Flattening Firms and Wage Distribution

Xin Jin*

University of South Florida

September 8, 2014

Abstract

This article studies the consequences of firm delayering on wages and the wage distribution inside firms. I consider a job-assignment model with asymmetric information and a slot constraint. The model predicts that more efficient firms are not necessarily larger than less efficient firms if firms are allowed to adjust their internal organizational structure through delayering. After delayering, wages at all levels increase and the wage distribution becomes more unequal. These predictions match a set of empirical findings in recent studies that are not well explained by existing theories.

Keywords: Delayering, asymmetric information, wage distribution, slot constraint

JEL: J31, M51

*Jin: Department of Economics, University of South Florida, Tampa, FL 33620, xjin@usf.edu.
I thank Michael Waldman, Julia Wulf, and Jan Zabojnik for their generous comments.

1 Introduction

Firms organize their workers into hierarchies to carry out production (Williamson, 1967; Calvo and Wellisz, 1979; Rosen, 1982). The hierarchical structure of firms is closely related to individuals' career and wage dynamics. In the past thirty years, firms became flatter, i.e., they got rid of some layers in their corporate hierarchy. This delayering trend is well documented in numerous studies using large-scale firm-level data sets from several developed countries (Colombo and Delmastro, 1999;2008; Rajan and Wulf, 2006; Caliendo et al., 2012). Widely identified and most convincing causes of this trend are the increased competition in the product market (Bloom et al., 2010; Guadalupe and Wulf, 2010), the improvement in corporate governance, and the advancement of information technology (Garicano, 2000; Bresnahan et al., 2002).

While the causes of this delayering trend are extensively studied, the consequences of this change, especially the impact on individuals' wages and the wage distribution within a firm, are not well explored.[1] In this essay, I build a model to explain two empirical regularities found in the recent literature. First, after firms delayer, wages at all levels increase (Bauer and Bender, 2001; Rajan and Wulf, 2006; Caliendo et al., 2012). Second, after delayering, the wage distribution becomes more unequal (Bauer and Bender, 2001).

To explain these wage patterns associated with delayering, I consider a job-assignment model with slot constraints and asymmetric information. In this set

[1] Numerous studies in the management and human resource literature explore the relationship between firm delayering and subsequent firm performance. The conclusion is mixed. For example, Carzo and Yanouzas (1969) find that tall organizations are more profitable, while Shaw and Schneier (1993), Cristini et al. (2003), and Kuhn (2011) find that the opposite is true.

up, workers compete for a single position at the upper level. Outside firms do not observe workers' outputs but can make inferences about workers' abilities using promotions as signals (Waldman, 1984a). In addition, I assume that workers sometimes leave their positions for exogenous reasons following Greenwald (1986). The main intuition in this model is that, holding firm size constant, the contestants' pool is larger in a flatter firm than in the firm with more layers. Therefore, winning a promotion in a flatter firm sends a more positive signal about the winner's ability.[2] In addition, losing a promotion in a larger contest sends a more positive signal about losers' abilities as well. This is because losing a larger contest does not necessarily mean that a worker is incompetent but rather that she is not the best among many workers. Thus, after delayering, wages at all levels go up. In addition, because the market expectation about the winner's ability goes up faster than the market expectation about losers' abilities, the wage gap between the winner and the losers widens after delayering.

Three key features of the model are essential to the resulting wage patterns. First, there is asymmetric learning among firms about workers' abilities. With asymmetric learning, workers' wages are determined by their "market value", which is based on their job assignments. Second, the upper level positions are characterized by a slot constraint. Without a slot constraint, after delayering, there would be an upward adjustment in the expectation about the workers' ability at the low level job and a non-upward adjustment in the expectation about the workers' ability at the upper-level job so that the wage distribution would become more equal.[3] Third, due

[2]This argument is related to discussions in Prendergast (1999) and Waldman (2013).

[3]For example, suppose there are three types of workers with high, regular, or low ability working for a three-layer firm. The high ability individual is at the top, the regular ability individual is in the middle, and the low ability individual is at the bottom. After delayering, since the regular worker

to the existence of exogenous job movers, my model alleviates the "winner's curse" problem that is common in models with asymmetric information (Milgrom and Oster, 1987). Without exogenous movers, if a worker is not promoted, the market would not want to offer a wage that is higher than the lowest expected productivity. If a worker's previous employer learns her ability perfectly after one period as assumed in Waldman (1984a) and Zabojnik and Bernhardt (1995), this lowest expected productivity does not change and so delayering would not affect the wage for low level workers.

This essay contributes to the literature in several different ways. First, it contributes to the delayering literature by exploring the effects of delayering on wage changes. Second, it contributes to the job assignment literature by considering how firms' organization structure affects wages. This essay also captures several empirical findings that are not well explained in the existing literature.

The organization of this essay is as follows. In Section 2, I review related literature. Section 3 discusses the model set up. In Section 4, I first analyze a model with two layers, and then compare the results to a three-layer model. Section 5 concludes.

is not good enough for the top level job, she would join the low ability worker on the low level job. As a result, the average ability at the top is unchanged while the average ability at the bottom rises. Suppose there is a continuum of workers' types, some middle level workers would join the top level job, and some middle level workers would join the bottom level job. As a result, the average ability at the top falls while the average ability at the bottom rises. Since workers' wages are largely attached to their expected ability (productivity), wages in the firm become more equal.

2 Related Literature

As I discussed in the introduction, most of the empirical literature on delayering has focused on the causes of delayering (Garicano, 2000; Bresnahan et al., 2002; Bloom et al., 2010; Guadalupe and Wulf, 2010). Several recent studies have documented some new wage patterns related to firm delayering. For example, using data from more than 300 US firms from 1986-1998, Rajan and Wulf (2006) find that after delayering, the division managers are paid more in salary and bonus after controlling for firm size and firm fixed effects. Internationally, using a comprehensive sample of French manufacturing firms, Caliendo et al. (2012) find that after delayering, wages at all levels of the firm increases. In an earlier study using a nationally representative linked employer-employee panel dataset from Germany, Bauer and Bender (2001) find that average wages increase after firms delayer. The wage distribution also becomes more unequal.

Two existing theories have addressed the relationship between delayering and wage distributions within firms. The first stream of literature is based on the command-and-control argument. Qian (1994) extends Calvo and Wellisz (1979) by endogenizing the number of layers in a hierarchy. The main mechanism is that since the entrepreneur's attention is limited, the further down a worker is in the hierarchy, the looser the control is and the lower the worker's effort is. Since the optimal number of layers decreases as the capital stock shrinks, workers who remain in the same position relative to the bottom receive a higher wage due to an increase in control with a shorter chain-of-command and workers who remain in the same position relative to the top receive a lower wage due to a decrease in monitoring with a larger span. That is, delayering is associated with a wage increase at the lower level and a wage decrease at the top. This prediction is inconsistent with the recent findings in the delayering literature and the more established stylized fact that the CEO-to-average-wage ratio has increased dramatically over the past thirty years.[4]

The other strand of literature takes the knowledge hierarchy approach (Garicano, 2000). Caliendo and Rossi-Hansberg (2012) study a model in which firms eliminate layers in response to negative demand shocks. Since the total knowledge for production is unchanged, as the number of layers decreases, the knowledge and thus wages in all pre-existing layers rise. In another study, Garicano and Rossi-Hansberg (2006) consider a model with homogeneous firms and heterogeneous workers. In their model, as the cost of acquiring knowledge decreases, fewer layers are preferred and managers acquire more knowledge. Since the knowledge increases more at the top, the overall wage inequality increases. While Caliendo and Rossi-Hansberg (2012) provide an explanation for why wages at all levels increase after delayering, they do not explain why wage inequality rises after delayering. On the other hand, Garicano and Rossi-Hansberg (2006) explain why wage inequality rises after delayering, but they do not explain why wages at all levels increase after delayering. My analysis captures both wage patterns associated with delayering under a single theoretical framework.

My model is built on the promotion-as-signal approach found initially in Waldman (1984a).[5] Using this approach, Waldman (1984a) and various extensions cap-

[4]For example, Murphy and Zabojnik (2004) find that from 1970 to 2000, the ratio of CEO cash compensation to average pay for production workers increased from 25 to 2000.

[5]The promotion-as-signal approach assumes asymmetric learning in the labor market which means a worker's current employer knows more about the worker's true ability than outside firms do. A competing modeling framework is to assume symmetric learning where all firms have the same information about a worker's type (Harris and Holmstrom, 1982; Gibbons and Waldman, 1999). Although both frameworks can explain various stylized facts concerning wage and promotion dynamics inside firms, many recent empirical studies have found evidence in favor of the asymmetric

ture many stylized facts about wage and promotion dynamics, such as large wage increases upon promotion (Bernhardt, 1995) and the wage-and-firm-size effect (Zabojnik and Bernhardt, 2001), etc. This paper adds to the promotion-as-signal literature by looking at the effect of delayering on wages.

This model's set up is closely related to Zabojnik and Bernhardt (2001). Zabojnik and Bernhardt (2001) consider a model with slot constraints to explore the firm-wage-size relation. The main focus in their model is on how promotion can induce optimal human capital investment. In this paper, on the other hand, I explore the implications of slot constraints and firms' hierarchical structural changes on wage distributions. I also show that, different from the prediction in Zabojnik and Bernhardt (2001), more efficient firms are not necessarily larger when we take into account the internal structure of the firm.

3 The Model

In this section I set up a two-period model. There are $F(>2)$ identical firms in the market. Each firm hires n risk-neutral young workers in period 1 (and firms can choose to vary their sizes in period 2). Different jobs in a firm have different production efficiencies, denoted by V. The production efficiency translates workers' ability into output. Workers are ex ante identical and have two-period careers. I refer to an individual in her first work period as young, and those who are in their second work period as old. Worker i's ability, θ_i, is drawn from a uniform distribution on $[\theta_L, \theta_H]$.

learning framework (Pinkston, 2009; DeVaro and Waldman, 2012; Kahn, 2013). I thus adopt the asymmetric learning framework in this essay.

An individual with ability θ who is assigned to a job with production efficiency V produces $s_t V\theta$ units of output in period t. $s_t = S > 1$ for an old worker who remains at her previous period's employer and $s_t = 1$ for a young worker or for an old worker who just starts to work for a new firm. s_t thus captures firm specific human capital. The total production at each firm is the sum of each worker's output. I refer to a worker's previous period's employer as the incumbent firm and all other firms as outside firms.

I assume there is over-supply of labor in the economy meaning that there are more than $n \cdot F$ workers. All the workers who are not hired by the firms stay self-employed. Following Waldman and Zax (2013), I assume that there is learning-by-doing in self-employment: a worker in her first period of self-employment produces $\bar{U}_1$ and a worker in her second period of self-employment produces $\bar{U}_2$, where $\bar{U}_2 > \bar{U}_1$. If a worker works for a firm in period 1 and becomes self-employed in period 2, she can only produce $\bar{U}_1$ in period 2. I assume $SV\theta_L > \bar{U}_1$, which means working for a firm in period 2 is better than self-employment if a worker has worked for a firm in period 1. This condition guarantees that workers do not change from working for a firm to self-employment between two periods. Furthermore, I assume $\bar{U}_1 + \bar{U}_2 > (1+S)VE(\theta)$, where $E(\theta)$ is the unconditional mean of a worker's ability. This condition guarantees that a firm hires a finite number of workers. This is because if firms hire an infinite number of workers, the total expected productivity and thus wage pay of each worker is $(1+S)VE(\theta)$. If this value is smaller than workers' total income from self-employment for two periods, $\bar{U}_1 + \bar{U}_2$, no worker would choose employment at a firm over self-employment. This regulates the firms not to hire an infinite number of workers.

Following Greenwald (1986), I assume a small probability, λ, that workers leave the firm for exogenous reasons. I consider the equilibrium behavior where $\lambda \to 0$. In addition, I assume that the firm-specific human capital is sufficiently large that an incumbent worker is always more productive than an outside worker.

The timing of the events is the following. At the beginning of period 1, nature assigns an ability type to each worker. Firms decide the optimal number of young workers to hire and offer wages accordingly. Workers choose the firm with the highest wage offer to work at. At the end of period 1, incumbent firms privately observe workers' outputs. At the beginning of period 2, incumbent firms update their beliefs about workers' types and one of the young workers is chosen to fill the upper level position if there is a vacancy at the upper level. Outside firms observe the incumbent firms' job assignment decisions and update their belief about workers' types. All firms then make wage offers simultaneously. Workers privately learn about their job-switching types and the exogenous movers depart. Workers then choose the firm with the highest wage offer to work at for period 2. If there are multiple firms offering the same highest wage, a worker chooses randomly among those highest-wage-offer firms but stays with her incumbent firm if her incumbent firm is one of the highest-wage-offer firms.

4 The Analysis

In equilibrium, firms assign workers to jobs and offer wages accordingly. I focus on the perfect Bayes-Nash equilibrium of this game. I first discuss equilibrium behavior in a two-layer firm and then I discuss and compare equilibrium behavior in a three-layer firm. Both models are solved by backward induction.

4.1 A two-layer model

Let us first consider the wage setting process in a two-layer firm. A two-layer firm consists of a CEO position (E) at the upper level and a number of laborer positions (L) at the lower level. Let $V^j, j \in \{E, L\}$ denote the production efficiency in job j. I assume $V^E > V^L$, which means the upper-level jobs have greater marginal returns to ability as in Sattinger (1975) and Rosen (1982).

In period 1, n young workers are hired into the laborer positions in each firm. In period 2, when the workers are old, one of them will be chosen to fill the CEO position (the CEO position remains unfilled in the first period). Since $V^E > V^L$, firms always have an incentive to assign the highest ability worker to the CEO position because this worker has the largest output increase if placed in the CEO position instead of a laborer position. In addition, due to the firm specific human capital, an incumbent firm can extract the highest rent from placing the most able worker in the CEO position.

In period 2, after observing a young worker's job assignment at her incumbent firm, outside firms form expectations about the worker's ability. Denote the expected ability of the promoted worker (i.e., the tournament winner) $\theta^E(n)$ and the expected ability of a non-promoted worker (i.e., a tournament loser) $\theta^L(n)$. Then $\theta^E(n) = E(\theta_i | i \text{ is the best among } n \text{ laborers})$ and $\theta^L(n) = E(\theta_i | i \text{ is not the best among } n \text{ laborers})$. Note that the expected ability is a function of the number of contestants in the promotion tournament only. $W^E(V, n)$ denotes the wage of the tournament winner who is promoted to the CEO position. $W^L(V, n)$ denotes the wage of the laborers who do not win the tournament. Proposition 1 describes the wages for the old workers in a two-layer firm. All proofs are reserved for the Appendix.

Proposition 1. *In equilibrium, an old worker's wage is equal to her expected productivity at an outside firm given her job assignment, i.e.,* $W^E(V,n) = V^L\theta^E(n)$ *and* $W^L(V,n) = V^L\theta^L(n)$*. The wage difference between the CEO and the laborers is* $\Delta W = W^E(V,n) - W^L(V,n)$*, which increases in the total number of workers in the laborer level n.*

Proposition 1 says that the old workers' wages are equal to their expected productivities at an outside firm, which are determined by their expected abilities as well as the production efficiency. Note that the wage paid to a tournament winner is evaluated at the laborer level. This is because firms are slot constrained at the upper level. In addition, due to the firm-specific human capital, an "insider" is more productive than an "outsider". Thus, firms do not replace an incumbent with an outside worker. Therefore, firms can only offer a wage that is consistent with assigning an outside worker to the lower level job.

Consider how the wage inequality, $\Delta W = V^L[\theta^E(n) - \theta^L(n)]$, changes with the total number of young workers hired into the laborer positions in period 1. When the pool of young workers (i.e. contestants) becomes larger, the winner's expected ability rises because the winner of a larger contest signals more strongly about her ability. On the other hand, the loser's expected ability also rises since not being the best among n laborers is a less bad outcome if n is larger.

To better illustrate this point, let us compare a two-person contest versus an n-person contest. Ex ante, the two groups of workers have the same expected ability. Ex post, since the expected ability of the winner from the two-person contest is above average, the loser's expected ability should be below average. That is, the "penalty" on losers' expected ability is fully absorbed by this one "loser". On the other hand, although the expected ability of the winner from the n-person contest is larger than that from the two-person contest, the penalty on losers' expected ability is averaged across many losers. The more the workers are in the promotion contest, the lower the penalty is on each losers' expected ability, and the closer the losers' expected ability is to the unconditional mean ability. In the extreme, when the size of the contestants' pool approaches infinity, the non-promoted workers' expected ability approaches the unconditional mean. Therefore, the expected ability of the losers from a larger contest is larger than the expected ability of the losers from a smaller contest.[6] When workers' ability is uniformly distributed, the expected ability for a tournament winner grows faster than the expected ability for a tournament loser as the size of the contestants' pool increases. To see this, note that $\theta^E(n) - \theta^L(n) = \frac{\theta_H - \theta_L}{2} \cdot \frac{1}{1+1/n}$. Thus, the wage gap between the winner and the losers widens monotonically with the size of the contestants' pool, i.e., $\partial W / \partial n > 0$.

Now, let us consider how firms choose the optimal number of young laborers to hire in period 1. Let n denote the number of young laborers. W_Y^L denote the first period's wage for young laborers, which is a function of n. The firms' problem is the following.

$$\underset{n}{Max} \quad nV^L E(\theta) - nW_Y^L(n) + [SV^E \theta^E(n) - W^E(V,n)] \tag{1}$$
$$+(n-1)[SV^L \theta^L(n) - W^L(V,n)]$$
$$s.t. \quad W_Y^L(n) + [\tfrac{1}{n} W^E(V,n) + \tfrac{n-1}{n} W^L(V,n)] \geq \bar{U}_1 + \bar{U}_2, \tag{2}$$

[6]Waldman (2013) makes a similar argument and he illustrates this argument by considering a binomial case where workers are either good or bad. See footnote 19 in Waldman (2013).

(2) is workers' participation constraint that a worker is better off choosing working for a firm for two periods than staying self-employed for two periods. That is, the overall lifetime income should be at least the amount of what the self-employment pays. In equilibrium, the participation constraint binds. The first-order condition to the above problem is

$$S(V^E - V^L)\frac{\theta_H - \theta_L}{(n+1)^2} = (\bar{U}_1 + \bar{U}_2) - (1+S)V^L E(\theta). \tag{3}$$

From (3), if we hold the production efficiency at the laborer level (V^L) constant but increase the production efficiency at the CEO level (V^E), the total number of young workers hired increases. The intuition is that as the CEO job becomes more efficient, firms can extract more rents from the CEO worker if this worker is more able. Firms achieve this sorting by increasing the contestants' pool. I summarize this property in Proposition 2.

Proposition 2. *The total number of young workers hired in period 1 increases with V^E, holding V^L constant.*

From Proposition 2, a firm with a more efficient production technology at the CEO level hires more young workers and the market expectation about the CEO's ability increases with the number of young workers competing in the promotion tournament. Thus, if the CEO job becomes more efficient, the firm grows holding the firm structure constant. As shown in Proposition 1 , when firms grow, wages at all levels go up and the wage distribution becomes more unequal. I summarize these relationships in Proposition 3.

Proposition 3. *Consider two firms with two layers. Suppose* $V_1^E > V_2^E, V_1^L = V_2^L$, *then* $n_1 > n_2, W_1^j > W_2^j, j \in \{E, L\}$, *and* $\Delta W_1 > \Delta W_2$.

The wage patterns described in Proposition 3 are driven by the difference in the production technology. The two-period model shows that firms will adjust their sizes in response to technology changes while firm structure is held constant. That is, firms with more efficient production technology at the CEO level are larger. However, Rajan and Wulf (2006) find that firm sizes are relatively stable over time in their dataset although there has been substantial technology advancements during their sample period. This suggests that firms adjust to technology changes through changes in firm structure while holding firm size constant. Before I discuss how firms' structure changes with technology, let me first consider a three-layer model in the following section.

4.2 A three-layer model

In a three-layer firm, the CEO (E) occupies the top level, managers (M) stay on the middle management level, and laborers (L) take up the lower level. Each manager heads a division with a team of laborers. All managers report directly to the CEO. The production efficiency at each job level is denoted by $\hat{V}^j, j \in \{E, M, L\}$. Similar to the two layer case, an upper level job is more efficient in utilizing workers' ability than a lower level job, i.e., $\hat{V}^E > \hat{V}^M > \hat{V}^L$. The production function is the same as before. An individual with ability θ who is assigned to job j produces $S\hat{V}^j\theta$ units of output if she is old and works for her incumbent firm; $\hat{V}^j\theta$ otherwise.

Firms live for two periods. In period 1, $\hat{m}$ young workers are hired into the management positions in each firm and $\hat{n}$ young workers are hired into the laborer positions in each management division, i.e., $\hat{m} \cdot (\hat{n}+1)$ young workers are hired in each firm in period 1. In period 2, when the workers are old, one of the managers will be chosen to fill the CEO position. After the tournament at the management level, the laborers in the promoted manager's division compete for the management vacancy. I assume that the CEO job requires some manager-specific human capital that only those who have worked as a manager can work as a CEO. I also assume that there is some division-specific human capital that only the laborers in a particular division can work as the manager in that division.[7]

The firm's problem at the management level in the three-layer model is very similar to the one in the two-layer model. In period 2, after observing a manager's job assignment decisions made by her incumbent firm, outside firms form expectations about the manager's ability. I assume that the management production efficiency is not very different from the laborer production efficiency so that there are no demotions.[8] Thus, a manager is either promoted to be the CEO or remains on the management level. Denote the expected ability of the promoted manager (i.e., the CEO) $\theta_P^E(\hat{m})$ and the expected ability of a non-promoted manager $\theta_N^M(\hat{m})$. Then $\theta_P^E(\hat{m}) = E(\theta_i | i \text{ is the best among } \hat{m} \text{ managers})$ and $\theta_N^M(\hat{m}) = E(\theta_i | i \text{ is not the best among } \hat{m} \text{ managers})$.

At the laborer level, only one laborer in the division with a management va-

[7]This assumption rules out cross-division promotions. Otherwise there are more laborers competing for a single management position than managers competing for the CEO position such that the wage of the winning laborer exceeds the wage of the CEO, which is counterfactual. See Friebel and Raith (2013) for a theoretical analysis where cross-division promotions are allowed.

[8]It is possible that firms would want to demote a manager and replace her with a more competent laborer after their true abilities are revealed. However, in doing so, the incumbent firm would send a strong signal to the market about the laborer's ability such that the productivity gain could not compensate for the increase in the wage bill. Thus, if the manager-level production efficiency is not sufficiently different from the laborer-level production efficiency, demotions will not occur.

cancy will be promoted to become a manager. Outside firms only observe workers' job assignments but they do not observe in which division a management vacancy becomes available. Since all firms are identical, firms know the number of equilibrium young managers that other firms hire. Thus, if a laborer is promoted to be a manager, the expected ability of the promoted laborer is $\frac{1}{\hat{m}}\theta_P^M(\hat{n}) = \frac{1}{\hat{m}}E(\theta_i|i$ is the best among $\hat{n}$ laborers). The expected ability of the non-promoted laborer is $\frac{1}{\hat{m}}\theta_N^L(\hat{n}) + (1-\frac{1}{\hat{m}})E(\theta) = \frac{1}{\hat{m}}E(\theta_i|i$ is not the best among $\hat{n}$ laborers$) + (1-\frac{1}{\hat{m}})E(\theta)$.

Now let us consider the second-period wages in a three-layer model. $\hat{W}_P^E(\hat{V},\hat{m})$ is the wage for a CEO who is promoted from a management position. $\hat{W}_N^M(\hat{V},\hat{m})$ is the wage for a manager who is not promoted to be a CEO from a management position. $\hat{W}_P^M(\hat{V},\hat{m},\hat{n})$ is the wage for a manager who is just promoted to be a manger from a laborer position. $\hat{W}_N^L(\hat{V},\hat{m},\hat{n})$ is the wage for a laborer who is not promoted to be a manger from a laborer position. Proposition 4 describes the wages for the old workers in three-layer firms.

Proposition 4. *In equilibrium, an old worker's wage is equal to her expected productivity at an outside firm given her job assignment. At the management level, a promoted manager's wage is* $\hat{W}_P^E(\hat{V},\hat{m}) = \hat{V}^L\theta_P^E(\hat{m})$ *and a non-promoted manager's wage is* $\hat{W}_N^M(\hat{V},\hat{m}) = \hat{V}^L\theta_N^M(\hat{m})$*. At the laborer level, a promoted laborer's wage is* $\hat{W}_P^M(\hat{V},\hat{m},\hat{n}) = \hat{V}^L[\frac{1}{\hat{m}}\theta_P^M(\hat{n})]$*; a non-promoted laborer's wage is* $\hat{W}_N^L(\hat{V},\hat{m},\hat{n}) = \hat{V}^L[\frac{1}{\hat{m}}\theta_N^L(\hat{n}) + (1-\frac{1}{\hat{m}})E(\theta)]$.

I now consider the first-period problem in each firm. In period 1, at the management level, firms choose a wage for young managers and decide how many managers to hire into the management position. Let $\hat{m}$ denotes the number of young managers to hire. $\hat{W}_Y^M$ denote the first period's wage for young managers, which is a function of $\hat{m}$. The problem at the management level is the following.

$$\begin{aligned} \underset{\hat{m}}{Max} \quad & \hat{m}\hat{V}^M E(\theta) - \hat{m}\hat{W}_Y^M(\hat{m}) + [S\hat{V}^E\theta_P^E(\hat{m}) - \hat{W}_P^E(\hat{V},\hat{m})] \quad (4) \\ & +(\hat{m}-1)[S\hat{V}^M\theta_N^M(\hat{m}) - \hat{W}_N^M(\hat{V},\hat{m})] \\ s.t. \quad & \hat{W}_Y^M(\hat{m}) + [\tfrac{1}{\hat{m}}\hat{W}_P^E(\hat{V},\hat{m}) + \tfrac{\hat{m}-1}{\hat{m}}\hat{W}_N^M(\hat{V},\hat{m})] \geq \bar{U}_1 + \bar{U}_2 \quad (5) \end{aligned}$$

(5) is a worker's participation constraint. $\bar{U}_1 + \bar{U}_2$ is a worker's expected lifetime income if she stays self-employed. That is, the young managers' wage is such that she is indifferent between working or staying self-employed for two periods. The first-order condition that characterizes the problem in (4) is

$$S(\hat{V}^E - \hat{V}^M)\frac{\theta_H - \theta_L}{(\hat{m}+1)^2} = (\bar{U}_1 + \bar{U}_2) - (1+S)\hat{V}^M E(\theta). \quad (6)$$

Note that the total number of young managers to hire is independent of the number of laborers and is determined only by the production efficiencies at the CEO level and the management level.

The problem at the laborers' level is similar to the problem at the management level. The only difference is that there might be no vacancy in a laborer's division. Let $\hat{n}$ denote the number of young laborers to hire. $\hat{W}_Y^L(\hat{n})$ denotes the first period's wage for young laborers. The problem at the laborer level is the following.

$$\underset{\hat{n}}{Max} \quad \hat{n}\hat{V}^L E(\theta) - \hat{n}\hat{W}_Y^L(\hat{n}) + (1-\frac{1}{\hat{m}})[\hat{n}S\hat{V}^L E(\theta) - \hat{n}\hat{W}_N^L(\hat{V},\hat{m},\hat{n})] \tag{7}$$

$$+\frac{1}{\hat{m}}\{[S\hat{V}\theta_P^M(\hat{n}) - \hat{W}_P^M(\hat{V},\hat{m},\hat{n})] + (\hat{n}-1)[S\hat{V}^L\theta_N^L(\hat{n}) - \hat{W}_N^L(\hat{V},\hat{m},\hat{n})]\}$$

$$s.t. \quad \hat{W}_Y^L(\hat{n}) + (1-\frac{1}{\hat{m}})\hat{W}_N^L(\hat{V},\hat{m},\hat{n})$$

$$+\frac{1}{\hat{m}}[\frac{1}{\hat{n}}\hat{W}_P^M(\hat{V},\hat{m},\hat{n}) + \frac{\hat{n}-1}{\hat{n}}\hat{W}_N^L(\hat{V},\hat{m},\hat{n})] \geq \bar{U}_1 + \bar{U}_2 \tag{8}$$

Similar to the manager's problem, (8) is a worker's participation constraint. The first-order condition that characterizes the problem in (7) is

$$\frac{1}{\hat{m}}S(\hat{V}^M - \hat{V}^L)\frac{\theta_H - \theta_L}{(\hat{n}+1)^2} = (\bar{U}_1 + \bar{U}_2) - (1+S)\hat{V}^L E(\theta). \tag{9}$$

From (9), the more young managers there are, the fewer young laborers there are in each division. The intuition is that if there are many managers, the probability that a vacancy occurs is low. Thus, it is less likely that a firm can extract the rent by placing a more able worker into the management position. As a result, firms hire fewer laborers.

Compare the period-1's problem at the management level and at the laborer level. If the difference in terms of production efficiency between the CEO job and the manager job is larger than the difference between the manager job and the laborer job, the CEO would have a larger span-of-control than the managers. That is,

Proposition 5. *If* $\hat{V}^E - \hat{V}^M > \hat{V}^M - \hat{V}^L$, *then* $\hat{m} > \hat{n}$.

Proposition 5 says that upper-level managers have a larger span than lower-level managers. This is due to the mechanism that firms have the incentive to assign more able workers to more efficient jobs and the sorting is achieved through increasing the number of contestant in a promotion tournament.

4.3 Technology, delayering, and wages

Now, we are ready to analyze the relationship between delayering and wages. Let us first look at the relationship between technology changes, firm sizes, and firm structure. We know from the delayering literature that one of the main causes of firm delayering is technology advancement. Thus, I consider an exogenous technology shock that changes the production efficiency at the CEO level in a three-layer firm from $\hat{V}^E$ to V^E and the firm decides to restructure into a two-layer firm. Since Rajan and Wulf (2006) find that firms' sizes are relatively stable over the years, I assume that the firm now hires the same number of workers as before, i.e., $N = \hat{N}$. I have the following Proposition.

Proposition 6. *If firm size and the production efficiency at the laborer level are unchanged, the CEO's production efficiency must be higher after delayering, i.e. if* $\hat{N} = N, \hat{V}^L = V^L$*, then it must be that* $V^E > \hat{V}^E$.

The logic behind Proposition 6 is the following. If a firm only has two layers, all workers are led by the CEO. If a firm has three layers, the CEO only leads the managers. If the firm has the same number of workers after delayering, it must be that the CEO in the two-layer firm leads a larger production team, which means she is more efficient. Note that Proposition 6 considers within firm changes rather than a cross firm comparison.

Now, let us consider delayering and wage changes. If we compare the non-promoted laborers' wage, $\hat{W}_N^L$, in the three-layer model to that in the two-layer model, W^L, we can see that the laborers' wage is higher in a two-layer firm, holding firm size and the production efficiency at the laborer level constant. The reason is that the two-layer firm has a larger contestant pool such that the expected ability of a loser is higher than the expected ability of a loser in a smaller contest. Since the production efficiency at the laborer level is unchanged, the flatter firm pays a higher non-promotion wage to the laborers. Similar argument applies to the wages paid to the CEOs in the two-layer firm and the three-layer firm. That is, after delayering, the CEO and the laborers are paid more because of an upward adjustment in their expected abilities.

With regard to the wage difference between the CEO and the laborer, when the market expectation about the non-promoted laborers' ability increases, the market expectation about the CEO's ability also increases. With a uniform ability distribution, the difference between these two expectations becomes larger as the number of contestants gets larger. That is, after delayering, the wage distribution within a firm becomes more unequal. I summarize the above argument in Proposition 7.

Proposition 7. *Suppose there is a technology change that results in a higher CEO production efficiency, i.e.,* $V^E > \hat{V}^E$*, and delayering. If firm size and the production efficiency at the laborer level are unchanged, both the laborers and the CEO's wages increase, i.e., if* $\hat{N} = N, \hat{V}^L = V^L$*, then* $W^E > \hat{W}_P^E, W^L > \hat{W}_N^L$*. In addition,* $\Delta W = W^E - W^L > \Delta \hat{W} = \hat{W}_P^E - \hat{W}_N^L$

The predictions in Proposition 7 capture the empirical findings in the delayering literature that when firms become flatter, wages at all level go up (Bauer and Bender, 2001; Rajan and Wulf, 2006; Caliendo et al., 2012) and wage inequality increases (Bauer and Bender, 2001).

In Zabojnik and Bernhardt (2001), more efficient firms are larger given that all firms have the same number of layers. From Proposition 6, a more efficient firm (measured by the production efficiency at the CEO level) is not necessarily larger than a less-efficient firm if we take into account firm restructuring. This result is consistent with the empirical finding in Rajan and Wulf (2006) that firms have fewer layers without much changes in firm size over time.

5 Conclusion

This paper develops a model with asymmetric learning and slot constraints to explore the relationship between firm delayering and workers' wages. It contributes to the literature in multiple ways. First, it contributes to the delayering literature by exploring the consequences of delayering on wages while most of the delayering literature focuses on the causes of the delayering trend. Second, it contributes to the job assignment literature by considering how firms' organization structure affects wages. Third, this model captures several empirical findings that are not well explained in the existing literature. My model shows that after delayering, workers' wages at all levels increase because they are now participating in a larger contest and thus their expected abilities are higher. In addition, since the workers' wage at the top increases faster than the wage increase at the bottom, the wage distribution becomes more unequal after delayering.

6 References

Bauer, T. K., & Bender, S. (2001). Flexible Work Systems and the Structure of Wages: Evidence from Matched Employer-Employee Data, IZA working paper.

Bernhardt, D. (1995). Strategic promotion and compensation. The Review of Economic Studies, 62(2), 315-339.

Bloom, N., Sadun, R., & Van Reenen, J. (2010). Does product market competition lead firms to decentralize? The American Economic Review, , 434-438.

Bresnahan, T. F., Brynjolfsson, E., & Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. The Quarterly Journal of Economics, 117(1), 339-376.

Caliendo, L., Monte, F., & Rossi-Hansberg, E. (2012). The Anatomy of French Production Hierarchies, Working Paper.

Caliendo, L., & Rossi-Hansberg, E. (2012). The impact of trade on organization and productivity. The Quarterly Journal of Economics, 127(3), 1393-1467.

Calvo, G. A., & Wellisz, S. (1979). Hierarchy, ability, and income distribution. The Journal of Political Economy, 87(5), 991.

Colombo, M. G., & Delmastro, M. (2008). The economics of organizational design: Theoretical insights and empirical evidence. Houndmills, UK: Palgrave MacMillan.

Colombo, M. G., & Delmastro, M. (1999). Some stylized facts on organization and its evolution. Journal of Economic Behavior & Organization, 40(3), 255-274.

DeVaro, J., & Waldman, M. (2012). The signaling role of promotions: Further theory and empirical evidence. Journal of Labor Economics, 30(1), 91-147.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. The Journal of Political Economy, 108(5), 874-904.

Garicano, L., & Rossi-Hansberg, E. (2006). Organization and inequality in a knowledge economy. Quarterly Journal of Economics, 121(4), 1383-1435.

Gibbons, R., & Waldman, M. (1999). A theory of wage and promotion dynamics inside firms. Quarterly Journal of Economics, 114(4), 1321-1358.

Gibbons, R., & Waldman, M. (1999). Careers in organizations: Theory and evidence. In O. C. Ashenfelter, & D. Card (Eds.), Handbook of labor economics, vol.3 (Orley C. Ashenfelter and David Card ed., pp. 2373-2437). Amsterdam: North-Holland:

Greenwald, B. C. (1986). Adverse selection in the labour market. The Review of Economic Studies, 53(3), 325-347.

Guadalupe, M., & Wulf, J. (2010). The flattening firm and product market competition: The effect of trade liberalization on corporate hierarchies. American Economic Journal: Applied Economics, 2(4), 105-127.

Harris, M., & Holmstrom, B. (1982). A theory of wage dynamics. The Review of Economic Studies, 49(3), 315-333.

Kahn, L. B. (2013). Asymmetric information between employers. American Economic Journal: Applied Economics, 5(4), 165-205.

Lazear, E. P., & Rosen, S. (1981). Rank order tournaments as optimum labor contracts. Journal of Political Economy, 89(5), 841-864.

Milgrom, P., & Oster, S. (1987). Job discrimination, market forces, and the invisibility hypothesis. The Quarterly Journal of Economics, 102(3), 453-476.

Murphy, K. J., & Zabojnik, J. (2004). CEO pay and appointments: A market-based explanation for recent trends. The American Economic Review, 94(2), 192-

196.

Murphy, K., & Zabojnik, J. (2007). Managerial capital and the market for CEOs. Available at SSRN 984376

Pinkston, J. C. (2009). A model of asymmetric employer learning with testable implications. Review of Economic Studies, 76(1), 367-394.

Qian, Y. (1994). Incentives and loss of control in an optimal hierarchy. The Review of Economic Studies, 61(3), 527-544.

Rajan, R. G., & Wulf, J. (2006). The flattening firm: Evidence from panel data on the changing nature of corporate hierarchies. The Review of Economics and Statistics, 88(4), 759-773.

Rosen, S. (1982). Authority, control, and the distribution of earnings. The Bell Journal of Economics, 13(2), 311-323.

Rosen, S. (1986). Prizes and incentives in elimination tournaments. The American Economic Review, 76(4), 701-715.

Smeets, V., Waldman, M., & Warzynski, F. (2013). Performance, career dynamics, and span of control. Working Paper.

Waldman, M. (1984). Job assignments, signaling, and efficiency. The Rand Journal of Economics, 15(2), 255-267.

Waldman, M. (1984). Worker allocation, hierarchies and the wage distribution. The Review of Economic Studies, 51(1), 95-109.

Waldman, M. (2003). Ex ante versus ex post optimal promotion rules: The case of internal promotion. Economic Inquiry, 41(1), 27-41.

Waldman, M. (2012). Classic promotion tournaments versus market-based tournaments. International Journal of Industrial Organization, (0), Available online 21

March 2012.

Williamson, O. E. (1967). Hierarchical control and optimum firm size. The Journal of Political Economy, 75(2), 123-138.

Zabojnik, J., & Bernhardt, D. (2001). Corporate tournaments, human capital acquisition, and the firm Size—Wage relation. The Review of Economic Studies, 68(3), 693-716.

7 Mathematical Appendix

Proof of Proposition 1 and 4. Consider a production unit with production efficiency V employing n young workers. In period 2, outside firms form expectations about old workers' ability denoted as θ^e. Due to the existence of the exogenous movers, the market expectation is free of the winner's curse. Since there are multiple firms, the standard result is that the market value of the worker is at her expected productivity, which equals to $V\theta^e$. Since the worker can produce more at her incumbent firm, her first-period employer has an incentive to match her market value.$\|$

Proof of Proposition 2. From (3) , since the RHS is unchanged, when V^E goes up, n goes up. $\|$

Proof of Proposition 3. From *Proposition* 2, since $V_1^E > V_2^E, V_1^L = V_2^L$, then $n_1 > n_2$. Since $\{\theta_i\}, i \in 1, \ldots, n$, follows uniform distribution, $f(\theta) = 1/(\theta_H - \theta_L), F(\theta) = (\theta - \theta_L)/(\theta_H - \theta_L)$. Prob.($\theta_i$ is the largest among n)=$F^n(\theta)$. Thus, $E(\theta_i|\theta_i$ is the largest among n$) = \int_{\theta_L}^{\theta_H} \theta \cdot nf(\theta)F^{n-1}(\theta)d\theta = (n\theta_H + \theta_L)/(n+1)$, which increases in n. Since $n_1 > n_2$, $E(\theta_i|\theta_i$ is the largest among $n_1) > E(\theta_i|\theta_i$ is the largest among $n_2)$. Thus, $W_1^E > W_2^E$. From the law of total expectation, $E(\theta) = E(\theta|\theta$ is not the largest among n)$\cdot P(\theta$ is not the largest among n)+$E(\theta|\theta$ is the largest among n)$\cdot P(\theta$ is the largest among n),$(\theta_H + \theta_L)/2 = E(\theta|\theta$ not largest)$(1 - 1/n) + (n\theta_H + \theta_L)/(n+1)(1/n)$.$E(\theta|\theta$ not the largest$) = [n(\theta_H + \theta_L) + 2\theta_L]/[2(n+1)]$, which also increases in n. Thus, the laborers' wage also increases, i.e., $W_1^L > W_2^L$. As for the wage difference, $\Delta W = (V^E - V^L)\theta^E(n) + V^L[\theta^E(n) - \theta^L(n)]$. We know $\theta^E(n)$ increases in n. $\theta^E(n) - \theta^L(n) = \frac{\theta_H - \theta_L}{2} \cdot \frac{n}{n+1}$, which also increases in n. Thus $\Delta W_1 > \Delta W_2$. $\|$

Proof of Proposition 5. Note that $\theta^E + (n-1)\theta^L = nE(\theta)$. Thus, the maximization problem at the two-layer firm can be re-write into

$$\underset{n}{Max} \quad nV^L E(\theta) + S(V^E - V^L)\theta^E(n) + SV_n^L nE(\theta) - n(\bar{U}_1 + \bar{U}_2).$$

I re-write as follows the three first-order conditions that pin down the equilibrium number of young workers in the three-layer firm and the two-layer firm.

$$S(V^E - V^L)\frac{\theta_H - \theta_L}{(N+1)^2} = (\bar{U}_1 + \bar{U}_2) - (1+S)V^L E(\theta) \quad (10)$$

$$S(\hat{V}^E - \hat{V}^M)\frac{\theta_H - \theta_L}{(\hat{m}+1)^2} = (\bar{U}_1 + \bar{U}_2) - (1+S)\hat{V}^M E(\theta) \quad (11)$$

$$\frac{1}{\hat{m}}S(\hat{V}^M - \hat{V}^L)\frac{\theta_H - \theta_L}{(\hat{n}+1)^2} = (\bar{U}_1 + \bar{U}_2) - (1+S)\hat{V}^L E(\theta) \quad (12)$$

From (12) over (11) , $\frac{\hat{V}^M - \hat{V}^L}{\hat{V}^E - \hat{V}^M} \cdot \frac{(\hat{m}+1)^2}{\hat{m}(\hat{n}+1)^2} = \frac{(\bar{U}_1+\bar{U}_2)-(1+S)\hat{V}^L E(\theta)}{(\bar{U}_1+\bar{U}_2)-(1+S)\hat{V}^M E(\theta)}$. Then the RHS>1. Suppose $\hat{n} > \hat{m}$. Since $\frac{\hat{V}^M - \hat{V}^L}{\hat{V}^E - \hat{V}^M} < 1, \frac{(\hat{m}+1)^2}{\hat{m}(\hat{n}+1)^2} < 1,$the $LHS < 1$. Contradiction. Thus, $\hat{n} < \hat{m}$.‖

Proof of Proposition 6. From (11),

$$\begin{aligned} S[(\hat{V}^E - \hat{V}^M) &- (\hat{V}^M - \hat{V}^L)]\frac{\theta_H - \theta_L}{(\hat{m}+1)^2} \quad (13) \\ &= (\bar{U}_1 + \bar{U}_2) - (1+S)\hat{V}^M E(\theta) - S(\hat{V}^M - \hat{V}^L)\frac{\theta_H - \theta_L}{(\hat{m}+1)^2}. \end{aligned}$$

From (10) over (14), we have

$$\frac{V^E - V^L}{\hat{V}^E - \hat{V}^L}\frac{(\hat{m}+1)^2}{(N+1)^2} = \frac{(\bar{U}_1 + \bar{U}_2) - (1+S)V^L E(\theta)}{(\bar{U}_1 + \bar{U}_2) - (1+S)\hat{V}^M E(\theta) - S(\hat{V}^M - \hat{V}^L)\frac{\theta_H - \theta_L}{(\hat{m}+1)^2}} > 1.$$

Since $\frac{(\hat{m}+1)^2}{(N+1)^2} < 1$, $\frac{V^E - V^L}{\hat{V}^E - \hat{V}^L} > 1, i.e., V^E > \hat{V}^E$. The CEO position at the flatter firm is more efficient. ‖

Proof of Proposition 7. $W^E = V^L \theta^E(N), W^L = V^L \theta^L(N), \hat{W}_P^E = \hat{V}^L \hat{\theta}_P^E(\hat{m}), \hat{W}_N^L = \hat{V}^L[\frac{1}{\hat{m}}\hat{\theta}_N^L(\hat{n}) + (1-\frac{1}{\hat{m}})E(\theta)]$. From Proposition 6,$V^E > \hat{V}^E$. From Proposition 3,$\theta^E(N) > \hat{\theta}_P^E(\hat{m})$. Let's consider $A := \theta^L(N)$ and $B := \frac{1}{\hat{m}}\hat{\theta}_N^L(\hat{n}) + (1-\frac{1}{\hat{m}})E(\theta)$. We know $N = \hat{N} = \hat{m}(\hat{n}+1)$. Then $A := \frac{\hat{m}(\hat{n}+1)(\theta_H+\theta_L)+2\theta_L}{2[\hat{m}(\hat{n}+1)+1]}$. $B := \frac{\hat{m}(\hat{n}+1)(\theta_H+\theta_L)+2\theta_L-(\theta_H+\theta_L)}{2[\hat{m}(\hat{n}+1)+1]-2}$. Since $A < \frac{\theta_H+\theta_L}{2}$, then $A > B$. That is, the expected ability of a non-promoted laborer in a two-layer firm is higher than the expected ability of a non-promoted laborer in a three-layer firm. Therefore, if $\hat{N} = N, \hat{V}^L = V^L$, then $W^E > \hat{W}_P^E, W^L > \hat{W}_N^L$.

From the proof of Proposition 3, the difference in the two expectations $E(\theta|\theta$ is the largest in $n) - E(\theta|\theta$ is not the largest in $n) = \frac{\theta_H-\theta_L}{2} \cdot \frac{n}{n+1}$ increases in n. $\Delta W = V^L[\theta^E(N) - \theta^L(N)] = V^L \cdot \frac{\theta_H-\theta_L}{2} \cdot \frac{N}{N+1}$. $\Delta\hat{W} = \hat{V}^L[\hat{\theta}_P^E(\hat{m}) - \frac{1}{\hat{m}}\hat{\theta}_N^L(\hat{n}) - \frac{\hat{m}-1}{\hat{m}}\frac{\theta_H+\theta_L}{2}]$. Let $C := \hat{\theta}_P^E(\hat{m}) - \frac{1}{\hat{m}}\hat{\theta}_N^L(\hat{n}) - \frac{\hat{m}-1}{\hat{m}}\frac{\theta_H+\theta_L}{2} < \frac{\hat{m}^2-\hat{m}+1}{\hat{m}(\hat{m}+1)} \cdot \frac{\theta_H-\theta_L}{2} = [1 - \frac{2\hat{m}-1}{\hat{m}(\hat{m}+1)}]\frac{\theta_H-\theta_L}{2}$. Let $D := \frac{N}{N+1} \cdot \frac{\theta_H-\theta_L}{2} = \frac{\hat{m}(\hat{n}+1)}{\hat{m}(\hat{n}+1)+1} \cdot \frac{\theta_H-\theta_L}{2} = [1 - \frac{1}{\hat{m}(\hat{n}+1)+1}]\frac{\theta_H-\theta_L}{2}$. We can see that $D > C$ as long as $\hat{m}, \hat{n} > 1$. Therefore, $\Delta W > \Delta\hat{W}$. ‖