A Theory of Factor Allocation and Plant Size

Thomas J. Holmes[*]and Matthew F. Mitchell[†‡]

October 25, 2007

Abstract

This paper develops a theory of how capital, skilled labor, and unskilled labor interact at the plant level. The theory has implications for the relationship between factor allocation and plant size and the effects of trade and growth on the skill premium. The theory is consistent with certain facts about factor allocation and factor price changes in the 19th and 20th centuries.

1 Introduction

This paper develops a theory of how capital, skilled labor, and unskilled labor interact at the plant level. In the equilibrium of the model, the allocation of factors at the plant level depends upon plant size. One outcome in the model is for larger plants to employ a larger fraction of skilled workers than small plants do, what we call a positive *size-skill* relationship. As documented below, there is strong evidence that such a positive relationship has existed in recent decades. It is also possible in the model for there to be a negative size-skill relationship where larger plants employ a smaller fraction of skilled workers. As discussed further below, a century ago the size-skill relationship was actually negative. The assembly lines in the large factories opened by industrialists such as Henry Ford were operated by legions of unskilled workers. In contrast, small plants during this time period tended to hire skilled artisans.

Our main result connects the size-skill relationship—a property of the cross-section of plants at point in time—with changes over time in the skill premium, an aggregate relationship. We consider the effects of forces over time that enable firms to increase plant size. The expansion of markets made possible through reductions in transportation costs and other trade barriers is one such factor that we consider. Chandler (1990) has emphasized the important role this factor played a century ago with the expansion of railroads and the advent of new communication techniques. This process continues to this day.

We show that if the size-skill relationship in the cross-section is negative, then an expansion of markets necessarily leads to a reduction in the skill premium. If, alternatively, the size-skill relationship is positive, then an expansion of markets might increase the skill premium. We show in numerical examples that an expansion of markets only raises the skill premium when the size-skill relationship is sufficiently pronounced. These implications are consistent with the historical record, including work by Goldin and Katz (2000) on the time pattern of the skill premium. A century ago, the skill premium was falling while the size-skill relationship was negative. In recent years, the skill premium has been rising and the size-skill relationship has been significantly positive.

In our model there is an analogy between the way capital relates to unskilled labor and

the way unskilled labor relates to capital. Capital can do relatively simple *mechanical* tasks that unskilled labor would otherwise do, but only if high setup costs are incurred. Analogously, unskilled labor can do complex *brain* tasks skilled labor would otherwise do, but only if high setup costs are incurred.[1] To explain, consider a simple mechanical task such as emptying the trash or moving a box from point A to point B. Unskilled labor has *general* ability to undertake such simple tasks. An unskilled worker hired just five minutes ago could first empty the trash and then move a box with virtually no training. It may be possible to obtain a machine to take out the trash, but this would in general require extensive setup costs, e.g. to construct a conveyer belt that would have to be designed to fit a particular space. Moreover, we expect that a different machine would have to be obtained to move the box. Machines tend to be specific in tasks they can be used for, at least as compared to the general ability of the human body to undertake simple mechanical tasks. Though advances in robotics and computer-controlled machinery have certainly made capital more flexible, it is still not as flexible as the human body.

Next consider a more complex task such as the management of the production process in an assembly line. A skilled worker with a degree in engineering can be put in charge of the production line; this worker has general knowledge to make appropriate decisions when problems arise. Alternatively, fixed costs can be incurred to redesign the production process to reduce the number of problems that arise. In conjunction with this redesign of the production process, manuals can be developed and protocols devised to make it possible for a unskilled worker managing the line to know what to do in the event of a (now rare) problem. The tradeoff here is that by paying fixed cost it may be possible to *design-out* the need for skill. Perhaps the most famous example of designing out skill is what Henry Ford did with the production process of automobiles. Following the principles of Frederick Taylor, he paid fixed cost to design a system that *routinized* tasks so that skilled craftsman could be replaced by unskilled laborers. This *de-skilling* through Taylorist principles that took place early in the 20th century has attracted much attention in the literature.[2]

---

[1] The view that skilled labor is not only more productive, but also more adaptable, in performing complex or new tasks goes back at least to Nelson and Phelps (1966). This idea has been revived recently by Galor and Moav (2000) and Aghion et al. (2002).

[2] See for example, Brown and Philips (1986). There is a literature in sociology that emphasizes the

To incorporate these ideas, we develop a model with the following features. To produce output at any plant a variety of tasks needs to be performed. The firm must decide which inputs (capital, unskilled labor, or skilled labor) should do which tasks. Tasks vary in *complexity* and more complicated tasks require more *setup costs*. Skilled workers, with their high level of general-purpose knowledge, have low setup costs. The setup costs of unskilled workers are higher, and the setup costs of capital are higher still. Thus capital can be thought of as an extreme form of unskilled labor.

In the optimal assignment of tasks there is a partition. Skilled workers are assigned complex tasks that would require extensive setup if undertaken by unskilled workers or capital. Capital is assigned the relatively easy to master tasks such as those that involve the movement of objects. Unskilled labor is assigned the in-between tasks. Thus on one margin capital substitutes for unskilled labor while on the second margin unskilled labor substitutes for skilled labor.

Historians know well that the size-skill relationship was negative in 1900. And labor economists know well that the size-skill relationship is positive today. But no previous analysis has tried to address both facts at the same time like we do here. In our theory, the relationship can go either way. Larger plants tend to substitute capital for unskilled labor and unskilled labor for skilled labor, because the larger scale makes it more worthwhile to pay fixed costs to lower marginal costs. Thus the net effect of plant size on the skilled labor share is ambiguous. We are able to derive a simple condition determining the direction of the net effect. What makes our result have content is that we relate the condition to how an expansion of markets or productivity growth affects the skill premium.

Our analysis of changes over time in the skill premium is closely related to previous work. Goldin and Katz (1999), Caselli (1999), Mobius (2000) and Mitchell (2001) all have models where changes in technology lead first to a reduction and then to an increase in the skill premium. And there is a large literature about capital skill complementarity. (See Griliches (1969), Krusell et al. (2000), Goldin and Katz (1998) and Autor, Levy, and Murname (2003).) What distinguishes our work from these papers is our attempt to connect changes in the skill premium to cross section relationships between plants of varying size.

---

de-skilling. See, in particular, Braverman (1974).

Furthermore, we show how the observed U-shaped pattern of the skill premium can be generated in a model even when there is no technological change. Expansion of markets and capital deepening, forces that raise plant size, are sufficient to obtain this result.

We note that an expansion of markets in our model is the same thing as an increase in trade. The channel through which increased trade affects the skill premium in our model is very different from the channel in the standard model, which is based on Hecksher-Ohlin arguments. In our paper, we interpret trade as simply the merging of multiple, identical countries, so it is simply a scaling up of market size. As a result, trade has no effect on the skill premium through the conventional channel. Here, trade allows plants to enjoy scale economies and as plants expand relative factor demands are affected. This is consistent with the plant level evidence from Bound, Berman and Griliches (1994), who find that the increase in demand for skilled labor is within industries and not due to a reallocation across industries, as in Hecksher-Ohlin. Our analysis of the effect of trade between similar countries is similar in spirit to Acemoglu (2003) who also identifies a channel (in his case endogenous technological change) through which increases in market size affects the skill premium.

2 Supporting Evidence

This section provides supporting evidence for assertions made in the introduction.

2.1 The Size-Skill Relationship

Suppose for now that an individual's pay can be used as a proxy for his or her skill. It is a well-known and robust fact that in today's economy larger plants have higher paid employes (Brown and Medoff (1989)).[3] It is not as well appreciated that the size-pay relationship has changed over time. Using micro data from the Census of Manufactures over the 1963-1986 period, Davis and Haltiwanger (1991) show there was a sharp upward trend in the relationship over this period. Idson (2001) also reports recent increases. Atack, Bateman, and Margo (2004) analyze Census micro data from the late nineteenth century and report a fundamentally different relationship between size and pay. In a simple linear regression of

---

[3]See also Oi and Idson (1999) for a survey of the literature.

log wage on log size, they find a *negative* relationship. In a regression with a quadratic term, they find an inverted U-shaped relationship that is first increasing and then decreasing.

These results are illustrated in Table 1. While output size is our preferred size measure since we use it in our theory, here we use employment to measure plant size because of data availability. The Census has published tabulations by employment size class in a consistent way over a long period enabling us to examine the long-run trend. To construct the table, we first calculated average pay for each plant size category by dividing total payroll in the category by total employment. For example, in the 1997 Census, average pay calculated this way for plants with 2,500 employees or more equaled $52,100. We then normalized by dividing by average pay in the entire manufacturing sector. The mean in 1997 was $33,900, so the normalized wage in 1997 in the 2,500 plus size category is 1.54=52.1/33.9, which is the figure reported in the table. Thus average pay in the in the largest size category is 54 percent higher than the average wage, a substantial premium.

Going from right to left in the table we move forward in time. Observe that for the largest two size classes, the premium increases monotonically with time. Thus the table replicates Davis and Haltiwanger's (1991) previous findings that apply for the time period beginning with the 1960s. But note there is also a substantial increase from 1947 to 1967 of 1.13 to 1.26. We conclude that the upward trend in the size wage premium began well before the 1960s.

The Census did not publish payroll by establishment size before 1947, so it is not possible to use Census tabulations to extend the table before that year. However, if we go back to 1880, we can use the 5 percent sample of the micro Census data collected by Atack and Bateman (1999).[4] This is the data used by Atack, Bateman, and Margo (2004) discussed above. With this sample data, it is possible to estimate payroll and employment by size class and extend the table to 1880, only we need to change the size groupings.[5] Plants in

---

[4]This data was obtained from raw manuscript data that is publicly available. The raw Census data becomes publicly available 72 years after it is collected.

[5]Our aim here is to constructing the variables for 1880 to make them as consistent as possible with the way the comparable variables are constructed for the later years. The mean is employment weighted, just as it is in the years. The mean also uses the sampling weights (Atack and Bateman oversampled small states). We use average annual employment in the denominator. Atack and Bateman make a correction

1880 were dramatically smaller than today or in 1947, as the median in 1880 had only 3 employees. This fact, combined with the fact that we have only a 5 percent sample, means there are very few observations in some of the cells in the original groupings. To deal with this fact, we have aggregated the larger size groupings and disaggregated the smallest grouping. With these groupings, we have over 200 observations in each cell.[6]

The results in Table 2 illustrate the inverted U-shaped pattern reported by Atack, Bateman, and Margo (2004). Pay rises with size for very small plants, but then flattens out over the range from 20 to 100 to 110 percent of the average wage. Pay then falls to 5 percent below the average wage, a drop of 15 percent. It is worth noting that if we were to look at the larger plants for which we have relatively few observations the drop is even larger. For the 50 plants with more than 250 employees, the average pay is 14 percent below the average. For the 15 plants with more than 500 employees, the average pay is 20 percent below the average.

Our interest here is in the size-skill relationship. Pay may depend upon other factors besides skill, so now we consider other measures of skill. Brown and Medoff (1989) and Troske (1999) both find that adding controls for worker quality such as education reduces the coefficient on establishment size in a wage regression. This indicates plant size is correlated with observable worker quality measures. Abowd and Karmarz (1999) used matched worker firm data and show that plant size is positively correlated with measures of worker quality. All of these studies use data from the 1970s or later.

In empirical work, it is common to classify production workers as unskilled workers and nonproduction workers as skilled workers. Table 3 presents nonproduction worker share by plant size, normalized by average nonproduction worker share. In 1997, the share for plants in the 2,500 category was .41 while the average share was .28, so the normalized share is 1.44=.41/.28. In 1997 the share in the largest size class was substantially larger than the average, but otherwise the relationship is relatively flat. In 1987, 1977, and 1967, the relationship is steep at the second highest class as well as the top class. There is a clear

_____

for the number of months the plant was operated. If we make this correction, it does not qualitatively affect the results.

[6]The cell counts are 11,750, 770, 263, and 209.

pattern that this relationship has steepened over time.

In the 1880 data, there is no classification by production/nonproduction worker status. Employment is divided up by men, women, and children. Since women mainly worked as production workers during this time period and since this is certainly true about children, we expect that the women/children share of employment to be positively correlated with the production worker share. Table 4 shows that this share increases sharply with plant size, with the largest plant having almost half of their workers being women and children. Goldin and Sokoloff (1982) show this same pattern holds in Census data from 1820-1850 and they argue that women and children factory workers in this period were unskilled workers.

Unlike the Censuses taken before it and after it, the 1890 Census collected information that was directly related to skills. Workers were classified into five categories based on the type of work that they did. Three of these categories can be regarded as *skilled* work. These include officers, clerks, and skilled workers. The other two categories, unskilled laborers and pieceworkers, can be classified as unskilled. Unfortunately, micro data from the 1890 Census is unavailable. However, state level data is published for 9 important industries.[7] For each state we calculated the percent of male workers in the state that were skilled workers as well as average employment size for plants in the state. Figure 1 plots these data points for the carriage and wagon industry as well as a regression line. There is a strong negative relationship. The R-squared of the regression is .54. An analogous pattern occurs in the other industries. In eight out of nine industries the regression line is negative and in the one case where it is not (the paper industry) the slope is not statistically significant. If we aggregate the data in a regression in logs with state and industry fixed effects, the estimated elasticity of skill share with respect to average size is -.16 with a standard error of .02. If we use average sales instead of average employment as a measure of size the elasticity falls somewhat to -.11 (standard error .02), but still continues to be quite high, especially considering the large observed variations in plant size. Given the clean definition of skill used here, we regard this as our strongest evidence that the size-skill relation was negative

---

[7]The tabulations are in Table 8 of U.S. Census (1895). The industries are: Agricultural Implements, Boots and Shoes, Carriages and Wagons, Dairy, Flour, Leather, Paper, Meatpacking, and Slaughtering. These nine industries accounted for 10 percent of 1890 employment.

in the late nineteenth century.

2.2   The Skill Premium

It is now well known that the skill premium has increased in recent decades (Katz and Murphy (1992)). Goldin and Katz (2000) show that if we look at the earlier 20th century, the skill premium, defined in various ways, was actually falling. Figure 2 shows their time series for the skill premium for the return to one year of college, over the century. We report the return to college because of the availability of a century-long series for it, but the primary features of the series are consistent with other measures of the return to skill and wage dispersion. For instance, the 90-10 or 80-20 wage ratio, looked at from the perspective of studies on various portions of the century, are similar: a fall in premium in the first half of the century, followed by an increase in the last quarter of the century. The return to high school, also reported in Goldin and Katz (2000) for the century, moves in a very similar pattern.

3   The Model

A fundamental component of the model is the existence of setup costs. Given the scale economies, the firms in our model have market power. In particular, firms sell differentiated products. Some firms have more desirable products than other firms, introducing variation in size across firms.

All firms face the same production technology. Each firm does a continuum of tasks that are ranked by the degree of complexity. More complex tasks require more setup. There are three inputs, capital, unskilled labor, and skilled labor that are in fixed supply to the economy. These inputs vary in the setup cost required to undertake any particular task.

3.1   Preferences and Technology

A representative household consumes a continuum of differentiated products indexed by $u \in [0, m]$. The differentiated goods are aggregated to a composite good through a CES

production function with elasticity of substitution σ:

$$Q = \left[\int_0^m \theta(u) q(u)^{\frac{\sigma-1}{\sigma}} du \right]^{\frac{\sigma}{\sigma-1}}. \tag{1}$$

Observe that the differentiated goods vary in the weight $\theta(u) > 0$ that they enter the CES function. Assume that higher u goods have higher weight, $\theta'(u) \geq 0$. With this specification, a consumer will buy more of the higher u good than a lower u good if the two goods have the same price. Normalize the scaling so that $\theta(0) = 1$. Assume $\sigma < 1$, so firms face inelastic demand. We make this assumption for convenience, as it simplifies the pricing formulas (with inelastic demand firms limit price).[8]

The technology for producing each differentiated product is the same. There are a continuum of tasks indexed by z on the unit interval $z \in [0, 1]$. Let $x(z)$ denote the level of activity of task z. The quantity of differentiated product produced given $x(\cdot)$ is CES with elasticity of substitution ω,

$$q = \left(\int_0^1 x(z)^{\frac{\omega-1}{\omega}} dz \right)^{\frac{\omega}{\omega-1}} \tag{2}$$

with the usual Cobb-Douglas specification in the limit $\omega = 1$.

There are three factors of production indexed by $j \in \{1, 2, 3\}$ in increasing order of "skill." Capital is $j = 1$, unskilled labor is $j = 2$, and skilled labor is $j = 3$. The total endowment in the economy of factor j is $\overline{X}_j$.

Undertaking each task entails a variable cost component and a fixed cost component. The variable cost is constant. Assume all three input types are equally efficient at the variable cost component in that one unit of the input is needed to undertake one unit of the task. Let $x_j(z)$ denote the use of factor j at task z. Then

$$x(z) = \sum_{j=1}^3 x_j(z)$$

is the total amount of task undertaken.

The three factors differ in setup cost. Assume that type 3 has zero setup. Let $\phi_2(z)$ be the setup cost for type 2 and $\phi_2(z) + \phi_1(z)$ be the setup cost of type 1, where $\phi_j(z) > 0$ for $z > 0$. Thus setting up factor 1 requires all the fixed costs needed to set up factor 2 plus

---

[8]We have also worked out the case where $\sigma > 1$ and obtain similar results.

additional fixed costs. Assume $\phi_j(z)$ is continuously differentiable and that for $j \in \{1, 2\}$ and $z > 0$,

$$\phi_j'(z) > 0. \tag{3}$$

Thus higher z goods require more setup. Higher z tasks are more complex.

The idea that capital has high fixed costs is not controversial; indeed, it is common to consider fixed costs for capital but none for labor. The fixed cost of capital has a natural interpretation in this model. In order to be able to do a particular task at a particular plant, it is essential that capital be specially designed for the task and be custom made. We recognize that one can come up with examples of capital goods that can be obtained "off the shelf" that may not incorporate much firm-specific design, e.g. a forklift truck. But most capital goods, e.g., assembly lines, machinery, factory buildings, do require firm-specific design investments and that is what we want to emphasize here.

Less standard is our treatment of fixed costs across different skill levels of labor. What we have in mind is that skills give workers general knowledge that allow them to move between tasks easily. Unskilled workers must be taught to do each task, at relatively high cost. Skilled workers can figure out how to do tasks without much difficulty. Another way to interpret the assumption is that it implies that unskilled workers are relatively efficient, compared to skilled workers, when they have very specialized jobs. The routinization of jobs on the assembly line allowed unskilled workers to be engaged where skilled workers had been necessary. The narrow scope of each job meant that even unskilled workers could pick up the necessary understanding to do the job properly.

Our model has the stark property that once fixed costs are paid, the inputs are perfect substitutes. We could add a second role for capital to allow for other possibilities. Suppose that the output of our differentiated-good production function had to be combined, Leontief, with capital (at no additional fixed cost) in order to make output. This would give a role to capital that is both non-substitutable with labor and at zero fixed cost. None of our results would change, since this second use of capital would simply be proportional to size. Moreover, we will consider in Section 6 an example where setup costs are prohibative for some range of tasks; in other words, output can only be produced if skilled workers are available to do some critical functions.

10

Our analysis will depend heavily on an elasticity concept. Define the setup cost elasticity for j to be:

$$\eta_j(z) \equiv \frac{\phi_j'(z)(1-z)}{\phi_j(z)}$$

This elasticity relates the percentage change in setup cost to the percentage change in the $1-z$ tasks above z. If the setup cost takes the following functional form,

$$\phi_j(z) = \alpha_j(1-z)^{-\theta_j} \tag{4}$$

then setup cost elasticity is constant, $\eta_j(z) = \theta_j$. We will be interested in cases of both constant and non-constant setup cost elasticity.

3.2 The Cost Minimization Problem

Before defining equilibrium, it is useful to study the cost minimization problem of a firm. Let the numeraire be the composite good and let (w_1, w_2, w_3) be the vector of input prices. Let w_j denote the price of a unit of factor j in terms of the numeraire. Since factors have identical productivity in the variable component of each task, but higher j factors have uniformly lower setup costs, it must be the case that, in any equilibrium, $w_1 < w_2 < w_3$.

Consider the cost minimization problem of a firm producing q units of output. The firm must choose how much of each task z to undertake and which factor to employ at this task. (Because of setup costs, each task is assigned to only one factor). Since the factors are equally productive at the variable component but higher j have higher wages, higher j are more costly in the variable component. But higher j have lower fixed costs since $\phi_j(z) > 0$, so there is a tradeoff. Since total setup increases in z, $\phi_j'(z) \geq 0$, it is immediate that the optimal assignment of tasks will consist of a pair of cutoff rules (z_1, z_2), such that factor 1 is assigned $z < z_1$, factor 2 is assigned $z \in (z_1, z_2)$ and factor 3 is assigned $z > z_2$. Within each range, the intensity is constant. Let x_j denote the intensity of factor j, in the range where it is used.

It is useful to decompose the cost minimization problem into two parts. The first part takes as given that the firm uses cutoffs (z_1, z_2) and determines the optimal mix across tasks. Recall that aside from the setup cost, the production function for the differentiated product is the constant returns CES form (2). Fixing (z_1, z_2) and given constant returns conditional

11

on these cutoffs, the cost minimizing input mix does not depend upon q. Let $\tilde{x}_j$ be the cost minimizing level at which to operate those tasks assigned to factor j, to produce a *single* unit of output. The demand $\tilde{x}_j$ is implicitly a function of the wages and the cutoffs. The per unit input demands satisfy the following problem,

$$\tilde{c}(z_1, z_2) \equiv \min_{\{(x_1, x_2, x_3) \text{ such that } q=1\}} \left[z_1 x_1 w_1 + (z_2 - z_1) x_2 w_2 + (1 - z_2) x_3 w_3 \right]. \tag{5}$$

The solution for this CES case is standard. The ratio of task intensities satisfies

$$\frac{\tilde{x}_j}{\tilde{x}_k} = \left(\frac{w_j}{w_k} \right)^{-\omega} \tag{6}$$

and the minimized cost is

$$\tilde{c}(z_1, z_2) = \left(z_1 w_1^{1-\omega} + (z_2 - z_1) w_2^{1-\omega} + (1 - z_2) w_3^{1-\omega} \right)^{\frac{1}{1-\omega}} \tag{7}$$

Note that minimized cost per unit is written as a function of the cutoffs but it also depends implicitly on the wages as well. Since $w_1 < w_2 < w_3$, it is immediate that

$$\frac{\partial \tilde{c}(z_1, z_2)}{\partial z_j} < 0$$

for $j \in \{1, 2\}$. Increasing the cutoff z_j replaces input $j + 1$ with input j which is less costly and equally productive in the variable component of the task. Thus increasing the cutoff lowers variable cost per unit.

The second part of the cost minimization problem is to choose the cutoffs z. Given a choice of cutoffs, the total expenditure on setup cost across all tasks is

$$f(z_1, z_2) = \left[\int_0^{z_1} \phi_1(z) dz + \int_0^{z_2} \phi_2(z) dz \right].$$

Observe that the firm pays $\phi_2(z)$ on all tasks done by either 1 or 2. In addition, it must pay $\phi_1(z)$ on all tasks done by 1. Given output level q, the firm chooses z_1 and z_2 to minimize the sum of variable costs plus setup costs,

$$c(q) = \min_{z_1, z_2} q\tilde{c}(z_1, z_2) + f(z_1, z_2) \tag{8}$$

This is a strictly convex problem since $\tilde{c}(z_1, z_2)$ is convex and $f(z_1, z_2)$ is strictly convex under assumption (3). The first order condition for the choice of z_j is

$$q\frac{\partial \tilde{c}(z_1, z_2)}{\partial z_j} + \phi_j(z_j) = 0. \tag{9}$$

12

The first term is the reduction in variable cost from increasing z_j, weighted by output q. The second term $\phi_j(z_j)$ is the marginal increment in total fixed cost f from shifting task z_j away from factor $j+1$ to factor j.

For later use, we rewrite the FONC for the choice of cutoff z_j. In the Cobb Douglas case ($\omega = 1$) the slope is

$$\frac{\partial \tilde{c}}{\partial z_j} = \tilde{c} \left[\ln w_j - \ln w_{j+1} \right]. \tag{10}$$

In the Cobb-Douglas case, total expenditure on each task is the same. With a unit measure of tasks, total expenditure overall must equal expenditure on any individual task. In particular, it equals that on a task assigned to factor 3,

$$q\tilde{c} = w_3 x_3. \tag{11}$$

Substituting (10) and (11) into the first order condition (9) for the choice of z_j yields

$$w_3 x_3 \left[\ln w_j - \ln w_{j+1} \right] + \phi_j(z_j) = 0. \tag{12}$$

Using an analogous derivation for the general ω case, we can rewrite the first-order necessary condition as[9]

$$\frac{1}{1-\omega} w_3^\omega x_3 \left(w_j^{1-\omega} - w_{j+1}^{1-\omega} \right) + \phi_j(z_j) = 0. \tag{13}$$

3.3   Equilibrium

Before discussing equilibrium, we first note a consequence of our assumption that for factor 3 (skilled labor), there is no setup cost for any task. If one unit of skilled labor were equally divided across all the tasks, one unit of output would result. With no setup costs, producing output in this way is constant returns to scale, and the cost per unit of output is w_3. We assume this constant returns avenue for obtaining output is freely available to consumers.

Now consider the behavior of producers. There is a single producer of each differentiated good u. A producer cannot charge consumers a price greater than w_3 since consumers would use the constant returns alternative just discussed to obtain the product at a cost of w_3 per unit. By the assumption that $\sigma < 1$, demand is inelastic for prices below w_3. Hence, it is

---

[9]This derivation, contained in the appendix, uses the marginal rate of substitution condition to write x_1 and x_2 as functions of w_3 and x_3.

immediate that producers will set a limit price up to the consumers reservation price of w_3; i.e., $p(u) = w_3$ for each differentiated product.

We exploit this structure of a constant limit price across all products to simplify our definition of an equilibrium. For simplicity of the definition, define $n_j(u)$ to be the number of tasks that the producer of product u assigns factor j:

$$n_1(u) \equiv z_1(u)$$
$$n_2(u) \equiv z_2(u) - z_1(u)$$
$$n_3(u) \equiv 1 - z_2(u)$$

Definition 1 *An equilibrium is a list of functions $(p(u), q(u), z_j(u), x_j(u))$ and factor prices w_j such that*

(1) *Limit Pricing*: $p(u) = w_3$

(2) *Marginal Rate of Substitution Condition*: $q(u) = \theta(u)^\sigma q(0)$

(3) *Cost Minimization*

$$(z_1(u), z_2(u)) = \arg\min_{z_1, z_2} q(u)\tilde{c}(z_1, z_2, w_1, w_2, w_3) + f(z_1, z_2)$$
$$x_j(u) = q(u)\tilde{x}_j(z_1(u), z_2(u), w_1, w_2, w_3)$$

(4) *Differentiated Goods Market Clearing*

$$\left(n_2(u)(x_1(u))^{\frac{\omega-1}{\omega}} + n_2(u)(x_2(u))^{\frac{\omega-1}{\omega}} + n_3(u)(x_3(u))^{\frac{\omega-1}{\omega}} \right)^{\frac{\omega}{\omega-1}} = q(u)$$

(5) *Factor Market Clearing*:

$$\int_0^m n_j(u)x_j(u)du = \bar{X}_j, \text{ for each } j,$$

(6) *Household's Budget Constraint*

$$Q - \int_0^m f(z_1(u), z_2(u))du = \bar{X}_1 w_1 + \bar{X}_2 w_2 + \bar{X}_3 w_3 + \Pi$$

Condition (2) follows from consumer utility maximization. It is the marginal rate of substitution condition between good u and good 0. (recall $\theta(0) = 1$ and that prices $p(u)$ are constant for all u). Since it is optimal for firms to set the limit price, the analysis of

14

the firm's problem reduces to minimizing the cost of producing the quantity demanded at the limit price. This is condition (3). Conditions (4), (5), and (6) are market clearing conditions. The left hand side of the household's budget constraint (6) reflects the fact that households consume all of the final good output (Q) except that used in the production of fixed costs. Profits from the differentiated goods firms are denoted Π.

It is easy to derive an explicit formula for the differentiated product price p (and also w_3 since $p = w_3$). Let $\tilde{q}(u)$ denote the cost minimizing quantity of differentiated good u required to produce a *single* unit of the composite. From the Marginal Rate of Substitution Condition we have $\tilde{q}(u) = \theta(u)^\sigma \tilde{q}(0)$. Combining this with the production function for one unit of the composite gives us

$$\tilde{q}(u) = \theta(u)^\sigma \left[\int_0^m \theta(u)^\sigma du \right]^{-\frac{\sigma}{\sigma-1}}$$

Since the composite is the numeraire, the total price of this cost minimizing bundle must equal one. This allows us to solve out for the equilibrium differentiated product price p,

$$p = \left[\int_0^m \theta(u)^\sigma du \right]^{\frac{1}{\sigma-1}}. \tag{14}$$

We next present our result for existence of equilibrium. Here we restrict attention to the case where there are a finite number of different product types.

Proposition 1 *Suppose the set $\{\theta(u), u \in [0, m]\}$ is finite. An equilibrium exists.*

Proof. See Appendix. ∎

4 Factor Allocation and Plant Size

In this section we consider the relationship between plant size and factor allocation. In this economy, more desirable goods (higher u) have higher q. All firms in the economy face the same wages and have access to the same technology. So in order to study how factor allocation depends upon plant size, we need to study how the cost-minimizing factor demands vary with q.

Our first step is to determine how the cutoffs z_1 and z_2 vary with plant size. The two cutoffs solve the two first-order necessary conditions,

$$q\frac{\partial \tilde{c}(z_1, z_2)}{\partial z_1} + \phi_1(z_1) = 0 \tag{15}$$

$$q\frac{\partial \tilde{c}(z_1, z_2)}{\partial z_2} + \phi_2(z_2) = 0$$

Our result is

Lemma 1 *In the solution to the cost minimization problem (8) z_1 and z_2 both increase in q.*

Proof. Observe from (10) for $\omega = 1$ and from (28) in the appendix for $\omega \neq 1$ that

$$\frac{\partial \tilde{c}}{\partial z_2}\frac{\partial^2 \tilde{c}}{\partial z_1 \partial z_2} - \frac{\partial \tilde{c}}{\partial z_1}\frac{\partial^2 \tilde{c}}{\partial z_2 \partial z_2} = 0. \tag{16}$$

Totally differentiating the two first-order conditions (15) for z_1 and z_2, using Cramer's rule and (16) yields

$$\frac{\partial z_1}{\partial q} = -\frac{1}{|H|}\frac{\partial \tilde{c}}{\partial z_1}\phi_2' = \frac{1}{|H|}\frac{1}{q}\phi_1\phi_2' > 0 \tag{17}$$

where H is the Hessian, $|H| > 0$, and where the first-order necessary condition is used to substitute in ϕ_1/q. Analogously,

$$\frac{\partial z_2}{\partial q} = \frac{1}{|H|}\frac{1}{q}\phi_2\phi_1' > 0. \tag{18}$$

∎

An increase in the target quantity q places more weight on the benefit of cost reduction for increasing the cutoffs. So it is intuitive that a higher q would increase the optimal cutoffs. With greater economies of scale, a larger firm substitutes capital for unskilled labor and unskilled labor for skilled labor, raising the complexity (i.e. average z) of the tasks undertaken by both factors.[10]

Define the *size-skill* relation as the ratio of the demand for skilled and unskilled workers as a function of plant size:

$$s(q) \equiv \frac{X_3(q)}{X_2(q)} = \frac{(1 - z_2(q))\, x_3(q)}{(z_2(q) - z_1(q))\, x_2(q)}$$

$$= \frac{1 - z_2(q)}{z_2(q) - z_1(q)}\left(\frac{w_3}{w_2}\right)^{-\omega}$$

---

[10]Eisfeldt and Rampini (2004) show that larger firms purchase relatively less used and more new capital than smaller firms. Our model is consistent with this fact if more complex (i.e. higher z) tasks are more likely to require new capital, which seems plausible.

16

Observe that the ratio of intensities $\frac{x_3}{x_2}$ is the same for firms of different sizes because this depends upon the ratio of factor prices through (6), the same for all firms. Thus to understand how the skill ratio varies with q, we need only to look at the behavior of the ratio of numbers of tasks assigned.

It is immediate that there are two conflicting forces at work. Higher q raises z_1, so capital replaces unskilled workers. It raises z_2, so unskilled workers replace skilled. The net effect on the size-skill relationship is ambiguous. It turns out to depend upon a simple comparison of the elasticity of setup costs, as demonstrated in the following proposition.

Proposition 2 *The slope $s'(q)$ of the size-skill relationship is positive, zero, or negative as $\eta_2(q)$ is greater than, equal to, or less than $\eta_1(q)$.*

Proof. The slope $s\prime(q)$ has the sign of

$$-\frac{dz_2}{dq}(z_2 - z_1) - \left(\frac{dz_2}{dq} - \frac{dz_1}{dq}\right)(1 - z_2)$$
$$= \frac{dz_1}{dq}(1 - z_2) - \frac{dz_2}{dq}(1 - z_1)$$

Substituting in (17) and (18) and multiplying by $q|H|$, the slope $s\prime(q)$ has the sign of

$$\phi_1\phi_2'(1 - z_2) - \phi_2\phi_1'(1 - z_1)$$

or, dividing by $\phi_1\phi_2$,

$$\eta_2 - \eta_1.$$

∎

It is intuitive that the relative setup cost elasticities should play a crucial role. If η_1 is small relative to η_2, it is relatively cheaper to shift the z_1 margin (capital replacing unskilled labor) than the z_2 margin (unskilled labor replacing skilled) so the size-skill relationship increases. Note that our model does not have an unambiguous prediction for the size-skill relationship, which gives it a chance to replicate the experience of the changes from the late 19th to the late 20th centuries.

5 Factor Prices

This section examines the impact of changes in the stock of endowments on factor prices. In particular, how does the skill premium change when the the size of the economy increases, through trade or productivity growth? We proceed in two parts. The first part derives a characterization for a special case of the model. The second part uses the charactization to derive the comparative static results.

5.1 A Characterization of a Limiting Case

To make the general equilibrium analysis as simple as possible, this section examines the limiting case where all plants are the same size. By continuity, our results apply when plant sizes differ, but a sufficiently large amount of the probability weight is concentrated near a particular plant size type. By focusing on this limiting case, we are able to simplify the general equilibrium analysis down to a simple equation that has a natural interpretation as demand equating supply for capital.

Assume that the total measure of products is $m = 1$, so that $p = 1$ from formula (14) and therefore $w_3 = 1$. Assume further that $\omega \geq 1$. The $\omega < 1$ analysis is the same except that case requires us to take into account the possibility of a corner solution where all of factor 1 or 2 is completely disposed of. With $\omega \geq 1$, things are simpler since all factor prices will be strictly positive and there is no disposal.

Let z_1 and z_2 denote the cutoffs of the representative firm. We will construct an equilibrium by beginning with an arbitrary z_2 and derive the equilibrium demand for capital $\hat{D}_1(z_2)$ that is consistent with this level of z_2. We then compare demand to the exogenous supply $\overline{X}_1$. An equilibrium is where demand equals supply, $\hat{D}_1(z_2) = \overline{X}_1$. This demand equal supply condition will prove useful for the analysis of general equilibrium effects, since there is a close relationship between the skill premium and the equilibrium z_2.

Given a cutoff z_2 and given that in any equilibrium factor 3 is equally distributed among the unit measure of firms, the intensity of tasks undertaken by factor 3 at each firm must then be

$$\hat{x}_3(z_2) = \frac{\overline{X}_3}{(1 - z_2)},$$

18

which we write as an explicit function of z_2. Substituting this into the FONC (13) for z_2, noting that $w_3 = 1$ and rearranging yields w_2 as an explicit function of z_2,

$$\hat{w}_2(z_2) = \left[1 + (\omega - 1)\frac{1 - z_2}{\overline{X}_3}\phi_2(z_2)\right]^{-\frac{1}{\omega - 1}}. \tag{19}$$

Observe our assumption that $\omega \geq 1$ implies that $0 < \hat{w}_2(z_2) < 1$ for all z_2.

Next we back out the z_1 that is implied by the marginal technical rate of substitution condition. Cost minimization implies

$$\frac{x_3}{x_2} = w_2^\omega$$

But given that a measure $1 - z_2$ tasks are assigned to factor 3 and $z_2 - z_1$ to factor 2, the implied levels of x_3 and x_2 yield

$$\frac{\overline{X}_3}{\overline{X}_2}\frac{z_2 - z_1}{1 - z_2} = w_2^\omega. \tag{20}$$

Solving this leads to an expression for z_1 as a function of z_2,

$$\hat{z}_1(z_2) = z_2 - (1 - z_2)\frac{\overline{X}_2}{\overline{X}_3}\hat{w}_2^\omega(z_2) \tag{21}$$

Observe this is negative at $z_2 = 0$. It is also immediate that

$$\lim_{z_2 \to 1} \hat{z}_1(z_2) = 1.$$

So define z_2° by

$$z_2^\circ = \max\{z_2: \text{ such that } \hat{z}_1(z_2) = 0\}$$

Observe that $\hat{z}_1(z_2) > 0$, for all $z_2 > z_2^\circ$. This is the range of z_2 that we will consider.

The next step is to determine the level of w_1 that is implied by the first-order condition in the choice of z_1. This equals

$$\hat{w}_1(z_2) = \left[w_2^{1-\omega} + (\omega - 1)\frac{1 - z_2}{\overline{X}_3}\phi_1(\hat{z}_1(z_2))\right]^{-\frac{1}{\omega - 1}}, \tag{22}$$

which satisfies $0 < \hat{w}_1(z_2) < \hat{w}_2(z_2)$, given $\omega \geq 1$.

Having determined all this, we can calculate the intensity used by factor 1,

$$\hat{x}_1(z_2) = \hat{x}_3(z_2)\hat{w}_1(z_2)^{-\omega}.$$

19

Putting this all together leads to what we call the *demand for capital given* z_2,

$$\hat{D}_1(z_2) \ = \ \hat{z}_1 \hat{x}_1 \tag{23}$$
$$= \ \hat{z}_1 \hat{x}_3 \hat{w}_1^{-\omega}.$$

At this point we make two observations. First, by definition of the cutoff z_2°, $\hat{D}_1(z_2^\circ) = 0$. Second, since z_1 goes to 1 and w_1 is bounded above by 1, and since $\hat{x}_3$ goes to infinity, the limit of demand is infinite near z_2 equal to one,

$$\lim_{z_2 \to 1} \hat{D}_1(z_2) = \infty.$$

These two observations and continuity of $\hat{D}_1(z_2)$ imply an equilibrium exists where demand equals supply,

$$\hat{D}_1(z_2) = \overline{X}_1.$$

If demand $\hat{D}_1(z_2)$ is everywhere upward sloping in z_2 the equilibrium is unique. It is intuitive that it should be upward sloping. An increase in z_2 means fewer tasks are assigned to factor 3, increasing work for the other two factors. For the case of constant setup cost elasticity given by equation (4), we can show that if $\eta_1 \geq 1$ and $\eta_2 \geq 1$ then demand is strictly monotonic. Figure 3 illustrates this for the parameters $\eta_1 = \eta_2 = 1$ and $\bar{X}_2 = \bar{X}_3 = 1$.[11] The demand curve as well as two supply curves are illustrated. Observe that if the supply of capital $\bar{X}_1 = 1$, then the equilibrium $z_2 = .35$. Skilled labor undertakes a fraction $1 - z_2 = .65$ of all the tasks, but accounts for only $\frac{1}{3} = \frac{\bar{X}_3}{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}$ of the labor stock. The other two factors are concentrated in disproportionately fewer tasks to keep setup costs low.

While an upward sloping relationship appears to be the typical case, for extreme parameters it is possible to construct examples where a portion of the demand relationship is non-monotonic. Figure 4 is an example with $\eta_1 = \eta_2 = .01$ and $\alpha_1 = \alpha_2 = 30$ where this occurs. In such a case, depending on where the supply line cuts, there may be multiple equilibria. If there are multiple equilibria, there will be at least one *irregular* equilibrium where demand cuts supply from above as well *regular* equilibria where demand cuts supply from below. For our comparative statics analysis, we will strict attention only to regular equilibria:

---

[11]The multiplicative constant for setup cost is $\alpha_1 = \alpha_2 = 1$.

Regularity Condition. Restrict the set of equilibria to include only those z_2^e that satisfy

$$\frac{d\hat{D}_1(z_2^e)}{dz_2} > 0,$$

in addition to $\hat{D}_1(z_2^e) = \bar{X}_3$.

This is analogous to a stability condition. Note a regular equilibrium will always exist. In our comparative statics analysis, in the neighborhood of regular equilibrium z_2^e we will assume a continuous equilibrium selection around z_2^e. Our comparative statics results are meant to be interpreted as local results.

5.2 Market Expansion and Growth

Consider scaling up all inputs in a proportation way,

$$\bar{X}_j = \lambda \xi_j$$

for a vector of constants (ξ_1, ξ_2, ξ_3). As we increase the scaling parameter λ, we make the economy bigger. We have in mind two interpretations of the scaling parameter λ. First, an increase in trade possibilities is equivalent to an increase in λ. If two separate identical economies are merged together by the expansion of trade, it is identical to a doubling of λ. Second, an increase in total factor productivity is equivalent to an increase in λ.

The question addressed in this subsection is how a change in λ affects factor prices. The first step in our analysis is the following lemma which determines how the equilibrium cutoff z_2^e varies with λ.

Lemma 2 *The cutoff z_2^e in a regular equilibrium strictly increases the scaling parameter λ.*

Proof. Define adjusted demand in a way that removes the scaling,

$$\widehat{AD}_1(z_2, \lambda) = \frac{\hat{D}_1(z_2^e, \lambda)}{\lambda} = \hat{z}_1 \frac{1}{\hat{w}_1} \frac{\hat{x}_3}{\lambda} = \hat{z}_1 \frac{1}{\hat{w}_1} \frac{\xi_j}{1 - z_2}.$$

Analogously, adjusted supply is

$$\widehat{AS}_1(z_2, \lambda) = \frac{\overline{X}_1}{\lambda} = \xi_1.$$

In equilibrium adjusted demand equals adjusted supply. Since adjusted supply is a constant, it is sufficient to show that $\widehat{AD}_1(z_2, \lambda)$ decreases in λ. Observe first that fixing z_2, $\hat{w}_2(z_2, \lambda)$ in (19) strictly increases in λ. From (21), $\hat{z}_1(z_2, \lambda)$ strictly decreases in λ. Since $\phi'_1(z_1) > 0$ by assumption, all results imply that $\hat{w}_1(z_2, \lambda)$ in (22) strictly increases in λ. Together, these results imply that $\widehat{AD}_1(z_2, \lambda)$ is strictly decreasing in λ. ∎

The lemma shows that when all factors increase proportionately, the number of tasks $1 - z_2$ done by skilled labor decreases. Skilled labor is allocated a disproportionate number of tasks to save on fix costs. As the scale of the economy increases, this disadvantage of capital and unskilled relative to skilled labor decreases.

The next step is to determine what happens to factor prices. The effect on the skill premium depends in large measure on the shape of the setup cost functions. When λ increases, there are more resources available to produce the fixed set of products. As a result, plants grow, which tends to make firms want to substitute capital for unskilled labor and unskilled labor for skilled labor. The change in the demand for unskilled labor (relative to skilled labor) depends on how costly it is, in terms of the ϕ functions, to substitute at the two margins z_1 and z_2. Our earlier analysis showed that the size-skill relationship provided information about the relative shapes of ϕ_1 and ϕ_2. Our next result ties this together.

Proposition 3 *Suppose at given market size λ° the size-skill relationship is nonpositive. Then an increase in market size λ around λ° strictly increases w_2, so the skill premium falls.*

Proof. We will show that if w_2 does not strictly increase in λ around λ° then $\eta_1(z_1^\circ) < \eta_2(z_2^\circ)$ must hold. Proposition 2 then implies that the size-skill relationship must be strictly positive, which is sufficient to prove the claim.

We begin with some preliminaries. First, we show that z_1 must strictly increase. Recall formula (21) for z_1:

$$z_1 = z_2 - (1 - z_2) \frac{\xi_2}{\xi_3} w_2^\omega \tag{24}$$

Differentiating with respect to λ yields

$$\frac{dz_1}{d\lambda} = \frac{dz_2}{d\lambda}(1 + \frac{\xi_2}{\xi_3} w_2^\omega) - \frac{\xi_2}{\xi_3}(1 - z_2)\omega w_2^{\omega-1} \frac{dw_2}{d\lambda} > 0.$$

This is strictly positive since $\frac{dz_2}{d\lambda} > 0$ from Lemma 2 and since $\frac{dw_2}{d\lambda} \leq 0$ by hypothesis. Since both z_1 and z_2 strictly increase, it is then immediate from a formula analogous to (20) for w_1^ω that w_1 strictly increases.

Next observe that

$$\frac{\frac{dz_1}{d\lambda}}{\frac{dz_2}{d\lambda}} \geq \left(1 + \frac{\overline{X}_2}{\overline{X}_3} w_2^\omega\right) \tag{25}$$
$$= 1 + \frac{\overline{X}_2}{\overline{X}_3}\frac{\overline{X}_3}{\overline{X}_2}\frac{z_2 - z_1}{1 - z_2} = \frac{1 - z_1}{1 - z_2}.$$

The weak inequality follows from differentiating (24) and using the fact that w_2 is weakly decreasing by hypothesis. The second line substitutes in w_2^ω from (20).

We now turn to the main step of the proof. From (13), the two first-order necessary conditions for the choice of z_1 and z_2 can be written

$$\frac{1}{1 - \omega}x_3\left(w_2^{1-\omega} - w_1^{1-\omega}\right) = \phi_1(z_1)$$
$$\frac{1}{1 - \omega}x_3\left(1 - w_2^{1-\omega}\right) = \phi_2(z_2)$$

Dividing yields

$$\frac{\left(1 - w_2^{1-\omega}\right)}{\left(w_2^{1-\omega} - w_1^{1-\omega}\right)} = \frac{\phi_2(z_2)}{\phi_1(z_1)}$$

Since w_2 weakly decreases while w_1 strictly increases, the left-hand side must strictly increase. Differentiating the right-hand side and multiplying through by a positive factor, we have

$$0 < \phi_2'(z_2)\phi_1(z_1)\frac{dz_2}{d\lambda} - \phi_1'(z_1)\phi_2(z_2)\frac{dz_1}{d\lambda}$$

A rearrangement including multiplying through by $(1 - z_2)$ yields

$$0 < (1 - z_2)\frac{\phi_2'(z_2)}{\phi_2(z_2)} - (1 - z_1)\frac{\phi_1'(z_1)}{\phi_1(z_1)}\frac{(1 - z_2)}{(1 - z_1)}\frac{\frac{dz_1}{d\lambda}}{\frac{dz_2}{d\lambda}}$$
$$\leq \eta_2(z_2) - \eta_1(z_1),$$

where the weak inequality uses (25). ∎

This result provides a link between the size-skill relationship and the effect of market expansion and growth on the skill premium. In particular, a negative or zero size-skill relationship implies that expansion of markets reduces the skill premium. Chandler (1990)

has argued that in the late nineteenth century, the expansion of markets brought about by advances in transportation led firms to increase plant size and adopt mass production technologies. Goldin and Katz (1998) have argued that the advent of mass production led to reductions in the skill premium. Our theory connects these two separate points, using the size-skill relationship to draw the connection. The negative size-skill relationship of the late nineteenth century was a signal that expansion of markets would *necessarily* have a negative impact on the skill premium.

For some parameters of our model, an expansion of markets can increase the skill premium, as we discuss in the numerical example below. When that happens, our result above tells us that the size-skill relationship must be strictly positive. The process of market expansion has continued throughout the twentieth century. The positive size-skill relationship that has become increasingly steep implies that the effect of an expansion of markets today on the skill premium *may* be positive.

We have assumed so far that factor supplies are exogenous. Another option would be to make factor supply respond to the relative prices of factors. For instance, consider the decision to develop skills. Suppose that at a particular point in time the total stock of workers, skilled and unskilled, is fixed at $\bar{X}_2 + \bar{X}_3 = L$. Let workers choose to be unskilled and make w_2 or be skilled and make $w_3 = 1$. Let θ denote the cost of attaining skills (in terms of the consumption good) and assume a continuous distribution $F(\cdot)$ of θ. The indifferent worker $\bar{\theta}$ will be where $w_2 = 1 - \bar{\theta}$ and the resulting factor supplies will be $\bar{X}_3 = F(\bar{\theta})L$ and $\bar{X}_2 = (1 - F(\bar{\theta}))L$. Our analysis implicitly fixes $\bar{\theta}$; however, it is conceptually straightforward to relax that assumption. We conjecture that Proposition 3 can be extended to this more general setup. Holding fixed factor supplies fixed at what they would be at $\bar{\theta}$, we know from this result that if the size skill relationship is negative, an expansion of the market raises the unskilled wage, say from $w_2(\bar{\theta})$ to $w_2'(\bar{\theta})$. Then, obviously, agents have less incentive to acquire skills, so when we make factor supply endogenous we expect $\bar{\theta}' < \bar{\theta}$. Of course, this cannot completely undo the increase in w_2; if $w_2(\bar{\theta}) = w_2'(\bar{\theta}')$, then $\bar{\theta}' < \bar{\theta}$ implies that the last agent $\bar{\theta}'$ to acquire skills did so at a gain. As usual, quantity changes can mitigate price effects but not eliminate them.

As noted above, we can interpret the increase in market size through higher λ as coming

from in the integration of two similar economies that were previously separated by trade barriers. Or it can come from the expansion of an existing economy through proportionate growth in inputs or in total factor productivity growth (growth in "effective inputs"). Another case to consider is where there is exogenous labor augmenting technical change, where the effective labor inputs scale up exogenously by a factor λ, $\bar{X}_2 = \lambda \xi_2$, and $\bar{X}_3 = \lambda \xi_3$, but there is no exogenous technical change for capital. In standard growth models that allow savings, the stock of capital will grow endogenously to match the exogenous growth in labor productivity. So we expect that in an economy with exogenous labor productivity growth and capital accumulation the results would be similar to what we have here.

6　An Example

This section presents a particular example of our model that exhibits the broad trends that we discussed in the introduction. The example is meant to capture the following scenario. In the late nineteenth century, mechanical tasks such as those involved in the manufacture of a vehicle were undertaken by both unskilled workers and craftsmen who were highly trained and who were considered skilled workers. As market size expanded, it was relatively easy to substitute unskilled workers for skilled workers at these mechanical tasks. Over time, unskilled workers pushed skilled workers out of mechanical tasks altogether; skilled workers shifted entirely to working on nonmechanical tasks like management. While in some cases it may be possible to substitute unskilled for skilled in these "brain tasks" (like the computer helpline discussed in the introduction), in general this substitution is quite difficult. As a result, unskilled labor today is faring relatively poorly on its two fronts. In its rear front where it competes with capital, it is losing out as capital displaces its mechanical tasks. But in its forward front where it competes skilled labor, it is unable to displace skilled labor at management-like tasks.

In this example, we keep the technology the same, including the ϕ_j functions. We also hold fixed the ratio of skilled workers to unskilled workers. We recognize that there have been major technological changes over the course of a century that would affect the ϕ_1 and ϕ_2 functions and these changes would affect the skill premium. We also recognize that their

25

have been changes in the ratio of skilled to unskilled through schooling and these would affect the skill premium. Our intent here is to abstract from these well-known forces so that we can focus on the potential role of market expansion and capital deepening in accounting for the observed patterns.

The particular assumptions we make about how the economy changes over time are:

$$X_1(t) = \xi_1 e^{\gamma t + \kappa t}$$
$$X_2(t) = \xi_2 e^{\gamma t}$$
$$X_3(t) = \xi_3 e^{\gamma t}.$$

for $\gamma > 0$ and $\kappa > 0$. Here, skilled and unskilled labor grow at the same rate γ so the ratio stays fixed. Capital grows at a higher rate that exceeds labor by an amount κ. We view the γ parameter as capturing the force of market expansion and productivity growth; the κ parameter captures capital deepening.

The specification of the setup cost functions is crucial for determining the dynamics. If we were to choose the constant elasticity specification (4) for both ϕ_1 and ϕ_2, Proposition 2 would then imply the sign of the size-skill relationship would be constant over time. We would have no hope of capturing the switch in sign, from negative to positive, that has occurred in the data. This leads us to adopt the following non-constant elasticity specification of ϕ_2,

$$\phi_2(z) = \alpha_2 (1 - z - \beta_2)^{-\theta_2},$$

for $\beta_2 > 0$. Here the elasticity is

$$\eta_2(z) = \theta_2 \left[\frac{1-z}{1-z-\beta_2} \right], \tag{26}$$

which increases in z and goes to infinity as z goes to $1 - \beta$. This captures the idea that as we move up in the hierarchy of tasks, it becomes increasingly more difficult to substitute unskilled for skilled. The set of tasks $[1 - \beta, 1]$ at the top end of the complexity scale can be thought of as management tasks that only skilled workers are capable of doing.

For simplicity, we assume that ϕ_1 takes the constant elasticity form (4). Having the elasticity of ϕ_2 change relative to that of ϕ_1 is enough to drive the patterns we will exhibit in the model.

As t goes to infinity, the quantity of each factor gets arbitrarily large, since $\gamma > 0$. Moreover, capital's share of the productive units gets arbitrarily large, since $\kappa > 0$. Hence $z_2(t)$ must approach $1 - \beta$. (To ensure that $z_2(t)$ goes to $1 - \beta$ is why we set $\kappa > 0$. If we set $\kappa = 0$, this will also happen if β is small enough.) As we go back in time and t goes to minus infinity, it is clear that $z_2(t)$ must go to zero, since setup cost becomes prohibitive. Assume

$$\eta_2(0) = \frac{\theta_2}{1 - \beta_2} < \min\{\theta_1, 1\}$$

Then if we go back far enough in time $\eta_2(t) < \eta_1 = \theta_1$. From Proposition 2, the size-skill relationship is negative. From Proposition 3, the effect of the market expansion (γ) on the skill premium is negative. We can show that $\eta_2(0) < 1$ implies that an increase in capital lowers the skill premium. Together, this implies that far enough back in time the skill premium is falling.

The setup elasticity η_2 for factor 2 given by (26) strictly increases over time and goes to infinity. There is a critical time period $\hat{t}$ where $\eta_2 = \eta_1$. Before this point the size-skill relationship is positive, after this point it is negative. There is also a point where η_2 exceeds unity. In simulations of numerical examples we have found that far enough into the future the skill premium increases in market size. Unskilled labor bangs into the constraint that it cannot cut management jobs. But expansion of markets enables capital to cut into tasks done by unskilled workers.

Figure 5 plots the evolution over time of the skill premium (defined by $(w_3 - w_2)/w_2$) and the size-skill relationship, defined as the elasticity of the skill share with respect to output size, $\varepsilon_s \equiv \frac{s'(q)q}{s(q)}$. The numerical example satisfies the restrictions above. The monotonic increase over time in the size-skill relationship and the U-shape of the skill premium is a robust pattern across various parameter values satisfying these restrictions. Note that at the bottom of the U where the skill premium begins to rise, the size-skill relationship is strictly positive (as must be the case from Proposition 3). This pattern is consistent with the U.S. experience, as the size-skill relationship was small, but strictly positive at midcentury (see Tables 1 and 3).

Analogous to the size-skill relationship $s(q)$, we can define the size-capital relationship $k(q) = X_1(q)/(X_2(q) + X_3(q))$. This is the ratio of capital total labor (size-skill is the ratio

27

of skill to total labor). Figure 5 plots this relationship in elasticity form in the same graph with the size-skill relationship. Like the size-skill relationship, the size-capital relationship is increasing over time. But while the size-skill relationship is negative early on, the size-capital relationship is always positive. It is intuitive that the size-capital relationship should be positive; our working paper Holmes and Mitchell (2003) presents conditions under which this is in fact the case. In particular, we show that a positive size-skill relationship is sufficient (but not necessary) for a positive size-capital relationship. It also presents empirical results that the size-capital relationship has been increasing over time, but never negative, consistent with Figure 5, and the fact that the size-skill relationship is eventually positive in the data.

In the discussion so far, we have focused on an education-based measure of skill. In the past four decades there has been an increase in the residual inequality, variance in wages not accounted for by education or other measured characteristics (Juhn, Murphy, and Pierce (1993)). In our paper, high-skill workers have the ablity to do a variety of tasks, including complex ones, with relatively low setup costs. This kind of general knowledge is something we expect people to obtain in college. But it can also be derived from innate characteristics and experiences that may not be easy to measure. Suppose in our model it is impossible to measure skill. But we can measure pay and higher skill workers will receive higher pay. Then in the model, if residual inequality is increasing, then it must be the case that the size pay relationship is positive.

7 Conclusion

We have introduced a model of task assignment at the plant level. Increased scale of production makes it worthwhile to substitute capital for unskilled labor and unskilled labor for skilled labor. The model delivers several empirically relevant results. It connects the positive size-skill relationship and positive size-capital relationship of the late 20th century to one another and it connects the effect of market expansion on the skill premium to the size-skill relationship. The model is flexible enough to account for the differences between early industrialization and post-Fordism while at the same time connecting time series movements to observable cross section implications. Indeed, as we showed in the last section, the

model is even capable of delivering the qualitative trends without any changes in technology, but simply through changes in the stock of inputs which effectively increases the size of the market.

We abstract from changes in technology in order to isolate the impact of increases in market size. We recognize that changes in technology have played a central role in the issues we are looking at. Some of these technological changes reinforce what happens in our model. For example, the advent of the assembly line in the early twentieth century made it easier to substitute unskilled labor and capital for skilled labor. Such technogical changes made plants bigger and lowered the return to skill, reinforcing what happens in the model economy of the previous section in the early periods. Other technological changes work against what happens in our model economy. There are new innovations making it easier to substitute capital for very high skilled labor. Extreme examples are computers that play chess and that make diagnoses of appropriate heart attack treatments. Such innovations would reduce the return to skill, working against what happens in the model in later periods.

In recent years, advances in technology have made capital more flexible. Manufacturers use computer-controled machines that can be quickly reprogramming to do different tasks. In our theory, skill is at the top of the hierarchy in flexibility, unskilled is below that and capital is at the bottom. This hierarchy probably has more relevance looking backward over the past century than going forward into the new century. We think that the modelling approach we take here can potentially be extended to account for these developments. In any event, we anticipate that for the foreseable future, skilled labor will continue to tend to be allocated complex tasks with small batch sizes, to keep setup costs low.

References

[1] Abowd, J. and F. Karmarz (1999). "The Analysis of Labor Markets Using Matched Employer-Employee Data," in *Handbook of Labor Economics*, vol. 3.

[2] Acemoglu, Daron (2003), "Patterns of Skill Premia," *Review of Economic Studies 70*, 199-230.

[3] Aghion, Philippe, Peter Howitt, and Giovanni L. Violante (2002), "General Purpose Technology and Wage Inequality," *Journal of Economic Growth* 7, 315-345.

[4] Atack, J. and F. Bateman (1999). "Nineteenth Century American Industrial Development Through the Eyes of the Census of Manufactures: A New Resource for Historical Research," *Historical Methods*, 32, 177-188..

[5] Atack, J., F. Bateman, and R. Margo (2004). "Skill Intensity and Rising Wage Dispersion in Nineteenth-Century American Manufacturing," *Journal of Economic History* v64, pp 172-192.

[6] Berman, E., Bound, J., and Griliches, Z. (1994). "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures" *Quarterly Journal of Economics* v109, 367-97

[7] Bound, J. and G. Johnson (1992). "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations," *American Economic Review* 82, pp. 371-92.

[8] Braverman, Harry (1974). *Labor and monopoly capital; the degradation of work in the twentieth century.* New York: Monthly Review Press

[9] Published New York, Monthly Review Press [c1974]

[10] Brown, C. and J. Medoff (1989). "The Employer Size Wage Effect," *Journal of Political Economy* 97, pp. 1027-1059.

[11] Brown, Martin and Peter Philips (1986), "Craft Labor and Mechanization in Nineteenth-Century American Canning," *Journal of Economic History* Vol. XLVI, No. 3, 743-756.

[12] Caselli, F. (1999). "Technological Revolutions." *American Economic Review* 89, p.78.

[13] Chandler, Alfred (1990). *Scope and Scale: The Dynamics of Industrial Capitalism.* Cambridge, Mass: The Harvard University Press.

[14] Davis, S. and J. Haltiwanger (1991). "Wage Dispersion Between and within U.S. Manufacturing Plants, 1963-86," *Brookings Papers: Microeconomics*, pp. 115-200.

[15] Eisfeldt, Andrea L., and Adriano Rampini (2004) "New or Used? Investment with Credit Constraints." Northwestern University Working Paper.

[16] Galor, Oded and Omer Moav (2000), "Ability-Biased Technological Transition, Wage Inequality, and Economic Growth," *Quarterly Journal of Economics*, vol. 115, no. 2, pp. 469-97.

[17] Goldin, C. and L. Katz (1998). "The Origins of Technology-Skill Complementarity," *Quarterly Journal of Economics* 113, pp. 693-732.

[18] Goldin, C. and L. Katz (2000). "The Return to Skill in the 20th Century," working paper, Harvard University.

[19] Goldin, Claudia and Kenneth Sokoloff (1982), "Women, Children, and Industrialization in the Early Republic: Evidence from the Manufacturing Censuses," *Journal of Economic History* v42, 741-774.

[20] Holmes, Thomas and Matthew Mitchell (2003). "A Theory of Factor Allocation and Plant Size," NBER working paper 10079

[21] Idson, Todd (2001). "Skill-Biased Technical Change and Trends in Employer Size Effects," Columbia University manuscript.

[22] Katz, Lawrence F. and Kevin M. Murphy (1992)."Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics,* Vol. 107, No.1 pp 35-78.

[23] Mobius, M. (2000). "The Evolution of Work," working paper, MIT.

[24] Mitchell, M. (2001). "Specialization and the Skill Premium in the 20th Century," Federal Reserve Bank of Minneapolis Staff Report 290.

[25] Nelson, Richard R. and Edmund S. Phelps (1966), "Investment in Humans, Technological Diffusion, and Economic Growth," *American Economic Review 56,* 69-75

[26] Oi, W. and T. Idson (1999). "Firm Size and Wages," in *Handbook of Labor Economics*, volume 3B, 2165-2214.

[27] Troske, K. (1999). "Evidence on the Employer Size-Wage Premium from Worker-Establishment Matched Data," *The Review of Economics and Statistics*, Vol. 81, 15-26.

[28] U.S. Department of Commerce, Bureau of the Census (1895). *Report on Manufacturing Industries in the United States at the Eleventh Census: 1890, Part 1. Totals for States and Industries.* Washington, D.C. GPO.

[29] U.S. Department of Commerce, Bureau of the Census (1926). *Biennial Census of Manufactures 1923.* Washington, D.C. GPO.

[30] U.S. Department of Commerce, Bureau of the Census (1950). *Census of Manufactures: 1947, Vol. 1 General Summary.* Washington, D.C. GPO.

[31] U.S. Department of Commerce, Bureau of the Census (1955). *County Business Patterns First Quarter 1953. Part 1, United States Summary.* Washington, D.C. GPO.

[32] U.S. Department of Commerce, Bureau of the Census (1957). *Census of Manufactures: 1954, Vol. 1 Summary Statistics.* Washington, D.C. GPO.

[33] U.S. Department of Commerce, Bureau of the Census (1971). *1967 Census of Manufactures, Vol.1 Summary and Subject Statistics,* Washington, D.C. GPO.

[34] U.S. Department of Commerce, Bureau of the Census (1977). *County Business Patterns 1977, United States CBP/97-1.* Washington, D.C. GPO.

[35] U.S. Department of Commerce, Bureau of the Census (1981). *1977 Census of Manufactures, Vol.1 Summary and Subject Statistics,* Washington, D.C. GPO.

[36] U.S. Department of Commerce, Bureau of the Census (1993). *1987 Economic Census. CD-Rom. Volume 1 Report Series Release 1E*

[37] U.S. Department of Commerce, Bureau of the Census (1999). *County Business Patterns 1997, United States CBP/97-1.* Washington, D.C. GPO.

[38] U.S. Department of Commerce, Bureau of the Census (2001). *1997 Economic Census. CD-Rom. Disc 1F-1*

8 Appendix: Proofs

Calculations for Section 3

Recall that the unit cost function for the general ω case can be written as

$$\tilde{c} = \left(z_1 w_1^{1-\omega} + (z_2 - z_1) w_2^{1-\omega} + (1 - z_2) w_3^{1-\omega} \right)^{\frac{1}{1-\omega}} \tag{27}$$

Now

$$\frac{\partial \tilde{c}}{\partial z_j} = \frac{1}{1 - \omega} \tilde{c}^{\omega} \left(w_2^{1-\omega} - w_3^{1-\omega} \right). \tag{28}$$

From the marginal rate of substitution condition for cost minimization we have

$$w_j = w_3 \left(\frac{x_3}{x_j} \right)^{\frac{1}{\omega}}. \tag{29}$$

Substituting (29) for w_1 and w_2 into (27) and taking it to the ω power yields

$$
\begin{aligned}
\tilde{c}^{\omega} &= \left(z_1 \left(w_3 \left(\frac{x_3}{x_1} \right)^{\frac{1}{\omega}} \right)^{1-\omega} + (z_2 - z_1) \left(w_3 \left(\frac{x_3}{x_2} \right)^{\frac{1}{\omega}} \right)^{1-\omega} + (1 - z_2) w_3^{1-\omega} \right)^{\frac{\omega}{1-\omega}} \\
&= w_3^{\omega} x_3 \left[z_1 x_1^{\frac{\omega-1}{\omega}} + (z_2 - z_1) x_2^{\frac{\omega-1}{\omega}} + (1 - z_2) x_3^{\frac{\omega-1}{\omega}} \right]^{\frac{\omega}{1-\omega}} = w_3^{\omega} x_3 q^{-1}
\end{aligned}
$$

Substituting this into the first-order necessary condition then yields the expression reported in the text

$$
\begin{aligned}
0 &= q \frac{\partial \tilde{c}(z_1, z_2)}{\partial z_j} + \phi_j(z_j) \\
&= \frac{1}{1 - \omega} w_3^{\omega} x_3 \left(w_j^{1-\omega} - w_{j+1}^{1-\omega} \right) + \phi_j(z_j) \tag{30}
\end{aligned}
$$

Proof of Proposition 1. Suppose there is a finite set of product types indexed by $i \in \{1, 2, ..., I\}$ and let θ^i be the weight of product i and m^i the mass of products of this type. Let z_j^i denote the j-cutoff of product j. Let $z = (z_1^1, z_2^1, z_1^2, z_2^2, ...z_1^I, z_2^I)$ be the vector of all cutoffs. Define the space of feasible cutoffs $Z \subset R^{2I}$ to be any z such that such and $0 \leq z_j^i \leq z_2^j \leq 1$ for $j \in \{1, 2\}, i \in \{1, 2, ..., I\}$. Clearly Z is compact.

Start with an arbitrary $\hat{z} \in Z$. Define $\hat{n}_j^i$ to be the associated measures of tasks performed by factor j at firm i, $\hat{n}_1^i = \hat{z}_1^i$, $\hat{n}_2^i = \hat{z}_2^i - \hat{z}_1^i$, $\hat{n}_3^i = 1 - \hat{z}_2^i$. For $\hat{z}$ such that $\hat{n}_j^i > 0$ for all i and j, calculate the unique vectors $\left(\lambda^1, \lambda^2, ... \lambda^I\right)$, $(q^1, q^2, ... q^I)$, and (x_1^1, x_2^1, x_3^1), that solve

$$x_j^i = \lambda^i x_j^1 \tag{31}$$

$$\sum_{i=1}^{I} m^i \hat{n}_j^i x_j^i = \bar{X}_j, \ j \in \{1, 2, 3\} \tag{32}$$

$$\left(\hat{n}_2^i \left(x_1^i\right)^{\frac{\omega-1}{\omega}} + \hat{n}_2^i \left(x_2^i\right)^{\frac{\omega-1}{\omega}} + \hat{n}_3^i \left(x_3^i\right)^{\frac{\omega-1}{\omega}}\right)^{\frac{\omega}{\omega-1}} = q^i$$

$$\left(\theta^i\right)^{\sigma} q^1 = q^i,$$

where $\lambda^1 = 1$. Observe that in any equilibrium, given the CES production function the ratio of intensities across a firm of type i and a firm of type 1 is a constant ratio across factors. The first condition above imposes this constant ratio. The construction determines the input vectors that satisfy this constant ratio, as well as conditions (1), (2), (4) and (5) in the definition of equilibrium.

Denote the intensity levels $x_j^i(\hat{z})$, a continuous function by the implicit function theorem. Extend this to all of Z by taking the limit of the continuous $x_j^j(\hat{z})$. Compute $w_1(\hat{z})$ and $w_2(\hat{z})$ according to $w_j(\hat{z}) = w_3 \left(\frac{x_3^1(\hat{z})}{x_j^1(\hat{z})}\right)^{\frac{1}{\omega}}$ for $w_3 = p$ where p is defined by (14).

Next, solve the cost minimization problem for type i,

$$z^i(w_1, w_2) = \arg \min_{0 \le z_1 \le z_2 \le 1} q^i \tilde{c}(z_1, z_2; w_1, w_2) + f(z_1, z_2)$$

Where $\tilde{c}$ is the unit cost function defined in (5). Since the problem is strictly convex and continuous, the solutions $z^i(w_1, w_2)$ are continuous functions.

Define

$$z^*(\hat{z}) \equiv z^1(w_1(\hat{z}), w_2(\hat{z})) \times z^2(w_1(\hat{z}), w_2(\hat{z})) \times .... \times z^I(w_1(\hat{z}), w_2(\hat{z})).$$

Since this is a continuous function on a compact set, there is a fixed point $\hat{z} \in z^*(\hat{z})$. Construct $(q^i, p^i, w_j, z_j^i, x_j^i)$ according to (31), $p^i = p$ from (14), and $x_j^i = \lambda^i x_j^1$. By construction, this satisfies conditions (1) through (5) of equilibrium whenever $\hat{n}_j^i > 0$ for all i and j. Notice that, if $\hat{n}_j^i = 0$, any x_j^i will maintain market clearing (since it is always multiplied by zero), so the fixed point is an equilibrium in those cases as well. The household's budget constraint condition (6) holds by Walras law. $\blacksquare$

Table 1
Payroll Per Employee by Establishment Size and Year
(Normalized relative to average across all establishment sizes)

Employment Size Category	1947	1954	1967	1977	1987	1997
1-99	.91	.86	.88	.85	.82	.85
100-249	.96	.94	.89	.87	.89	.92
250-499	.98	.96	.91	.89	.92	.95
500-999	1.01	1.02	.98	.99	1.02	1.02
1,000-2499	1.05	1.09	1.09	1.14	1.20	1.20
2,500+	1.13	1.19	1.26	1.39	1.48	1.54

Source: Authors calculations from tabulations by establishment size in U.S. Census (1950, 1957, 1971, 1981, 1993, 2001).

Table 2
Plant Characteristics by Plant Size
1880 Census of Manufactures
Atack-Bateman Sample

Employment Size Category	Pay per Employee (normalized)	Women/Children share of workforce	Capital Intensity (normalized)
1-19	.98	.09	1.07
20-49	1.10	.19	1.30
50-99	1.10	.25	1.27
100+	.95	.40	.74

Source: Authors' calculation with the Atack-Bateman sample data discussed in Atack and Bateman (1999)

Table 3

Nonproduction Worker Share by Establishment Size and Year

(Normalized relative to average across all establishment sizes)

Employment Size Category	1947	1954	1967	1977	1987	1997
1-99	1.03	.95	.86	.92	.92	.96
100-249	.94	.94	.93	.92	.90	.95
250-499	.92	.94	.90	.92	.89	.93
500-999	.95	.99	.96	1.01	.93	.94
1,000-2499	1.05	1.04	1.09	1.10	1.12	1.03
2,500+	1.07	1.14	1.27	1.21	1.42	1.44

Source: Authors calculations from tabulations by establishment size in U.S. Census (1950, 1957, 1971, 1981, 1993, 2001).

Figure 1
Skilled Worker Share and Average Size in the Carriage and Wagon Industry
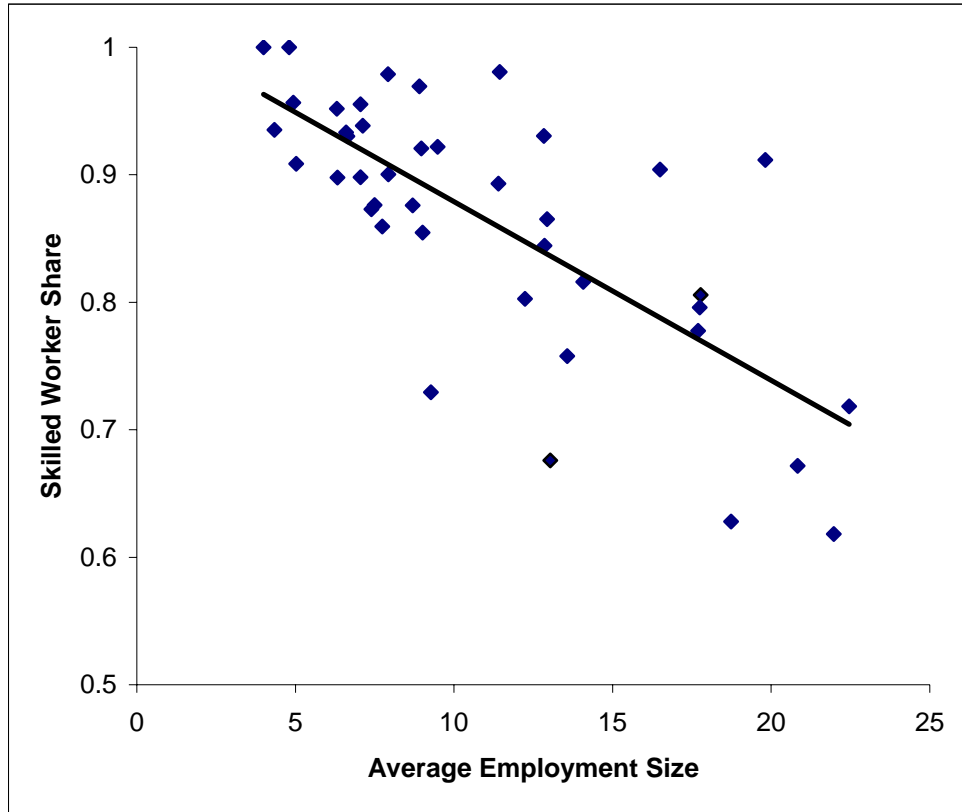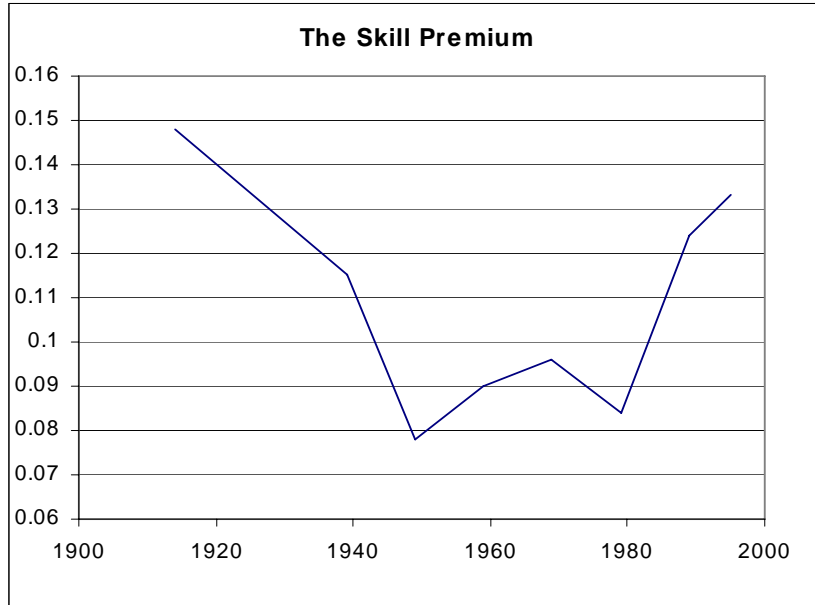by State
1890 Census

Figure 2
The Skill Premium (Return to College) in the 20th Century

The Skill Premium

Source: Goldin and Katz (2000).
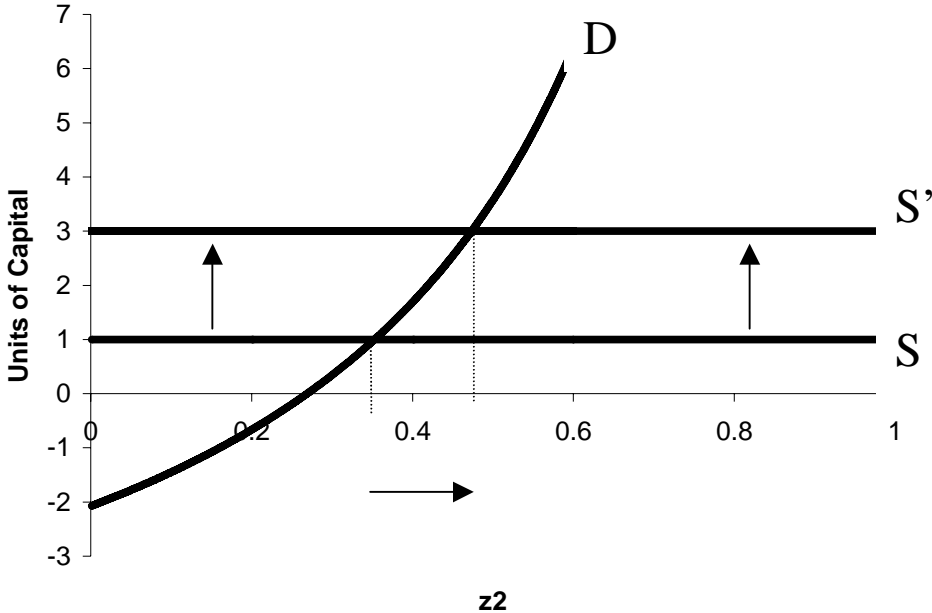
Figure 3
Supply and Demand for Capital
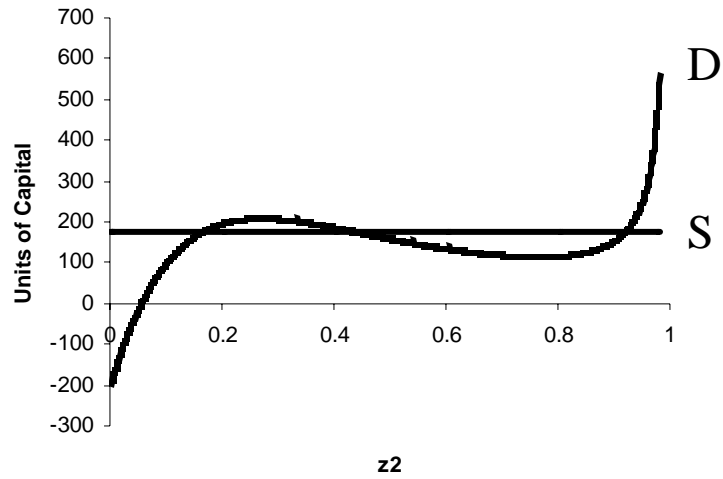
Figure 4
Example of Nonmonotone Demand

Figure 5
Factor Allocation and Prices Over Time
A Numerical Example