14.662 Spring 2018

Lecture Note: Market Structure, Organizational Structure and Wage Structure

David Autor, MIT and NBER

February 28, 2018

1 Why is Information Technology Associated with Greater Worker Discretion and Teamwork (i.e., less 'Taylorism')?

We've generally talked about technological change in the abstract without putting much concrete content into that term. Many narrative accounts posit a dramatic change in the organization of work but do not tell a detailed story about how specifically work organization has changed nor why this change has been especially demanding of skilled labor. It would be helpful to have a substantive theory of the nature of recent technological or organizational change that didn't ultimately boil down to the assumed sign on a derivative or the magnitude of some (ultimately unmeasurable) elasticity.

Caroli and Van Reenan study—in a reduced form manner—the impact of work reorganization on skill demands. They provide evidence from British and French establishments that measures of organizational change—such as decentralization of authority—have a strong predictive power for the demand for skills at the establishment level, even after controlling for other determinants of the demand for skills, such as computers. Bresnahan, Brynjolfsson and Hitt (2000) provide U.S. based evidence in a similar vein. Neither of these studies get a great deal further than documenting correlations between reorganization and skill input, but the correlations are striking.

There is also a variety of theoretical and empirical work that offers explicit hypotheses on the link between computerization and changes in the content of work and organization of production. Some recent careful case studies are in my view among the richest material in this literature (which may suggest that economic theory would benefit if economic theorists would leave their desks more often). Ann Bartel, Casey Ichniowski, and Kathryn Shaw have performed detailed case studies of changing production technology in three industries: steel, medical devices, and valves. The 2007 *QJE* paper is specifically about the valve industry. I read their paper as making the following points:

1. Information technology has removed much of the mechanized, repetitive, rote components of production. Many repetitive tasks are now performed by machines that coordinate and monitor production. In the BIS valve study, computer-guided industrial lasers inspect completed valves for specification compliance with precision measured in microns. Previously, this time-consuming inspection step was done by hand. Similarly, the process of machine setup for new runs and the coordination of production as products move between machining phases have also been extensively computerized. This automation is feasible because computerized machining tools are far more flexible than older forms of capital. Much of the cumbersome reconfiguration of the assembly line needed for different products and batches is now handled directly by the machinery. All of these production changes are *process improvements*.

2. One consequence of the increased flexibility of the *process* is a change in the set of *products* produced. The valve firms studied by BIS, which are those that have continued producing in the U.S., have increasingly moved out of commodity production and into 'mass customization.' They exploit the flexibility of the new capital to do shorter runs of more complex

products. This in turn requires greater flexibility on the part of the workforce. But notice that commodity valve manufacturing, which is increasingly performed overseas, may not be undergoing similar changes in skill demand. [A 2005 paper by Chong Xiang of Purdue in the *ReStat* ("New Goods and Rising Skill Premium: An Empirical Investigation") presents detailed evidence that *new goods* are increasingly skill intensive. This suggests that product as well as process innovations may contribute to the evolution of skill demands.]

3. Workers are increasingly required to use abstract reasoning to manage production. Whereas workers used to spend much of their time in contact with physical materials, much of the 'work' is now performed from the 'control pulpit' of highly automated factories where the key task is to monitor multiple assembly lines via large displays. The challenge is to be alert to problems as they arise and make subtle adjustments that improve efficiency, reduce error, enhance quality. In this setting, a good decision can have enormous value added and a bad decision can destroy large amounts of capital.

4. Production work is frequently reorganized into teams where workers have responsibility for quality control and are required to solve problems and develop better ways to organize production. As observed by BIS, and Bresnahan, Brynjolfsson and Hitt (2002), and Caroli and van Reenan (2001) problem solving teams and other 'lean production techniques' are often paired with information technologies, suggesting that these Human newer Resource Practices (HRM) and IT are complements. (More on why this might be below...)

These observations are drawn from the manufacturing sector, which provides a declining share of employment in almost all advanced countries. How relevant are they to the service sector? Similar case studies for the service sector are not in abundance. One by Autor, Levy, Murnane (2002, ILRR) on the computerization of the back office check-processing function of a bank provides descriptive evidence that is consistent with the observations in Bartel, Ichniowski and Shaw (though the examples are not nearly as dramatic as those in manufacturing). In the bank studied by ALM, the introduction of Optical Character Recognition and electronic imaging of paper checks reduced the rote component of check processing for the 'exceptions processors,' reducing time spent on routine 'paper chase' activities and increasing time spent on problem solving and account management. Notably, for the 'check preparers' who perform the physical handling of the checks, there was little net change in skill demands—but there was a dramatic reduction in employment.

BIS specifically investigate four hypotheses:

1. New IT-enhanced machines improve production process efficiency. Setup time, run time, and inspection time fall after new IT-enhanced equipment in these stages is adopted.

2. New IT promotes product customization and innovation. New 3D-CAD technologies should directly affect the plant's capabilities of designing more customized valves, while other technologies that reduce setup time would also promote customization.

3. IT adoption may increase (or decrease) skill demand.

4. IT adoption may require new HRM practices.

The theoretical foundation for these hypotheses is BIS' observation that IT reduces setup time, which is otherwise particularly costly for customized, small batch jobs. This cost reduction differentially reduces the price of producing customized relative to commodity products. What happens when IT prices fall:

1. Firms purchase more IT

2. Production efficiency rises, setup time, run time and inspection time time fall.

3. Firms make a strategic move towards producing more customized products. This is due to fall in setup times.

4. Optimal skill demand changes, but the direction is ambiguous. If setup is the most skill-intensive task (as seems likely), when setup time falls, skill demand falls. But the move to more setup-intensive products exerts a countervailing force. Third, IT-based machinery increasingly displaces routine tasks, thereby raising skill content of labor input. This goes in the direction of increasing skill requirements.

5. Finally, new HRM practices may complement use of higher skill levels or new machinery, though a micro-foundation for this idea is not given.

2 Technology and the organization of production

A 2007 *JPE* paper by Wouter Dessein and Tanos Santos ("Adaptive Organizations") offers an ingenious hypothesis on the subtle linkages between information (or communication) technology and work organization. Quoting Dessein-Santos (p. 17):

> Adam Smith's famous observation that the "the division of labor is limited by the extent of the market," has been challenged by both the management literature as well as economists such as Becker and Murphy (1992). These two strands of the literature have emphasized that perhaps, more importantly, specialization is mainly constrained by the need to coordinate activities. In particular, a straightforward comparative static prediction in Becker and Murphy (1992 [QJE]) is that as coordination costs increase, one should see less specialization."

The 'Smithian' view would suggest that the expansion of markets via population growth, international trade, etc. would increase specialization by enlarging the extent of the market. The Becker-Murphy view would also suggest that the falling price of information technology would increase specialization by lowering communications/coordination costs. Yet, we seem to see a world where jobs are becoming broader and workers are increasingly asked to exercise discretion. What is going on?

One possibility is that casual empiricism is simply wrong: work *is* becoming more compartmentalized, perhaps not in advanced countries but in the world at large (though why *not* in advanced countries?). A second possibility is that the theory is wrong. The Dessein-Santos' paper suggests the latter.

2.1 Adaptation versus Coordination

Dessein-Santos consider a model of organizational design in which there are a number of tasks, n, to be performed. Performance of each task is optimized through two actions: (1) adaptation and (2) coordination.

Adaptation is the adjustment that should be made to a particular task in response to current conditions (which are assumed not entirely forecastable). For example, when cooking a Japanese dinner, the optimal quantity of water to add to the rice cooker depends in part on the current humidity and temperature as well as the season when the rice was harvested relative to the cooking date.

Coordination is the adjustment of other complementary tasks to the adaptations made on the primary task. For example, adding water to the rice cooker might mean that the meal will be delayed, in which case the vegetables should be sautéed several minutes later so they do not turn to mush.

One way to accomplish both objectives simultaneously (adaptation *and* coordination) is to assign one person to perform both tasks (*bundle*). The authors assume that *coordination* is costless (or very low cost) if one person performs all tasks (that is, there are no within-person coordination costs). In the limit, one can push coordination costs to zero by having all tasks performed by a single worker.

But the downside to bundling of tasks is the loss of the gains to *specialization*. Productive efficiency at each task is probably greatest if one chef cooks the rice and another steams the vegetables (picture a large restaurant). The authors therefore assume that the more tasks done by one worker, the lower her productivity at each. One can maximize the gains to specialization by having a separate person do *each* task. But coordination suffers, leading to lower quality or higher coordination costs to achieve the same quality. If tasks are highly specialized, there is a significant risk of coordination failures—information may be lost in coordinating among the workers performing complementary tasks. If the rice chef adjusts the water perfectly but neglects to inform the vegetable chef (or the information is lost with some probability) then we'll end up with better rice but perhaps a worse meal.

There is therefore a trade-off between allowing discretion to maximize adaptation (which comes at the cost of greater coordination failures) and reducing discretion to improve coordination (which comes at the cost of reduced adaptation). One can also bundle more tasks into a single job, which improves both adaptation and coordination. But this reduces gains from specialization.

Dessein-Santos map out two general overarching organizational strategies:

1. 'Ex ante coordination'—The goal here is to reduce coordination failures by *limiting discretion.*

So, the rice chef does not adapt the recipe by much to current conditions. This means the rice is perhaps of lower quality but it is well timed with the vegetables. In this case, it is clear that a firm will also want to subdivide tasks finely to take advantage of gains to specialization since there is little need to coordinate (i.e., the coordination is built-in since there is no discretion). This the so-called 'rigid' organization. In this type of organization, there will be little investment in communication (meetings, teams, etc.) because such communication is not needed. One could call this production mode 'Fordism.'

2. 'Ex post coordination'—Here, the firm allows workers discretion to adapt to local conditions but then must rely on high quality *ex post* communications to coordinate among adapted tasks. The firm will therefore want to invest in high quality communications to make ex post coordination work. Notice also that firms may want to use *broad* job assignments in this model. The reason is that if communication is costly but workers can still coordinate costlessly *within* their groups of tasks, it is valuable to both have high quality communications and to reduce the degree of specialization so that these communications are not too costly *in* practice (e.g., not too many meetings). One might label this organizational model as 'flexible production.'

Now consider what happens when communications costs fall (i.e., due to computerization). Firms will find ex post coordination cheaper (i.e., communications among adapted tasks). This may increase the value of giving workers discretion to adapt their tasks (less coordination failure ex post). But if so, it will also increase the value of bundling tasks so that workers can costlessly coordinate within tasks they are adapting. Thus, broad task assignments and intensive 'horizontal communications' are complements under some circumstances.

2.2 Task structure of production

Production can be thought of as taking place on an $n \times n$ grid of tasks:

a^{11}	a^{12}	a^{13}	...	a^{1n}
a^{21}	a^{22}	a^{13}	...	a^{2n}
a^{31}	a^{12}	a^{33}	...	a^{3n}
...	...	...	...	a^{4n}
a^{n1}	a^{n2}	a^{n3}	...	a^{nn}

The diagonal elements of this grid are the *primary* actions. The off-diagonal actions are the coordinating complementary actions.

Workers are assigned one or more *rows* of tasks, with the number of rows t assigned to each worker determined in equilibrium. For each row that a worker is assigned, it is her responsibility to select both the primary and coordinating actions for the tasks in that row. For simplicity, assume that task assignments are symmetric, so that all workers perform the same number of tasks.

Although workers are responsible for rows, output is a function of the elements of a column. In particular, the complementary actions in each row must be coordinated with the actions in their respective columns *not* the actions in their row. This means that workers responsible for different rows must coordinate with one another to choose tasks optimally within a column.

Before discussing how tasks are chosen, it's useful to understand the sequencing of the model, which goes as follows:

1. The firm determines the number of task per agent, t.
2. Local circumstances θ^i for $i = 1, 2, ..., n$ are realized and observed by employee(s) in charge of task i.
3. Workers communicate the realizations of local information, and each attempt at communication is successful with probability p. Workers conveying this information cannot determine whether or not communication was successful.
4. For each row i, the employee in charge of i chooses actions a^{ij}, where $j \in \{1, 2, ..., n\}$, to maximize the objective function given his information.
5. Profits are realized given the realization of local information, the success of communication, and the chosen values of all tasks.

Now let's consider how actions are chosen.

2.3 Adaptation and coordination

Task i consists of undertaking a primary action a^{ii}, whose effectiveness depends on how well it is adapted to the local environment. Adaptation calls for use of *local information*, which pertains only to this task and can only be observed by the worker assigned to the task.

The local information is a random variable θ^i with mean $\hat{\theta}^i$ and common variance σ_θ^2. To achieve perfect adaptation, the primary action a^{ii} should be equal to θ^i. The realization of θ^i is independent across tasks.

To ensure that task i is coordinated with all tasks $j \neq i$, the employee in charge of task i must perform a sequence of $n-1$ actions $\left\{a^{i1}, a^{i2}, ..., a^{in}\right\}$ that are complementary to the primary actions of tasks $j \neq \dot{i}$. To achieve perfect coordination between tasks i and j, action a^{ij} of task i should be set equal to the primary action a^{jj}. Notice (as above) that worker in charge of task i must coordinate the complementary actions with the tasks in rows $j \neq i$.

So, if the organization consists of two tasks, then adaptation and coordination losses amount to:

$$\phi\left[\left(a^{11}-\theta^1\right)^2+\left(a^{22}-\theta^2\right)^2\right]+\beta\left[\left(a^{12}-a^{22}\right)^2+\left(a^{21}-a^{11}\right)^2\right],$$

where the parameters $\phi \geq 0$ and $\beta \geq 0$ determine the importance of adaptation and coordination respectively. Here, a^{11} and a^{22} are the primary tasks, and actions a^{12} and a^{21} are the subordinate tasks for each primary action respectively.

2.4 Specialization

Let $T(i)$ equal the set of tasks bundled with worker i. For simplicity, assume that organizational design is symmetric, so that all workers in an organization have an identical number of tasks t assigned to them, so that $T(i) = t$ for all workers (and the total number of workers is n/t).

Task variety is costly in the Smithian sense in that forgoes the gains from specialization. Specifically, the labor cost of carrying out task i, denoted by $h(t, \alpha)$, is increasing in the level of task bundling t and in α. Thus $h(\cdot)$ is a per-task cost, with:

$$h(\bar{t}) - h(\underline{t}) \geq 0, \text{ and } h_\alpha(\bar{t}) - h_\alpha(\underline{t}) \geq 0, \text{ and } \bar{t} > \underline{t}.$$

(Put more simply $h(0) \geq 0$ and $h_t(\cdot) > 0$.) Bundling tasks raises per unit costs of execution by sacrificing gains from specialization. (The parameter α reflects the returns to specialization. When α is larger, the gains from specialization are larger because the costs of non-specialization are higher.)

2.5 Communication

To improve coordination, workers can communicate the realization of the local information θ^i prior to the actual implementation of the actions. If tasks i and j are assigned to different employees, then with probability p the value of θ^i will be communicated successfully and with probability $(1-p)$, it will be pure noise. Thus, p is a measure of the quality of the communication channel. It is assumed that agents who receive communication know whether communication was successful or pure noise. Agents sending this information do not know whether their communication was successful.

2.6 Profits

Profits of a firm are given by:

$$\begin{aligned}
\pi &= -\sum_{i=1}^{n} C^i\left(a^{1i}, a^{2i}, ..., a^{ni}, t|\theta^i\right) \\
&= -\sum_{i=1}^{n}\left[\phi\left(a^{ii} - \theta^i\right)^2 + \sum_{j \notin T(i)}^{n} \beta\left(a^{ji} - a^{ii}\right)^2 + h(t, \alpha)\right]
\end{aligned}$$

You can think of this profit function as a loss function that the firm is attempting to minimize. Hence it might be more intuitive to write $\pi^* = p \times (y + \pi)$, where p is the market price of a perfectly produced unit of output and π is the reduction in quality incurred by imperfections in adaptation and coordination during production.

2.7 Optimal choice of actions

One can show that employees optimally choose the following primary actions:

$$a^{ii}(\tau) = \hat{\theta}^i + \left[\frac{\phi}{\phi + \beta(n-t)(1-p)}\right]\left(\theta^i - \hat{\theta}^i\right).$$

Notice that the degree of adaption is increasing in ϕ, decreasing in β, increasing in the quality of communication p, and increasing in task bundling t.

Complementary actions are chosen as:

$$a^{ji}(t) = \begin{cases} a^{ii(t)} \text{ when task } j \text{ learns } \theta^i \\ \hat{\theta}^i \text{ when task j does not learn } \theta^i \end{cases}.$$

The covariance between local circumstances and the primary action are:

$$\sigma_{a\theta}(t) = Cov\left[a^{ii}(t), \theta^i\right] = \left[\frac{\phi}{\phi + \beta(n-t)(1-p)}\right]\sigma_\theta^2.$$

Thus, $\sigma_{a\theta}(t)$ characterizes how strictly employees adhere to the ex ante strategy ($\hat{\theta}^i$) versus tailoring their actions to local circumstances. Notice that $\sigma_{a\theta}(t)$ is increasing in the *variability* of local circumstances (in addition to the other comparative statics above).

A key observation is that $\sigma_{a\theta}(t)$ is increasing in task bundling:

$$\sigma_{a\theta}(\bar{t}) > \sigma_{a\theta}(\underline{t})$$

Given t, expected profits are:

$$\Pi(t) = n\phi\left[\sigma_{a\theta}(t) - \sigma_\theta^2\right] - nh(t,\alpha),$$

$$t^* = \arg\max_{t \in T} \Pi(t),$$

as a function of $\phi, \alpha, \sigma_\theta^2, \beta$ and p.

2.8 Some results

2.8.1 Specialization

Task specialization is decreasing in the importance of adaption, ϕ, and the variance of local circumstances σ_θ^2, but increasing in the returns to specialization, α.

2.8.2 The relationship between coordination costs and specialization

The Becker-Murphy 1992 *QJE* article would suggest that as coordination costs fall, specialization increases. In the current paper, coordination costs operate through β (holding communication technology fixed). When β is higher, tasks are more interdependent, so coordination costs are higher.

How does a rise in β affect task bundling? Becker-Murphy suggests that increase bundling (reduce specialization). In this model, the answer depends on the importance of adaptation ϕ. An increase in task interdependence affects two margins:

1. Holding worker flexibility/adaptation constant, a rise in β makes it more important to improve coordination, which leads to increased bundling.

2. But a rise in β may also spur the organization to reduce employee flexibility and become less adaptive. This reduces the need for task bundling. If β is very large, an employer may eliminate all flexibility, in which case it is optimal to fully specialize $t^* = 1$. Conversely, if β is very small so that tasks are virtually independent, then coordination can be ignored and so it is also optimal to specialize.

Accordingly, the relationship between t^* and β is non-monotone. For $h(n, \alpha) < \infty$ and given α, there exists a unique $\hat{\phi}$ such that:

$$\lim_{\beta \to \infty} t^* = \begin{cases} n \text{ if } \phi > \hat{\phi} \\ 1 \text{ if } \phi < \hat{\phi} \end{cases}$$

Moreover, when $\phi < \hat{\phi}$, t^* is non-monotone in β, increasing and then decreasing. Initially, a rise in β leads to more bundling to improve coordination. As β gets larger, firms simply reduce adaption and make jobs very narrow. This resolves coordination problems, though also makes firms less adaptive.

2.8.3 Quality of communication

Intuition would suggest that one would see more specialization as communication costs $(1 - p)$ fall since coordination gets easier. But organizations may also take advantage of improved communications to become more adaptive. In this case, by bundling, the organization increases its adaptiveness to local environmental variables. This increased adaptation is complemented by improved communication because it becomes more important to coordinate the other tasks when the primary action is allowed to vary more.

So, there is again a trade-off:

1. For given level of employee flexibility, $\sigma_{a\theta}(t)$, an improvement in communications makes it easier to coordinate, reducing the need for bundling.

2. But as coordination through communication improves, firms will want to increase employee adaptability. This favors additional task bundling.

When communication channels are very poor, task bundling often increases as communication technology improves. Firms go from being extremely inflexible to somewhat more adaptive. In this range, communications and bundling are complements.

But when communication technology gets sufficiently good, bundling looks less and less attractive since it forgoes Smithian gains from specialization but saves little in coordination errors (since the quality of communications is good). In this range, communications and bundling are substitutes.

2.8.4 Other comparative statics

There are many interesting extensions in the paper, which I leave for your further reading. An important choice variable that the paper studies is the optimal level of p when improvements in p can be purchased at a positive price.

3 The Economics of Superstars

The 1981 *AER* article by Sherwin Rosen, "The Economics of Superstars," is often cited as a prescient explanation for the rise in returns to skills that was to occur in the years immediately following its publication. Whether or not you believe that the Superstars hypothesis helps to explain the phenomenon of rising inequality, the paper offers a fascinating set of insights that has considerable currency as an explanation for why wages of CEOs, entertainers and athletes are incomparably higher than for other occupations. (That is, the Superstars model certainly offers a deep and "truthy" explanation for why the distribution of wages is so skewed in some occupations. Additionally, the hypothesis may or may not explain why the overall distribution of wages has become *more* skewed (more skewered?))

Certain services appear to be characterized by a demand structure that places considerable weight on quality versus quantity. For example, a patient would rather have one really good heart operation than two mediocre ones; an opera fan would rather see one Pavarotti concert than ten mediocre opera performances; no amount of Applebee's food amounts to a meal at Oleana. Hence, if there is widespread agreement on *who* is talented among the providers of such services, one can imagine that talented service providers could earn considerable rents.

There is also a natural check on the possibility that one 'talent' will control an entire market: congestion. A single heart surgeon cannot perform all of the surgeries; a great cook can only make so many dinners in an evening. It may be these congestion externalities that prevent single talents from controlling entire markets.

Now, imagine a technological improvement that reduces congestion, allowing holders of talent to more efficiently serve larger markets. Examples: television, recording technology, the Internet, etc. These types of technological advances may increase the rents accruing to the most talented individuals by increasing their market share (possibly to a global scale). Some professions in which this is likely to be important: all performing artists and entertainers; managers of large corporations (CEOs); athletes; fund managers; possibly some academics. Is is actually hard to make a very long list.

The Rosen model almost surely helps to explain why athletes and movie-stars earn such enormous salaries, and why these salaries have risen as mass communications and entertainment have improved. The 2008 *AER* paper by Terviö explores this idea rigorously.

3.1 What are CEO's Worth (and Why)?

CEO pay is a divisive topic on which there has been an enormous amount of research. While there is room for debate about whether this topic receives more attention than it merits, it still offers 'teachable moments.' One such moment is the 2008 *AER* by Terviö. This paper studies CEO pay and productivity through the lens of an assignment model. Assignment models were introduced by Nobel Laureate Jan Tinbergen in the 1950s, and they were 'popularized' (okay, made less obscure) by Sattinger's widely cited 1993 *JEL* survey. The defining feature of assignment models is the presence of indivisibilities among factors of production. In a situation where the amounts of two matching factors cannot be shifted across different units of production, factors are *not* paid their marginal products in the standard sense. As we will see, it's possible (indeed) likely for the wage paid to a given worker to rise or fall due to changes in the distribution of ability of other workers *without* any change in the productivity of that worker or in the value of his output.

Assignment models have defied widespread use, perhaps in part because they are typically rather non-intuitive and analytically difficult. The paper by Tervio presents a beautiful exposition of the intellectual foundation of assignment models by reformulating the basic assignment model using distributional ranks rather than actual distribution functions. This formulation is completely natural because assignment models are inherently ordinal rather than cardinal. The paper's substantive conclusions on CEO pay and its relationship to the distribution of CEO talent, are both insightful and intuitive.

3.1.1 Basic setup

There are three assumptions that simplify the model:

1. One-dimensional quality of factors

2. Continuity of quality distributions

3. Complementarity between the qualities of factors

The first two assumptions are simply for analytical convenience. The third assumption is essential to the conceptual framework. The paper further assumes, risk neutrality and symmetric information (which implies an absence of agency problems).

Worker quality is denoted as a. Firm quality is denoted by b, which will be referred to as firm size. There is a unit mass of workers and firms with continuos, finite support without gaps. (The unit mass should be thought of as a normalization of all units that are active in equilibrium.)

The production function is continuous and strictly increasing in both its arguments. The key assumption is that this function has a positive cross-partial between a and b, so these inputs are complements. Accordingly, efficiency requires positive assortative matching.

As is explained in the paper, one can write the production function *without loss of generality* as:

$$Y(a,b) = a \cdot b.$$

The notation in the paper is actually an important contribution. Order abilities by quantile so that $a[i]$ is the ability of the the i^{th} quantile individual, and $a'[i] > 0$. Denoting the distribution function by F_a, the profile of a is defined by

$$a[i] = a \;\leftrightarrow\; F_a(a) = i.$$

3.1.2 Equilibrium conditions

There are two types of constraints on the problem: a sorting constraint, which requires that no worker/firm pair wish to rematch; and an incentive compatibility constraint, which requires that all workers and firms earn at least their outside income:

$$\begin{aligned} Y(a[i], b[i]) - w[i] &\geq Y(a[j], b[i]) - w[j] \quad \forall i, j \in [0,1] \qquad SC(i,j) \\ Y(a[i], b[i]) - w[i] &\geq \pi^0 \quad \forall i \in [0,1] \quad PC\ b[i] \\ w[i] &> w^0 \quad \forall i \in [0,1] \quad PC\ a[i] \end{aligned}$$

It is assumed that π^0 and w^0 are the same for all units, though it would be sufficient to assume that outside opportunities increase slower along the profile than do equilibrium incomes. The lowest active pair breaks even:

$$Y(a[0], b[0]) = \pi^0 + w^0.$$

While there are in theory $2n!$ sorting constraints, where n is the number of units (if there were a finite number, which formally there are not), most constraints are redundant since for $i \geq j \geq k$, $SC(i,j) + SC(j,k)$ implies $SC(i,k)$. The binding constraints are therefore: (1) the marginal sorting constraints that keep firms from wanting to hire the next best individual; and (2) the participation constraints of the lowest types.

Ordering the binding sorting constraints, one can write:

$$\frac{Y(a[i], b[i]) - Y(a[i-\varepsilon], b[i])}{\varepsilon} \geq \frac{w[i] - w[i-\varepsilon]}{\varepsilon}. \tag{1}$$

This equation becomes an equality as $\varepsilon \to 0$, and yields the slope of the wage profile (since the limit of the above expression is the partial derivative):

$$w'[i] = Y_a(a[i], b[i])\, a'[i],$$

where Y_a is the partial derivative. This gives us the slope. To get the full profile, we need to integrate over the production function and add in the binding participation constraint:

$$w[i] = w^0 + \int_0^i Y_a(a[j], b[j])\, a'[j]\, d[j].$$

Similarly, the profile of profits must satisfy:

$$\begin{aligned} \pi'[i] &= Y_b(a[i], b[i])\, b'[i] \\ \pi[i] &= \pi^0 + \int_0^i Y_b(a[j], b[j])\, b'[j]\, d[j]. \end{aligned}$$

These conditions also imply that $y = \pi + w$ at each firm. Note that this set of sorting conditions could equivalently be written in terms of workers choosing firms rather than firms choosing workers; these cases are isomorphic.

This is an extremely tight set of constraints on the problem. It implies that the wages and profits of each factor depend at quantile i depend on the full profile of factors from quantile 0 to $i - \varepsilon$ (but not on the profile above i).

A number of useful observations follow (and you should read the paper, which is chock full of astute insights, and even more so in the footnotes):

1. There is no bargaining in this model because of the continuity of the distribution of both factors. If there was a jump at some point in the profile of one factor, then all of the surplus would go to the factor with the jump because the other side is perfectly competitive.

2. Payments to factors are only affected by the quality of those *below* them in the ranking. This is because the binding constraint on each worker or firm is that the quality/price of the worker/firm just below it in the distribution.

3. The unobserved productivity characteristics a and b are essentially ordinal. Any increasing transformation of the scale of measurement for a factor's quality, combined with the inverse change in the functional form of the production function changes nothing of substance in the model. In this sense $Y(a[i], b[i]) = a[i] \times b[i]$ is a general functional form (so long as we are assuming complementarity).

4. Quoting Tervio, "It would be incorrect to say that factors earn their marginal productivity by the usual definition of marginal productivity, because the increase in output if the individual of ability $a[i]$ were to increase in ability is proportional to $b[i]$. But if she were to increase in ability, then, in equilibrium, she would also move up in the ranking and be matched with a higher b— and other individuals would have to move down and experience a decrease in productivity. This means that individuals in fact do receive their marginal product, once the margin is defined correctly. As ability cannot conceivably be extracted from one individual and poured into another, the relevant margin here is whether an individual will participate in the industry or not—and if not, then the effect of the resulting rearrangement of remaining individuals is part of the marginal product."

3.1.3 Comparative statics

Consider a change by which the production function Y is multiplied by some constant G with no change in distributions of a and b. Uniform productivity growth is mathematically equivalent to changing the units of measurement for output. This gives rise to a useful lemma:

Scaling lemma: If $Y_t(a,b) = GY(a,b)$, $w_t^0 = Gw^0$ and $\pi_t^0 = G\pi$, then $w_t[i] = Gw_t[i]$ and $\pi_t[i] = G\pi_t[i]$ for all $i \in [0,1]$.

Figure 1 summarizes much of the intuition of the model. In the strictly multiplicative case, the output accruing from matching a worker of ability a and a firm of ability b is $a \times b$, which is simply a rectangle in a Cartesian graph. This graph is drawn as follows. Let $a = \varphi(b)$, defined by $a[F_b(b) = \{(a,b) \;\; st.F_a(a) = F_b(b)\}]$ with slope

$$\varphi'(b) = a'[F_b(b)]\, f_b(b) = \left.\frac{a'[i]}{b'[i]}\right|_{i=F_b(b)}.$$

$\varphi(b)$ is strictly increasing in b, and the slope is given by the relative steepness of a and b at each quantile i.

Be sure to understand Figure 1, since essentially all of the intuition of the assignment model is contained in this figure.

Here's an interesting exercise to build intuition. Let's check Tervio's assertion that "the relevant margin here is whether an individual will participate in the industry or not—and if not, then the effect of the resulting rearrangement of remaining individuals is part of the marginal product." Consider a hypothetical case where the highest ability worker (index $a[1]$) falls in ability to that of the lowest ability worker $a[0]$. Can we demonstrate that the reduction in total output is equal to this worker's wage?

- The "demotion" of the highest ranked worker means that each firm other than $b[0]$ will have to match with a slightly lower ranked worker. So, previously, total output was equal to:

$$Y = \int_0^1 Y(a[j], b[j])\, d[j]$$

- Now, each firm except for the lowest ranked firm $b[0]$ will have to pair with a slightly lower quality worker. Thus, the fall in output at each firm $Y[j]$ is

$$\Delta Y[j] \equiv Y[j] - \hat{Y}[j] = Y(a[j], b[j]) - Y(a[j-\varepsilon], b[j]). \tag{2}$$

  The output of the lowest ranked b is unaffected since it still pairs with the lowest ranked a. But for all units b above rank 0, they will now be pairing with a units that are ε below them in the prior ranking. Note that $a[\cdot]$ continues to refer to the values of a in the *original* distribution, not the new distribution.

- Dividing equation (2) by ε and letting $\varepsilon \to 0$, we take the limit of

$$\frac{\Delta Y[j]}{\varepsilon} = \frac{Y(a[j], b[j]) - Y(a[j-\varepsilon], b[j])}{\varepsilon},$$

  to get

$$Y'[j] = Y_a(a[j], b[j])\, a'[j].$$

- We integrate over the full distribution of units to obtain the total loss in output:

$$\Delta Y = \int_0^1 Y_a(a[j], b[j])\, a'[j]\, d[j]$$

  This is the net reduction in output caused by worker $a[1]'s$ demotion.

- Having solved for the reduction in output, let's compare this to the change in the wage bill. The wage of the previously highest ability worker falls from

$$w[1] = w^0 + \int_0^1 Y_a(a[j], b[j])\, a'[j]\, d[j]$$

  to

$$\hat{w}[1] = w^0.$$

  Thus the change in the wage for worker $a[1]$ is

$$\begin{aligned} w[1] - \hat{w}[1] &= w^0 + \int_0^1 Y_a(a[j], b[j])\, a'[j]\, d[j] - w_0 \\ &= \int_0^1 Y_a(a[j], b[j])\, a'[j]\, d[j], \end{aligned}$$

  which is identical to the fall in total output, ΔY. This confirms Tervio's assertion.

3.1.4 Applying this model to CEOs

There are numerous conceptual and practical issues to confront before bringing this model to the data. I will discuss these assumptions and results only briefly.

The issues:

1. The surplus created by the CEO-firm interaction is unobserved. The market value of the firm is affected by the current CEO and by expectations of future productivity (which also depend on future CEOs)—thus firm size, an outcome, cannot be treated as the b variable.

2. A second issue is that part of the market value of the firm will surely reflect the value of capital that can be readily transferred/resold among firms. This capital is *not* indivisible and so is not part of the surplus represented in $Y(\cdot)$.

3. A third issue is that the current market value of a firm depends on the quality of the current CEO and the quality of past and (in expectation) future CEOs. Thus, one cannot necessarily infer the current CEO's quality purely from contemporaneous data.

4. Productivity tends to grow over time, and the expectation of growth further affects current market value.

5. The distribution of CEO ability and latent (exogenously determined) firm size (not market value) can change over time.

6. Outside options may shift.

Thus, the subsequent analysis rests on a number of strong assumptions that necessarily make the exercise speculative (though perhaps still informative if one believes the foundational assumptions of the assignment model):

1. The distribution of $a[i]$ and $b[i]$ are time invariant.

2. Productivity grows deterministically at rate g at all firms (that's why the scaling lemma is needed)

3. The value of outside options grow at rate g. Tervio uses values between 0.2 and 0.025.

4. The discount rate is constant. Tervio uses values between 0.08 and 0.05.

5. The impact of past *and future* CEO quality on current firm performance decays at a constant rate $\alpha_{\tau+1} = \alpha_\tau \lambda / (1+\lambda)$. λ determines the decay rate. With $\lambda \to \infty$, only the current CEO affects contemporaneous earnings. Tervio uses values between ∞ and 0.1

6. Since adjustable capital must earn the market rate of return, it must be subtracted from Y. To determine the contribution of adjustable capital, Tervio assume's that the gross surplus has constant elasticity θ with respect to adjustable capital. Tervio sets the share of adjustable capital in Y at values between 0 and 0.8.

These assumptions are sufficient to pin down the contribution of firm size and ability to output up to an additive constant (one constant each for the two factors). The reason the constant is indeterminate is that one cannot infer the contribution of firm size and ability to surplus at the smallest firm in the sample since this depends on the full unobserved distributions of size and ability below this firm. However, these constants drop out when estimating counterfactuals of the form 'how much would surplus rise or fall *relative to* the current baseline if the distribution of CEO ability were modified in various ways?'

Tervio explores three counterfactuals for CEO ability: all firms are paired with a CEO of the ability of the lowest ranked CEO; all firms are paired with the median CEO; all firms are paired with the highest ranked CEO. One can also perform a similar counterfactual for firm size. See Tables 1 and 2. Several points:

1. The contribution of firm size to CEO rents is much larger than the contribution of CEO ability. That is, in the logic of this model, the high levels of CEO pay at the top are mostly due to the exogenous component of firm scale.

2. The distribution of CEO ability *relative to* the lowest ability CEO can be inferred from the differential equations that describe the equilibrium outcomes. See Figure 5. A key conclusion, and one that should be intuitive, is that the distribution of CEO ability is *not* wide. It is estimated that the CEO at the top ranked firm is only 5 to 8 percent more able than the CEO at the bottom ranked firm.

3. To see the role of firm size in CEO pay, consider figure 4 where all firms are replaced with the lowest, median or highest ranked firm while the CEO ability distribution is held constant. Panel B is perhaps most informative. If all firms were the median size, the top CEO would earn 12.5 million less than the 500th CEO than is currently the case and the bottom CEO would earn an additional 2.5 million less relative to the 500th CEO than is currently the case. (Why greater pay dispersion at the bottom? Because the flat profile of firm size, all surplus is allocated to workers–unlike in the continuous firm distribution case. This surplus redistribution effect can offset the leveling effect of greater equality of firm size. The relevant comparison here is not really among CEOs in these various scenarios but rather between scenarios–highlighting how changes in firm size hugely affect the level and dispersion of CEO pay, holding CEO talent constant.)

3.1.5 How the heck does Terviö calculate what part of the capitalized market value of firms is due to CEO vs. firm 'quality?'

This is pretty tricky.

- Under the assumptions of (1) deterministic, constant growth; (2) proportionate growth in the value of outside options; and (3) no change in the quality distribution of factors over time, one obtains the following relationship

$$\frac{w[i]}{1-B} + \nu[i] = \frac{a[i]\,b[i]}{1-B},$$

  where $B = (1+g)/(1+r)$ is the growth adjusted interest rate. The first term on the left is the present value of pay to *all* CEOs that firm i will ever employ and $\nu[i]$ is the market value of the firm. As shorthand, Terviö refers to $1/(1-B)$ as the price to earnings ratio.

- Adjustable capital can be partialled out from $\nu(i)$ using a Cobb-Douglas assumption. Only the part of ν that excludes adjustable capital is subject to the assignment process.

- The differential equations (24) and (25) in the paper permit inference on the quality of CEOs and firms at each percentile in the distribution *relative to* the lowest quality units $(a[0], b[0])$

participating in the market:

$$\frac{a[i]}{a[0]} = \exp\left\{\frac{\lambda}{\lambda+1-B}\int_0^i \frac{w'[j]}{w[j]+v[j](1-B)}dj\right\},$$

$$\frac{b[i]}{b[0]} = \exp\left\{\int_0^i \frac{v'[j]-w'[j]/\lambda}{w[j]+v[j](1-B)}dj\right\},$$

where λ is the decay rate of present CEO contributions to future firm value.

- I think a large part of the intuition of Figure 5 is that CEO pay increases very, very little from the smallest to largest firm, whereas market value increases by hundreds of billions (say from 19 billion at rank 1,000 to 750 billion at rank 1,000). This loosely implies that the slope of CEO talent must be relatively shallow.

3.1.6 Conclusions

You do not have to be convinced that this model or empirical exercise fully (or even partly) explains the level or trends of CEO pay to find much of value in this paper. As a labor economist, you should add the basic workings of assignment models to your intellectual toolkit. And this is best paper I've seen for acquiring those tools. As I will discuss in class, the assignment model can be used to interpret the Rosen 1981 *AER* superstar's model, though Rosen's original model is not constructed in the form of an assignment model.

I will not spend class time on the related paper by Gerbil and Lanier (*QJE* 2008) and the critique of Gabaix-Landier by Gordon and Dew-Becker ("Unresolved Issues in the Rise of American Inequality," *Brookings*). This is an important debate.

4 Terviö: Mediocrity in Talent Markets

A closely related paper (one could even say a 'companion paper') by Terviö proposes an alternative model that casts the same facts in quite a different light. The point of departure between the Rosen and Terviö articles is in the mechanism by which talent is discovered. Rosen takes it as given that talent is *known* to the market. Given full information, the wages paid to talent are likely to be efficient. Terviö's paper takes a step back by asking: how does talent become known? Terviö observes that to be 'discovered,' one must first 'audition.' This observation seems plausible for many talent-occupations. In the Terviö model, knowledge about worker quality is a joint output of production. (That is, to find out if a worker would make a good actor, he needs to appear in a movie.) Economist Richard Caves refers to this ex ante uncertainty about talent as the "Nobody Knows" property – one cannot evaluate talent without putting it to use. Now, add a second key assumption: there are positive production costs. That is, workers cannot audition 'for free' because some other scarce is used for the audition. This resource could be capital or, more realistically, it could be the time of skilled evaluators (talent agents, moviegoers, academics, passengers on an airplane, etc.). More generally, consumers cannot become informed about the entire array of talents potentially available in the market because time and attention devoted to discovery is scarce. To evaluate and disseminate information about talent is costly. If these assumptions are correct, there is a positive opportunity cost to auditioning a worker's talent by using him for production. The cost is that someone else better could fill the slot (therefore making better use of the same scarce resource).

Now assume: 1) learning is symmetric (i.e,. my talent is revealed to the public as the same time as it is revealed to the employer – this assumption can be weakened); 2) workers cannot commit to binding wage contracts (i.e., no indentured servitude); 3) workers are capital constrained so that they cannot simply 'buy' a job to audition their talents.

4.1 Setup

- Firms need one worker and one unit of capital at cost $c > 0$ to produce output in each period.
- Output of a firm is equal to the worker's talent: $Y = \theta$.
- Talent is drawn from a distribution with a continuous and strictly increasing cumulative distribution function, F, with positive support $[\theta_{min}, \theta_{max}]$.
- There is an unlimited supply of potential workers, all of who have an outside wage of w_0. Hence, *talent is not scarce.*
- Talent is industry-specific and becomes public knowledge after one period of work; the worker may then work in the industry up to T more periods, after which he ceases to be productive.
- Both workers and firms are risk neutral, and there is no discounting.
- Industry output faces a downward-sloping demand curve $p^d(q)$. We think of the "quantity" of talent as referring to the quality of output times the number of workers I. Thus, all else equal, the equilibrium price of output is lower when talent in the industry is higher (i.e., less scarce)
- The number of firms is "large" so that individual firms have no impact on total output, and there is no uncertainty about the realization of the distribution of talent.
- Let the number of firms (and jobs) be a continuous variable I equal to the mass of the industry workforce, normalize so that $I = 1$
- Long-term wage contracts are not enforceable; workers cannot commit to decline higher offers from other firms in the future.

Some key mechanics

- There will be a talent threshold ψ above which workers will be retained in the industry and below which they will not.
- The social planner would like to maximize average talent in the industry. This is because the opportunity cost of talent in the outside market is uncorrelated with talent. Thus, the higher the level of talent *in* the industry, the higher is aggregate social welfare. (Note that output is linear in talent, so the average is a sufficient statistic for the maximization problem.)
- In steady state, inflows of talent (new workers, that is) must equal outflows of talent.

Some key economics

In an efficient market allocation

1. The veteran worker of the lowest type ψ retained in the industry must be indifferent between exiting and remaining in the industry, and thus must earn exactly the outside wage w_0
2. Novices must be indifferent between entering the industry or pursuing the outside career, taking into account their option to exit for the outside career later on (that is, before $T+1$ periods have elapsed)
3. Firms must expect zero profits from hiring any talent
4. The output price must be market clearing

4.2 Talent threshold

- In both market equilibrium and in the social planner's optimal solution, individual careers will proceed in a simple manner:
  - *After one period of work, those whose talent is revealed to be below a certain threshold level ψ will exit the industry, while those above the threshold ψ will stay on for T more periods. This exit threshold will be the key variable in the model.*
- In steady state, **inflows must equal outflows**. Denote the fraction of novices in the workforce by i; a fraction $F(\psi)$ of them exit. The remaining fraction of jobs $1-i$ are held by veterans; a fraction $1/T$ of these, the oldest cohort, retires each period. Equating the flows of exits and entries yields:

$$iF(\psi) + \frac{1}{T}(1-i) = i$$

$$i(\psi) = \frac{1}{1+T(1-F(\psi))}$$

- **Average talent in the industry is therefore**

$$A(\psi) = i(\psi)\bar{\theta} + (1 - i(\psi)) E[\theta|\theta \geq \psi].$$

One can substitute the equation above for $i(\psi)$ into this expression to obtain

$$A(\psi) = \frac{1}{1 + T(1 - F(\psi))}\bar{\theta} + \frac{T(1 - F(\psi))}{1 + T(1 - F(\psi))} E[\theta|\theta \geq \psi].$$

Notice that a higher ψ means that a larger fraction of workers are novices (which lowers average quality) but the average quality of retained veterans is higher, which raises average quality.

- *This expression will be maximized when the average talent in the industry (averaging over new entrants and veterans) is equal to the marginal talent retained.* (When ψ is too low, raising it increases both the marginal and the average; when ψ is too high, raising it raises the marginal but reduces the average because there are too few veterans.)

- The equilibrium condition for ψ can be written as

$$\psi - \bar{\theta} = T(1 - F(\psi))(E[\theta|\theta \geq \psi] - \psi), \tag{3}$$

the solution to which is denoted as A^*. The LHS of this expression is the foregone output of hiring a novice rather than a veteran (invoking that average output is equal to marginal output ψ at A^*). The RHS is the expected future gain from hiring a novice, equal to the probability that the novice is retained for T additional periods times expected output (conditional on being retained) in each subsequent period (invoking that ψ is constant across future periods).

- The paper demonstrates that the solution to this condition is unique with no distributional assumptions on θ required (except continuity and boundedness). Here's a numerical example of the optimal choice of ψ with $\theta \sim N(0, 1)$ and $T = 15$:

TervioMediocrityCalc.pdf

4.3 Indifference condition for entrants and veterans

- Veterans must choose to exit if $w(\theta) < w_0$, therefore the marginal veteran is indifferent between exiting and remaining in the industry

$$w(\psi) = w_0.$$

- Similarly, novices must be indifferent about entering the industry, accounting for the option to exit if $\theta < \psi$:

$$w\left(\bar{\theta}\right) = T \times E\left[\max\left\{w\left(\theta\right), w_0\right\}\right] = \left(1+T\right) w_o.$$

  This expression may imply that $w\left(\bar{\theta}\right) < 0$, that is entrants will receive a *negative* wage. The optimal solution may therefore require an absence of binding credit constraints.

- If constrained individuals cannot accept a wage below $w_0 - b$ where $b > 0$, then the wage of novices will be given by

$$w\left(\bar{\theta}\right) = w_0 - b.$$

  This will yield substantial market-level inefficiencies as we shall see.

4.4 Zero profit condition

- Firms must expect zero profit from hiring any talent. Denoting the output price as P, this implies that

$$P\theta - c - w\left(\theta\right) = 0.$$

- Since there is no risk aversion and information is symmetric, novices can be treated as a known quantity with talent $\bar{\theta}$.

4.5 Market clearing

- The output price must clear the market. We've stipulated downward sloping demand, and we know that average talent is rising in ψ up to the optimum A^*. Therefore, there will be a fixed point for the market price equation where

$$P = p^d\left(I \times A\left(\psi\right)\right),$$

  where $I \times A\left(\psi\right)$ is total talent supplied.

- Since talent in the industry is a function of ψ, we can rewrite the wage function in terms of individual talent and the market talent cutoff $w\left(\cdot\right) = w\left(\theta|\psi\right).$

- We can use the zero profit condition to get an expression for P in terms of the wage.

  - A firm employing the threshold obtains revenue $P\psi$ and pays costs $w\left(\psi|\psi\right) = w_0 + c$.
  - Thus, the equilibrium output price must satisfy

$$P\left(\psi\right) = \frac{w_0 + c}{\psi}. \tag{4}$$

- Since firms must be indifferent about hiring any level of talent, the surplus must go to the worker, implying that $w(\theta|\psi) = w_0 + \{P\theta - (w_0 + c)\}$. Rearranging:

$$w(\theta|\psi) = w_0 + \left[\frac{\theta}{\psi} - 1\right](w_0 + c). \tag{5}$$

Thus, a worker at threshold talent of ψ earns w_0, which is her opportunity cost. A worker with $\theta > \psi$ earns a wage in excess of w_0 while a worker with $\theta < \psi$ would exit the market and take w_0 instead.

4.6 Solution with unconstrained workers

- Competitive equilibrium with unconstrained individuals is socially efficient, so the social planner's solution already tells us that the exit threshold must be $\psi = A^* - \bar{\theta}$.

- Inspection of the wage equation (5) makes clear that novices must accept less than the outside wage w_0 since they have a positive probability of earning talent rents in the future while in the worst case they get the outside wage. If the novice wage were not below w_0, novices would earn rents. (Concretely, the fact that the novice wage is below ω_0 is evident from the fact that $\left(\bar{\theta}/\psi\right) < 1$.)

- Market equilibrium pins down the wage function as $w(\theta|A^*)$ and $P^* = (w_0 + c)/A^*$. Intuitively, note that unconstrained individuals bid for the chance to enter the industry up to the expected value of lifetime talent rents. Since veterans of the threshold type are available at the outside wage, novices have to pay $P \times \left(\bar{\theta} - \psi\right)$ for this first period job, which exactly compensates a novice-hiring firm for its expected revenue loss (compared to what it would get by hiring a threshold type).

- In equilibrium, this novice payment must equal the expected lifetime rents: with threshold ψ , a novice has a probability $1 - F(\psi)$ of being retained, in which case he gets the excess revenue $P \times (\theta - \psi)$ as a rent on each of the T remaining periods of his career. So, the market equilibrium condition is

$$P \times (\theta - \psi) = P \times T(1 - F(\psi))(E[\theta|\theta \geq \psi] - \psi),$$

  where the P cancels out and hence this is identical to (3) above.

- Payments by unconstrained novices raise the exit threshold to the efficient level. Thus, the inability of workers to commit to long-term contracts does *not* cause problems *if* they are able to buy their jobs upfront.

- The unconstrained payment (the price of a job) reflects the economic cost of the efficient level of experimentation.

$$b^* = P^*\left(A^* - \bar{\theta}\right) = (w_0 + c)\left(1 - \frac{\bar{\theta}}{A^*}\right)$$

As Terviö notes, the fraction of the total costs of production, $w_0 + c$, that should be financed by the novice is increasing in $A^*/\bar{\theta}$, which is a measure of the upside potential of novices. (For small values of b^* the novice payment would merely be a wage discount below the outside wage.)

4.7 Equilibrium with constrained workers

- Suppose now that the ability of individuals to pay for their first period job is constrained at some $b < b^*$ due to an exogenous liquidity constraint. Now the novice payment is $w(\theta|\psi') = w_0 - b$, where ψ' denotes the threshold in the constrained case.
- Then

$$\psi'(b) = \begin{cases} \left(\frac{w_0+c}{w_0+c-b}\right)\bar{\theta} & \text{for } b < b^* \\ A^* & \text{for } b \geq b^* \end{cases} \tag{6}$$

- Clearly the exit threshold is increasing in b. It follows that the average talent in the industry is also increasing in b. When novices cannot "subsidize" their employers, then the price of output must adjust upwards to induce the hiring of novices into the industry.
- This implies that when there are liquidity constraints faced by novices:

  1. threshold quality will be too low
  2. turnover will be too low and average career lengths will be too long
  3. average quality will be too low
  4. output price will be too high
  5. those who enter the industry earn rents in expectation
  6. top wages will be 'too high' because the top talents will still be drawn from the same distribution as in the unconstrained case, but they'll face a higher output price.

- I won't go through the specifics of the solution in this setting, but they are straightforward provided that you are clear on the above. Key point: *Credit constraints yield, in equilibrium, mediocre veterans in the market*

$$\psi' \in \left(\bar{\theta}, A^*\right)$$

4.8 Two examples

Let's work with two simple examples. Both will have the same parameter values, but in one case we'll assume that workers are not credit constrained and in the other case, we'll assume that they are. For concreteness, let's say that this industry is the movie business. The parameter c refers to the cost of making a movie. Everything is measured in $1,000's$ of dollars. So let's say that $c = 4,000$, that is, the cost of making a movie is \$4 million.

Parameter values

- $T = 15$ (where the novice period does not count in T)
- Fixed cost of production is $c = 4,000$
- Workers' outside option is $w_0 = 0$
- The talent distribution is uniform on 0 to 100. $\theta \sim U[0,100]$, which implies that $\bar{\theta} = 50$ and $E[\theta|\theta \geq \psi] = 50 + \psi/2$
- Output is equal to talent
$$Y(\theta) = \theta$$
- Surprisingly, this is all the info needed to solve the model.

4.8.1 No credit constraints case

Start with the case where there are no credit constraints, meaning that the first period wage in the industry can be negative. Solving for equilibrium:

- Let's start with the talent threshold (equation 3)
$$\psi - 50 = 16\left(1 - \frac{\psi}{100}\right)\left(50 - \frac{\psi}{2}\right).$$
This generates a quadratic equation with two roots, one of which is $\psi = 80$ and the other of which is out of bounds (430/3). Hence, $\psi = 80$.
- We can now obtain the wage function
$$w(\theta|\psi) = w_0 + \left[\frac{\theta}{\psi} - 1\right](w_0 + c)$$
$$w(\theta|80) = \left[\frac{\theta}{80} - 1\right]4,000$$
$$w(\theta|80) = (\theta - 80)\,50$$
This function implies that $P = 50$. We know this because the worker obtains all of the surplus in excess of the threshold. So if the wage is increasing by 50 per unit θ, then output must be priced at 50.
- Another way to obtain P is to invoke the zero profit for hiring a marginal incumbent as in equation (4):
$$P(\psi = 80) = \frac{w_0 + c}{\psi} = \frac{0 + 4000}{80} = 50$$

- Now we need to solve for b, the cost that a novice pays to enter. From above, and using the fact that the marginal worker quality equals average worker quality, so $A^* = \psi = 80$.

$$b^* = P^* \left(A^* - \bar{\theta}\right) = (w_0 - c)\left(1 - \frac{\bar{\theta}}{A^*}\right)$$
$$b^* = 50\,(80 - 50) = -4000\left(1 - \frac{50}{80}\right) = 1,500$$

Novices pay $1.5 million for the their first job. Why is the worker paying to enter? Because if it turns out that $\theta_i > \psi$, the worker will earn talent rents in the subsequent T periods. Those rents *must* be zero in expectation, or else entry is inefficient: novices are not indifferent about entering, meaning they are paid too much, so other distortions must result.

- *What is the probability of a novice becoming a veteran?* It's $\Pr[\theta \geq \psi] = 0.20$

- *What fraction of active workers are novices?* This is

$$i(\psi) = \frac{1}{1 + T(1 - F(\psi))} = \frac{1}{1 + 15\,(1 - 0.8)} = 0.25$$

- *What are the expected career earnings of a worker who exceeds the talent threshold?*

$$15 \times (50 \times \{E[\theta|\theta \geq 80] - 80\}) = 7,500$$

- Expected lifetime earnings (rents) for a novice are therefore

$$\Pr[\theta \geq \psi] \times 7,500 - b = 0.20 \times 7,500 - 1,500 = 0.$$

- And finally, the top wage in the industry is

$$w\,(100|\psi = 80) = 50 \times (100 - \psi) = 1,000.$$

Thus, the top earner receives **one million dollars** per movie (yes, Austin Powers joke).

4.8.2 Equilibrium with constrained workers

- Now assume that $b = 0$. Workers cannot pay to enter the industry. Optimality no longer applies.

- One can calculate ψ' by using (6)

$$\psi'(b) = \begin{cases} \left(\frac{w_0 + c}{w_0 + c - b}\right)\bar{\theta} & \text{for } b < b^* \\ A^* & \text{for } b \geq b^* \end{cases}$$

With $b = 0$, $\psi'(0) = 50$.

- Why is the talent threshold equal to mean talent in this example? Because novices cannot bid for jobs, veterans with $\theta \geq 50$ will always generate (in expectation) weakly more revenue for firms than will novices. This means that vets with $\theta \geq 50$ will always out-compete novices for a job. Thus, the talent threshold falls to the mean of novices.

- We can solve the wage equation as above

$$w(\theta|\psi) = w_0 + \left[\frac{\theta}{\psi'} - 1\right](w_0 + c)$$
$$w(\theta|\psi) = \left[\frac{\theta}{50} - 1\right] \times 4,000$$
$$= [\theta - 50] \times 80$$

  This implies immediately that $P' = 80$

- We can also invoke the zero profit condition to obtain P (though the equation above already implies that $P = 80$):

$$P' \times \bar{\theta} = 4,000.$$

  Since $\bar{\theta} = 50$, this implies that $P' = 80$.

- What is the expected probability of being retained after the first period? It's $\Pr[\theta \geq \psi'] = 0.50$

- The average output of veterans is $E[\theta|\theta \geq 50] = 75$, since veterans are retained if their talent exceeds the mean.

- The fraction of novices in the industry is

$$i'(\psi) = \frac{1}{1 + T(1 - F(\psi'))} = \frac{1}{1 + 15(1 - 0.5)} = 0.118$$

- The expected career earnings of a worker who exceeds the talent threshold is:

$$15 \times (80 \times \{E[\theta|\theta \geq 50] - 50\}) = 30,000$$

- Expected lifetime earnings (rents) for a novice are therefore

$$\left(1 - \frac{50}{100}\right) \times 30,000 = 15,000.$$

- And finally, the top wage in the industry is

$$w\left(100|\psi' = 50\right) = 80 \times \left(100 - \psi'\right) = 4,000.$$

  Thus, the top earner receives **four million dollars** per movie (no Austin Powers joke).

4.8.3 Comparison

Let's compare the characteristics of these two equilibria

	Constrained Equilibrium	Efficient Benchmark
Talent Threshold ψ	50	80
Proportion Novices i	12%	25%
$E\left[\text{Talent}\vert\text{Veteran}\right] = E\left[\theta\vert\theta \geq \psi\right]$	75	90
Average talent A^*	72	80
Output price P	\$80K	\$50K
Top Wage $P \times (100 - \psi)$	\$4 million	\$1 million
E[Career Earnings| Veteran]	\$30 million	\$7.5 million
Wage of Novice	\$0	-\$1.5 million
E[Rents| Novice]	\$15 million	\$0

What is quite striking here is that the original equilibrium has *lower talent and higher wages* than the efficient benchmark. The reason is that the original equilibrium has artificial talent scarcity. Because talent rents accrue to workers not firms, firms have insufficient incentive to create new opportunities for talent revelation (in fact, they have none). So, there is under-experimentation, excess retention of mediocre talent, and inefficiently high rents.

This problem could be solved in either of two ways: (1) create binding indentured servitude contracts such that a worker's wage is set independent of output and the worker cannot quit once hired (but can be fired in the 1st period) – this used to happen in European Soccer through the transfer fees system; (2) eliminate credit constraints so that workers can bid for jobs. This bidding will generate an efficient market in talent revelation. You might ask yourself whether the markets for CEOs, athletes, movie stars and other celebrity professionals appears to satisfy (1) or (2).

An interesting point on which to speculate is whether this model is, in some sense, more general than the Rosen model; could one could potentially nest the Rosen model as a special case where talent revelation was costless? Almost surely the answer is yes.

4.9 Mediocrities: Conclusion

The key insight of this model is that talent can receive large rents even if that talent mediocre. The reason is that the (ex-post) realization of *known* talent may have artificial scarcity. It will be interesting to see whether this paper has an impact on the economic presumption that the market for superstars is in some intrinsic sense efficient.