $q > \frac{\alpha}{\alpha+n}$, the worker is no longer assigned to the bottom firms. This proves (iii).

For (iv), since $q_h > \frac{\alpha}{\alpha+n}$, we have $1 - q_h = s(n)(\eta_n^{q_h})^{-\alpha}$. Therefore,

$$\eta_n^{q_h} = (\frac{s(n)}{1-q_h})^{\frac{1}{\alpha}}.$$

To calculate $\eta_n^{q_l}$, there are two possible cases. First, if $q_l > \frac{\alpha}{\alpha+n}$, then we have

$$\eta_n^{q_l} = (\frac{s(n)}{1-q_l})^{\frac{1}{\alpha}}.$$

In this case, we have

$$\frac{W(\eta_n^{q_h})}{W(\eta_n^{q_l})} = \frac{\theta\sigma^{\theta+1}\lambda^{\frac{\theta+1}{\theta}}(\big(\frac{s(n)}{1-q_h}\big)^{\frac{(\theta+1)}{\alpha\theta}}+1}{\theta\sigma^{\theta+1}\lambda^{\frac{\theta+1}{\theta}}(\big(\frac{s(n)}{1-q_l}\big)^{\frac{\beta(\theta+1)}{\alpha\theta}}+1},$$

where $\theta = \frac{\gamma}{\alpha}$. This expression increases with $s(n)$, which increases with n.

Second, if $q_l \leq \frac{\alpha}{\alpha+n}$, then we have

$$W(\eta_n^{q_l}) = \frac{\alpha+n}{\alpha-1+n} > (\frac{s(n)}{1-q_l})^{\frac{1}{\alpha}}.$$

Now if $q_l < \frac{\alpha}{\alpha+n+1}$, then

$$W(\eta_{n+1}^{q_l}) = \frac{\alpha+n+1}{\alpha-1+n+1} < W(\eta_n^{q_l}).$$

Since $W(\eta_{n+1}^{q_h}) > W(\eta_n^{q_h})$, we get

$$\frac{W(\eta_{n+1}^{q_h})}{W(\eta_{n+1}^{q_l})} > \frac{W(\eta_n^{q_h})}{W(\eta_n^{q_l})}.$$

Now if $q_l \geq \frac{\alpha}{\alpha+n+1}$, then we have

$$\begin{aligned}
\frac{W(\eta_n^{q_h})}{W(\eta_n^{q_l})} &< \frac{\theta\sigma^{\theta+1}\lambda^{\frac{\theta+1}{\theta}}(\left(\frac{s(n)}{1-q_h}\right)^{\frac{(\theta+1)}{\alpha\theta}}+1}{\theta\sigma^{\theta+1}\lambda^{\frac{\theta+1}{\theta}}(\left(\frac{s(n)}{1-q_l}\right)^{\frac{(\theta+1)}{\alpha\theta}}+1} \\
&< \frac{\theta\sigma^{\theta+1}\lambda^{\frac{\theta+1}{\theta}}(\left(\frac{s(n+1)}{1-q_h}\right)^{\frac{(\theta+1)}{\alpha\theta}}+1}{\theta\sigma^{\theta+1}\lambda^{\frac{\theta+1}{\theta}}(\left(\frac{s(n+1)}{1-q_l}\right)^{\frac{(\theta+1)}{\alpha\theta}}+1} \\
&= \frac{W(\eta_{n+1}^{q_h})}{W(\eta_{n+1}^{q_l})}.
\end{aligned}$$

This finishes the proof.

It is convenient to prove Proposition 2, Corollary 3, and Proposition 5 together.

Proof of Proposition 2, Corollary 3, and Proposition 5: The condition $\eta > \frac{\alpha+n}{\alpha+n-1}\eta_l$ guarantees that the expected ability of the worker in any future periods will be greater η_l. Therefore, we have $W(\Gamma) - \frac{\alpha}{\alpha+\gamma}\eta_l = (\frac{\eta}{\lambda})^k\frac{\sigma\lambda}{k}$. 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 $\Gamma_{n+t} = Pareto(C(t,n)\eta, \alpha+n, \frac{\alpha+n}{\alpha+n+t})$, we have

$$\begin{aligned}
& E[W(\Gamma_{n+t}) - \frac{\alpha}{\alpha+\gamma}\eta_l|\Gamma_n = \eta] \\
= & \frac{\alpha+n}{\alpha+n+t}[\frac{\sigma\lambda}{k}(\frac{\eta}{\lambda})^k C(t,n)^k] + \int_{C(t,n)\eta}^{\infty}\frac{\sigma\lambda}{k}(\frac{x}{\lambda})^k\frac{t(\alpha+n)}{\alpha+n+t}(C(t,n)\eta)^{\alpha+n}x^{-(\alpha+n+1)}dx \\
= & \frac{\alpha+n}{\alpha+n+t}\{[\frac{\sigma\lambda}{k}(\frac{\eta}{\lambda})^k C(t,n)^k] + \frac{\sigma\lambda}{k}C(t,n)^k(\frac{\eta}{\lambda})^k(\frac{t}{\alpha+n-k})\} \\
= & \frac{\alpha+n}{\alpha+n+t}(C(t,n))^k(1+\frac{t}{\alpha+n-k})(W(\eta) - \frac{\alpha}{\alpha+\gamma}\eta_l)) \\
= & B(t,n,k)(W(\eta) - \frac{\alpha}{\alpha+\gamma}\eta_l).
\end{aligned}$$

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

and

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

where the inequality uses $k = 1+\frac{\alpha}{\gamma} > 1$. Therefore, $\frac{d\log(B(t,n,k))}{dk} > 0$ and thus $B(t,n,k) > 1$. This proves Proposition 2.

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

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

so we have proved Proposition 5.

Next, note that

$$\frac{d\log(B(t,n,k))}{dt} = -\frac{1}{\alpha+n+t} + k(\frac{1}{\alpha+n+t} - \frac{1}{\alpha+n+t-1}) + \frac{1}{\alpha+n-k+t},$$

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 $\frac{dE[W(\Gamma_{n+t})|\Gamma_n=\eta]}{dt} > 0$ and we have proved (i) in Corollary 3.

Finally, note that $\frac{d\log(B(t,n,1))}{dn} = 0$, $\frac{d^2\log(B(0,n,x))}{dndx} = 0$, and

$$\begin{aligned} \frac{d^3\log(B(t,n,k))}{dkdtdn} &= \frac{d^3\log(B(t,n,k))}{dt^2dk} \\ &= \frac{-1}{(\alpha+n+t)^2} + \frac{1}{(\alpha+n+t-1)^2} - \frac{2}{(\alpha+n-k+t)^3} \\ &< 0 \end{aligned}$$

because $\frac{d^3\log(B(t,n,1))}{dt^2dk} = \frac{1-3(\alpha+n+t)}{(\alpha+n+t)^2(\alpha+n+t-1)^3} < 0$ and $\frac{d^4\log(B(t,n,k))}{dt^2dk^2} = \frac{-6}{(\alpha+n-k+t)^4} < 0$.

Therefore, we have

$$\begin{aligned}
\frac{d\log(B(t,n,k))}{dn} &= \frac{d\log(B(t,n,1))}{dn} + \int_1^k \frac{d^2\log(B(t,n,x))}{dndx}dx \\
&= \frac{d\log(B(t,n,1))}{dn} + \int_1^k (\frac{d^2\log(B(0,n,x))}{dndx} + \int_0^t \frac{d^3\log(B(y,n,x))}{dydndx}dy)dx \\
&< 0.
\end{aligned}$$

And this implies (ii) in Corollary 3.

Proof of Proposition 3: For (i), since the matching is monotone,

$$\begin{aligned}
& sgn(d(\Pr(W(\Gamma_{n+1}) > W(\Gamma_n)|\Gamma_n = \eta)/dn) \\
= \; & sgn(d(\Pr(\Gamma_{n+1}) > \Gamma_n|\Gamma_n = \eta)/dn) \\
= \; & sgn(d(\frac{1}{\alpha+n+1}\left(\frac{(\alpha+n)^2-1}{(\alpha+n)^2}\right)^{\alpha+n})/dn) \\
= \; & sgn(d(\log(\frac{1}{\alpha+n+1}\left(\frac{(\alpha+n)^2-1}{(\alpha+n)^2}\right)^{\alpha+n}))/dn) \\
= \; & sgn(\frac{3-(\alpha+n)}{(\alpha+n)^2-1} + \log\left(\frac{(\alpha+n)^2-1}{(\alpha+n)^2}\right)),
\end{aligned}$$

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 $3-(\alpha+n) \leq 0$ because $\alpha \geq 2$ and $n \geq 1$.

For (ii), consider the case in which $\Gamma_{n+1} < \Gamma_n$. In this case,

$$1 < \frac{W(\Gamma_n)}{W(\Gamma_{n+1})} < \frac{W(\Gamma_{n+1}) - \frac{\alpha+n}{\alpha+n-1}\eta_l}{W(\Gamma_n) - \frac{\alpha+n}{\alpha+n-1}\eta_l} = (\frac{\Gamma_n}{\Gamma_{n+1}})^k.$$

Taking logs, we get

$$\begin{aligned}
\log(\frac{W(\Gamma_n)}{W(\Gamma_{n+1})}) &< k\log(\frac{\Gamma_n}{\Gamma_{n+1}}) \\
&\leq k\log(1+\frac{1}{(\alpha+n)^2-1}) \\
&< \frac{k}{(\alpha+n)^2-1},
\end{aligned}$$

where the first weak inequality follows from Equation (1) with $t=1$, and the last inequality follows because log is a concave function.