NBER WORKING PAPER SERIES

A THEORY OF DUAL LABOR MARKETS
WITH APPLICATION TO INDUSTRIAL
POLICY, DISCRIMINATION
AND KEYNESIAN UNEMPLOYMENT

Jeremy I. Bulow

Lawrence H. Summers

Working Paper No. 1666

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 1985

A Theory of Dual Labor Markets
with Application to Industrial
Policy, Discrimination
and Keynesian Unemployment

ABSTRACT

This paper develops a model of dual labor markets based on employers' need to motivate workers. In order to elicit effort from their workers, employers may find it optimal to pay more than the going wage. This changes fundamentally the character of labor markets. The model is applied to a wide range of labor market phenomena. It provides a coherent framework for understanding the claims of industrial policy advocates. It also can provide the basis for a theory of occupational segregation and discrimination which will not be eroded by market forces. Finally, the model provides the basis for a theory of involuntary unemployment.

Jeremy I. Bulow
Center for the Study of
the Economy and the State
University of Chicago
1101 East 58th Street
Chicago, IL 60637
(312) 962-8779

Lawrence H. Summers
Department of Economics
Harvard University
Cambridge, MA 02138
(617) 495-2447

Economists have a clear understanding of how perfectly competitive labor markets without any information problems would function. All workers would be paid their marginal products. Wages of equally productive workers would be equal and wage differentials would depend only on productivity differentials. Involuntary unemployment would not be observed because of flexible wages. Marginal workers would be indifferent to losing their jobs since wages would just equal the amount they could earn pursuing alternative opportunities. With perfectly competitive labor markets, subsidies to workers in specific industries or interferences with free trade would be very likely to reduce economic welfare.

Unfortunately, none of these characteristics of competitive labor markets appears to hold in practice. Wages differ across workers in ways that are inexplicable in terms of productivity differences. After controlling for both measurable and unmeasurable differences in productivity, econometric studies consistently suggest that factors such as occupation, firm size, race, sex, age and tenure all influence wages. Involuntary unemployment is observed and is found to rise at times when productivity and real wages fall. Workers value their jobs and regret losing them. Subsidies to workers in particular industries, and interferences with free trade are pervasive around the world.

This paper examines one theory of dual labor markets which can account for these phenomena.[1] The deviation from the standard competitive model which drives our analysis is the inability of firms to costlessly determine how much effort their workers are putting forth. Firms can elicit more effort from their

---

[1]Shapiro and Stiglitz (1984) analyze some related issues with a model which is formally very similar to the model presented in the first section of this paper. Yellen (1984) provides an informal discussion of dual labor markets paralleling parts of the formal discussion presented here. Our initial analysis was developed before we became aware of these contributions. The idea that

workers either by devoting resources to monitoring them more closely or by raising the cost to them of being caught shirking. The latter objective may require that firms pay workers more than their opportunity cost. A worker who is paid only his opportunity cost has little incentive to avoid shirking. Firms have a strong incentive to make workers value their jobs. They can only do this by paying more than the "going wage." It is the deviation of workers' wages from their opportunity costs that gives rise to our model's imperfectly competitive features.

The potential importance of the linkages between the level of wages and worker productivity which form the basis of our analysis is well illustrated by one of the most dramatic chapters in the history of American industrial relations. This episode also serves to highlight the importance of the effort elicitation problem as an explanation for non-market clearing wages. In 1914, the Ford Motor company introduced the five dollar day for industrial workers. At the time prevailing wages at Ford and other companies were between $2.00 and $3.00 a day. Immediately following the announcement, more than ten thousand persons gathered outside the Ford plant gates looking for jobs. Ford declared that the motive for the wage increase was "profit sharing and efficiency engineering".[2] A contemporary engineering study of production at Ford, Arnold and Faurote (1915) described the results of Ford's innovation:

---

firms need to elicit effort from workers can give rise to a labor market with imperfectly competitive features has recently been explored in Calvo (1979), Ishikawa (1984), and Bowles (1984). A particularly clear statement of the role of non-competitive wages in eliciting effort may be found in Becker and Stigler (1974). It dates back at least to Marx's "reserve army of the unemployed."

[2]Quoted in Nevins (1954), p. 538.

> The Ford high wage does away with all of the inertia and living
> force resistance... The workingmen are absolutely docile, and it
> is safe to say that since the last day of 1913, every single day
> has seen major reductions in Ford shops labor costs.[3]

Alan Nevins (1954) in his history of the early Ford Motor Company concluded

that:

> Ford and his associates freely declared on many occasions that
> the high wage policy had turned out to be good business. By this
> they meant that it had improved the discipline of the workers,
> given them a more loyal interest in the institution, and raised
> their personal efficiency... Once the Ford factory, like others
> had been called "The House of Correction"; now it was temporarily
> called "The House of Good Feeling".[4]

The only available quantitative estimate of the effects of Ford's high wage

policy appears to be the calculation of Ford's chief of labor relations, John

Lee (1916) that productivity increased by 51 percent in 1914 following the

introduction of the high wage policy. One estimate, Levin (1927) suggests that

the high wage policy halved absenteeism. And Nevins reports that discharges

for cause declined very sharply because of the "instant and unquestioning

obedience of men eager not to lose their five dollar day."

The increases in productivity appear to be attributable almost entirely to

increases in worker effort. Nevins (1954) after reviewing the Ford archives

concluded that:

> Suggestions that management was actuated by a desire to get the pick of
> Detroit mechanics and by anxiety to end a high turnover rate are
> demonstrably false and misleading...The mass production methods of the Ford
> plant made it unnecessary to search for picked mechanics...the turnover
> problem had been practically solved before the five dollar wage was
> adopted. The theory [that management was trying to avert the threat
> of unionization] is not supported by real evidence. The union had
> shot its feeble bolt long before the wage decision.[5]

---

[3]Arnold and Faurote (1915), p. 331.

[4]Nevins (1954), p. 549.

[5]Nevins (1954), pp. 551-554.

The contemporary importance of variations in effort and the costs of monitoring them may be conveyed in a number of ways. A large fraction of the US labor force is involved in supervision rather than the direct production of goods and services, and the maintaenance of worker morale is a major priority for most firms. Productivity varies widely across time and space in ways which are difficult to account for except in terms of differences in effort broadly defined.[6] An HEW task force on Work in America (1973) after reviewing hundreds of studies of individual plants concluded that "lowering business costs by reducing absenteeism,tardiness,turnover, labor disputes, sabotage, and poor quality can increase productivity by up to 40 percent."

Many economists take the position that markets will tend to make efficient micro-economic decisions about the compostition of national output, and the means by which it is produced but that stabilization policies of some sort are required to address the macro-economic problem of cyclical unemployment. An alternative view is that recessions are just the most obvious symptom of market imperfections which pervasively distort the allocation of resources. The model presented here provides a justification for this alternative view. It shows how involuntary unemployment can result from sources which also badly distort microeconomic aspects of performance.

The paper is organized as follows. Section I lays out our basic model of a dual labor market. It shows how equally productive workers can in equilibrium be allocated arbitrarily between a primary and secondary sector of the economy.

---

[6]Nelson (1981) surveys evidence suggesting that standard variables leave a great deal of productivity variation unexplained.

These sectors parallel institutional descriptions of the dual labor market such as that of Doeringer and Piore (1971). The primary sector is characterized by high wages and responsible career jobs, while the secondary sector has menial jobs, low wages and no job ladders. Although workers in the secondary sector envy those in the primary sector, and are equally productive there is no equilibrating market force which can erode wage differentials. The model also explains why firms are extremely conscious of their relative standing in setting wages, and how firms paying different wages can coexist in market equilibrium.

Section II shows that our dual labor market model provides a formal framework which can justify the arguments of American industrial policy advocates that the high wage/high value added sectors of an economy should be subsidized and protected from foreign competition. Unlike most of the arguments in the industrial policy debate, ours explains why workers and not just imperfectly competitive firms may benefit from industrial policies.

Section III discusses theories of wage discrimination -- the observation that workers' wages are affected by their demographic status as well as by their productivity. Unlike standard theories of discrimination, the model presented here explain why discrimination endures in spite of market forces tending to eliminate it. Our theory of discrimination also provides an explanation for the empirical observation that disadvantaged groups receive "equal pay for equal work" but unequal work. We also show that "affirmative action" policies which subsidize desirable jobs for members of disadvantaged groups are likely to raise total welfare.

Section IV extends the model to provide a theory of involuntary unemployment. We argue that it is a natural consequence of the rationing of

primary sector jobs for incentive reasons. By postulatinf that the primary
sector does not hire workers who are employed but only those who are out of
work,we are able to show that involuntary unemployment with Keynesian features
will result. The model also provides the basis for theories of the cyclical
upgrading of secondary workers, the absence of work-sharing arrangements, and
real business cycles.

Section V concludes the paper by suggesting directions for further
theoretical and empirical research.

I.   A Theory of Dual Labor Markets

Doeringer and Piore (1971) have described the American economy as having a
dual labor market. Jobs fall into either the primary or secondary sector. Jobs
in the primary sector are "good jobs" characterized by high wages, job security,
substantial responsibility and ladders where internal promotion is possible.
Jobs in the secondary sector are characterized by low wages, casual attachments
between workers and firms, and are menial. Workers in the secondary sector envy
those in the primary sector who have both better jobs and higher wages. A
typical example of a primary sector employer is a large manufacturing
establishment, while small service firms such as fast food outlets typify the
secondary sector.

As Doeringer and Piore and others have documented, these descriptions
accurately capture many aspects of the labor market. Yet they raise an obvious
question. If workers in the secondary sector envy those in the primary sector,
why are not wages in the primary sector bid down? The logic of competitive
economics denies the possibility of equally skilled workers receiving different

wages in different jobs. The model in this section provides an explanation for the existence of dual labor markets. The central idea is that primary sector firms may find it advantageous to pay more than the "going wage" it helps them in eliciting effort from their workers. By interpreting the secondary sector in the model presented below as home production, it can also be viewed as a theory of involuntary unemployment.[7]

The Model

We assume that the economy is comprised of N identical infinitely lived agents. Each can supply one unit of labor and produce $\bar{w}$ units of output in either sector of the economy. Consumers maximize lifetime welfare which is given by:

$$(1.1) \qquad U = \int_t^\infty U(x_1, x_2 - \alpha s) e^{-r(v-t)} dv$$

where x_i is the number of units of sector i output consumed in period t, r is the discount rate, and s is an indicator variable equalling zero when the worker works, and one when he shirks. Thus it is assumed that only two levels of worker effort are possible. Shirking workers are assumed to produce no output. Note that we have assumed that shirking and consuming extra units of x_2 are perfect substitutes. This assumption simplifies the analytic treatment considerably without altering any of the substantive conclusions. Note that in equilibrium no workers shirk.

We further assume that preferences are homeothetic and normalize so that $U(0,0)=0$. Risk neutrality is also assumed. It follows that

---

[7]Such a model is provided by Shapiro and Stiglitz (1984). We suggest an alternative and we believe more satisfactory theory of involuntary unemployment in Section IV.

$U(\lambda x_1, \lambda x_2) = \lambda U(x_1, x_2)$. Note also that our assumptions imply that all consumers will consume the economy's two goods in the same proportion. We assume that consumers can neither borrow nor lend. As we discuss below, this assumption restricts workers from posting bonds against the possibility of their shirking.

We normalize the price of secondary sector output, x_2 to be unity. Using the hometheticity of the utility function and production assumptions noted above we can write:

$$(1.2) \qquad p_1 = f(\frac{x_1}{x_2}) = f(\frac{E_1}{E_2}) = f(\frac{E_1}{N-E_1}) \equiv g(E_1); \ g' < 0$$

where E_i workers are employed producing $\bar{w}E_i$ units of output in sector i.

Competition will insure that secondary sector workers receive a wage equal to their marginal product, $w_2 = \bar{w}$. In keeping with the discussion above, we assume that secondary sector workers are monitored perfectly and thus have no possibility of shirking. This reflects the idea that secondary sector jobs are menial.

The key aspect of our model is that detection of shirkers in the primary sector is difficult. This is a natural consequence of the responsible character of primary sector jobs. Both false positives and false negatives may result as firms try to detect shirkers. We assume that a worker who is not shirking has an instantaneous likelihood d_1 of being falsely labelled a shirker. A worker who is shirking is assumed to have an instantaneous likelihood d_2 of being identified as a shirker. It follows that the probability of not being labelled a shirker over an increment of time is $(d_2-d_1)dt$ greater for those who are not shirking than it is for those who are shirking. It is also assumed that all employment relationships have an exogenous separation rate q. Separations occur

either due to worker quits in order to relocate of withdraw from the labor force or due to employer induced seperations caused by changing patterns-of product demand.

We assume in the discussion below that firms are restricted to paying all workers a common wage. Subsequently, more elaborate compensation schemes are discussed. Firms in the primary sector need to induce their workers not to shirk. The only penalty that they have at their disposal is the threat to fire workers caught shirking, since by assumption workers cannot post a bond. It will become apparent that it is never desirable to fire workers not caught shirking. Denoting the present value of lifetime utility for workers in the primary and secondary sectors respectively by PV_1 and PV_2, workers will shirk at any given instant unless the condition

$$(1.3) \qquad \alpha \leqslant (d_2-d_1)(PV_1-PV_2)$$

is satisfied. The left hand side is the instantaneous gain in utility from shirking while the right hand side is the product of incremental probability of being dismissed if a worker shirks and the loss in lifetime utility from being dismissed. The values of PV_1 and PV_2 may be calculated from the recursive equations:

$$(1.4) \qquad PV_1 - \frac{\bar{w}}{r} = (w_1-\bar{w}) \int_0^\infty e^{-(r+q+d_1)t} dt + (q+d_1)(PV_2-\frac{\bar{w}}{r})\int_0^\infty e^{-(r+q+e_1)t} dt$$

$$(1.5) \qquad PV_2 - \frac{\bar{w}}{r} = \frac{E_1}{E_2}(q+d_1)(PV_1-\frac{\bar{w}}{r}) \int_0^\infty e^{-[r + \frac{E_1}{E_2}(q+d_1)]t} dt$$

Equation (1.4) holds that the present value of the surplus from a primary sector job is the sum of the discounted surplus from the current job and the present value of the future surplus if in the secondary sector, discounted to reflect

the time until the secondary sector will next be entered. Since workers in the secondary sector get a wage $\bar{w}$, equation (1.5) is simply the present value of the future surplus if in the primary sector discounted to reflect the time until the primary sector is next entered. In forming equation (1.5) we have made use of the steady state assumption that the flow of workers into and out of the primary sector is equal, so that the flow rate out of the secondary sector is $\frac{E_1}{E_2}(q+d_1)$.

Solving equations (1.3) through (1.5) yields the no shirk condition on primary sector wages:[8]

$$(1.6) \qquad w_1 - \bar{w} = \frac{\alpha r}{d_2 - d_1} + \frac{\alpha(d_1 + q)N}{(d_2 - d_1)(N - E_1)}$$

Equation (1.6) has several plausible implications. As the utility from shirking, α, increases, firms must pay more to induce their workers not to shirk. As the probability of successfully detecting a shirker, d_2 declines, firms must also pay higher wages. Likewise, if the rate of turnover among non-shirkers, $d_1 + q$ increases firms must pay higher wages. This is because the value of maintaining a job is reduced if future turnover is more likely. The greater the number of primary sector jobs, E_1, the higher wages must be to maintain the opportunity cost of losing a job, because the time a worker must spend waiting to return to the primary sector if fired is reduced. Finally, increases in r raise required wages because they reduce the present value of the loss from being fired.

Notice that equation (1.6) implies that in equilibrium, primary sector wages will exceed wages in the secondary sector even though all workers are

---

[8]Another way to derive (1.6) is to note that the difference in present value of a primary and secondary sector job, $PV_1 - PV_2$, equals

$[w_1 - \bar{w}]/(r + d_1 + q + [E_1(d_1 + q)/N - E_1]$

identical. Workers in the secondary sector will envy those in the primary sector, but it will not be possible for them to bid for primary sector jobs by being willing to accept lower wages. For if they accepted lower wages, they would have an incentive to shirk. Hence firms will not offer lower wages.

The model here implies that firms will be extremely conscious of relative wages. Primary sector firms need to observe wages offered in the secondary sector in order to insure that their wages are set high enough so that their workers have no incentive to shirk. In standard competitive models, firms need only observe their own labor supply curves and have no need to learn about either wage or employment levels at other firms. Indeed, one of the major arguments for competitive markets as an allocation mechanism is that they reduce the costs of information acquisition because each agent can act knowing only the prices it faces.

In fact, firms are extremely concerned with assessing where they stand in the wage distribution. The Handbook of Wage and Salary Administration (1984) notes that "Salary surveys are indispensable to an effective compensation program...The establishment of sound compensation policies depends on an accurate assessment of an organization's place in the salary marketplace." It goes on to give an idea of volume of survey information available by noting that there are almost 100 available surveys on the compensation of accounting clerks, almost 50 surveys with salary data on the Chicago metropolitan area and 9

This is the incremental wage from being in a primary sector job, $w_1 - \bar{w}$, discounted by the interest rate, the chance that a primary sector worker will drop into the secondary sector, d_1+q, and the chance that a secondary sector worker will find employment in the primary sector, $E_1(d_1+q)/(N-E_1)$.

surveys with specific data on salary for accounting clerks in Chicago.[9]

Equation (1.6) also provides an explanation additional to standard human capital arguments for the existence of enduring attachments between workers and firms, which does not rely on the productivity enhancing effects of experience. Reductions in the exogenous seperation rate q, reduce the wage which primary sector firms pay in order to insure that workers will not shirk. Firms therefore will have an incentive to minimize this separation rate. At a minimum this means that they will not replace workers not determined to be shirking. Firms also have an incentive to give laid off workers first priority for new jobs, and to provide alternative jobs within the firm for poorly matched workers.

Equation (1.6) provides one relation between primary sector employment and wages. A second relation is necessary to characterize the market equilibrium. This relation (5) results from the requirement that workers in the primary sector are paid their marginal product. It holds that

$$(1.7) \qquad w_1 = p_1 \bar{w} = \bar{w} f(\frac{E_1}{E_2}) = \bar{w} g(E_1)$$

The determination of equilibrium by the intersection of these product market equilibrium condition and the no shirk condition is depicted in Figure 1. It is apparent that in competitive equilibrium, identical workers will be paid different wages. We defer until the next section an analysis of the efficiency properties of this equilibrium. It is clear however that the composition of output will be affected by factors such as the probability of detecting shirking

---

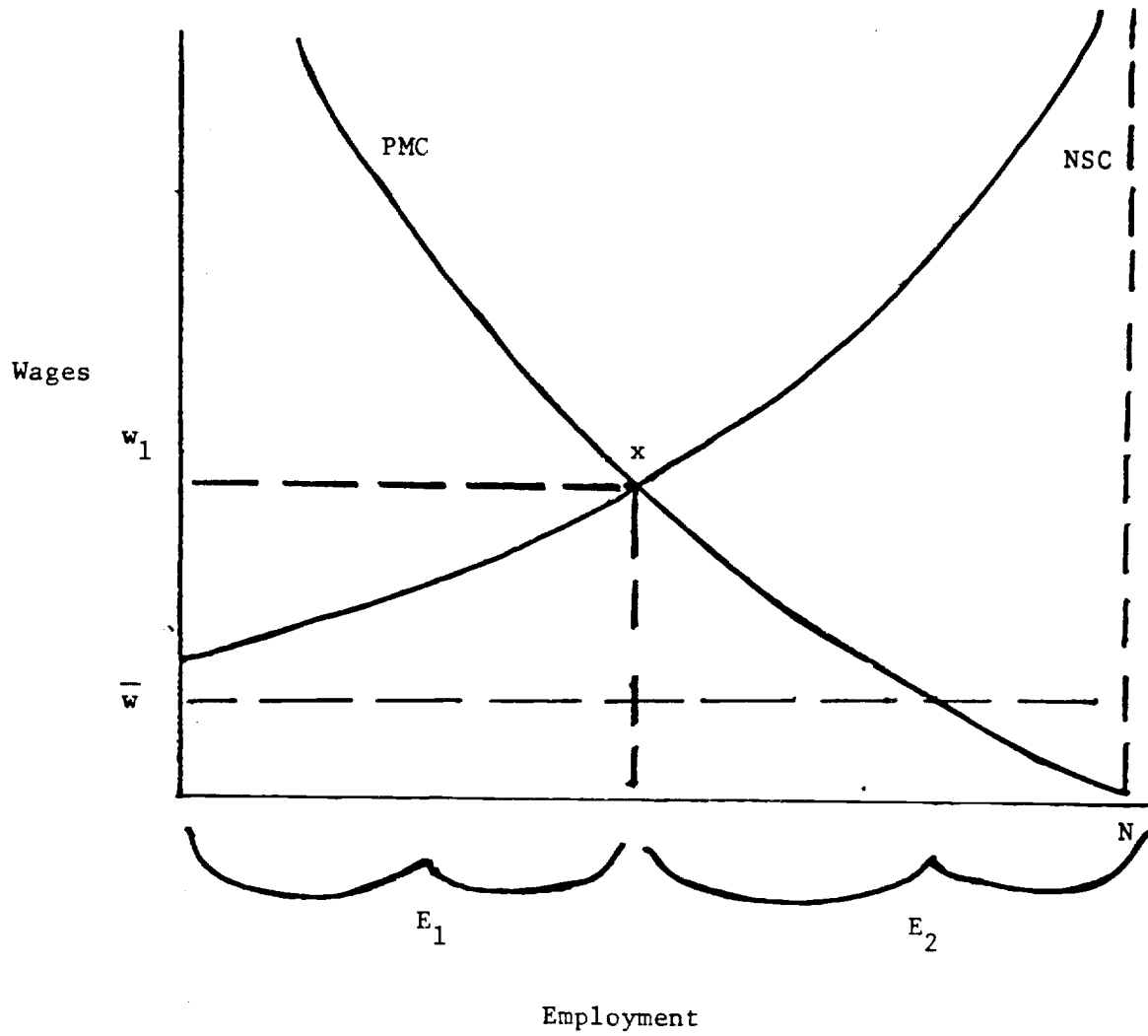[9]Handbook of Wage and Salary Adminsitration (1984), p. 323.

Figure 1

The Determination of Equilibrium

workers, the interest rate and the utility workers get from shirking. In equilibrium, of course no shirking will occur.

The analysis so far has assumed that the only sanction available to firms when workers are suspected of shirking is dismissal. One can imagine firms developing various other means for inducing workers not to shirk. For example, workers in primary sector jobs might be asked to post bonds which would be forfeited in the event they were detected shirking. Since the probability of detection is less than one, these bonds would have to exceed the value of the output foregone because of worker shirking. Bonds of this type are not observed in practice for two reasons. First, workers simply do not have the requisite liquidity to post bonds. Second, and probably more important, third party verification of detected shirking is impossible. Firms would have an incentive to falsely label workers as shirkers in order to collect their bond. Even if firms did not do this there would remain the problem of determining in which employment separations the bond should be forfeited.

While we do not observe the posting of bonds in the actual economy, a number of the features of actual primary sector firms may perform some of the same economic functions that bonding might perform. In particular, these firms are characterized by job ladders and rising age-wage profiles, while similar phenomena are not observed in the secondary sector. As emphasized by Lazear (1981) a rising age/wage profile can maintain the present value of a job as a worker's years to normal retirement diminish. As Medoff and Abraham (1981) have argued, it is difficult to ascribe rising age-wage and tenure profiles to rising productivity.[10] Rising age-wage profiles are of course subject to the same sort

---

[10]Their claim has been challenged by a number of authors, notably Braun (1984). It seems unlikely, however, that all of the observed returns to experience can be traced to this source.

of incentive problems as bonds with one important addition. Rising age-wage profiles unlike posted bonds will encourage workers to stay with firms even when they have higher productivity opportunities elsewhere. As long as firms cannot fully solve the effort elicitation problem with rising age-wage profiles, wage differentials will persist.

It is noteworthy that job ladders appear in parts of the economy where individual performance is difficult to disentangle but not in parts where it is easily monitored. Thus John Dunlop (1985) writes "Capital intensive production methods as in electric power generation, refineries and basic steel tend to be associated with...lengthy promotion ladders and elongated pyramids, while more labor intensive operations as in light assembly, grocery stores and sewing have few promotion ladders and flat pyramids." As Doeringer and Piore (1971) emphasize, promotion opportunities predominate in high rather than low wage sectors.

We have already shown how taking account of the effort elicitation problem can potentially explain two features of observed labor markets: the fact that there are good and bad jobs held by equivalent workers, and the existence of rising age-wage profiles in primary sector jobs. How important are these phenomena empirically? One way of thinking about the question is the following: How indifferent, even in good times, would the typical manufacturing worker be to the loss of his job? In a competitive market, firms would pay workers no more than their opportunity cost of working. Hence workers would be indifferent to losing their job. This is not true in the environment described here where good jobs are scarce and rationed.

Intra Industry Wage Differentials

The analysis presented so far provides an explanation for inter-industry and inter-occupation wage differences even for equally skilled workers. This analysis is clearly relevant to empirical observations such as the often emphasized finding that wages in US manufacturing exceed those in service jobs, even after controlling for many measures of worker quality. It is also relevant to the dualism that is said to characterize the labor market. Yet, a second type of anomaly in the wage structure also stands out. There appear to be substantial, persistent differences in wages within industries. Large literatures document the correlation of wage rates with unionism (e.g., Freeman and Medoff (1984)), and firm size (e.g., Brown and Medoff (1985))and there is substantial idiosyncratic variance as well. The analysis of these wage differentials poses a difficult issue. Leaving aside the question of why firms choose to pay different wages, how do high wage firms remain viable in the market place? The question is a critical one. For typical firms the wage bill is 4 times profits. If a firm paid even 10 percent above the going wage its profits would be reduced by 40 percent.

In part, we believe the answer lies in the monopolistically competitive character of product markets as described in Weitzman (1984). But a simple extension of our model can provide a further part of the explanation. Imagine that the probability of detecting a shirking worker in the primary sector is not given exogenously but depends on the expenditure of resources. These resources are consumed according to a function $\phi(d_2)$, where $\phi' > 0$. Then primary sector firms will solve the problem:

(1.8)   Min: $w + \phi(d_2)$ s.t. No Shirk Condition

to produce output most efficiently. It is immediately apparent from the first order conditions for the solution of (1.8) that first order changes in wages

will have no effect on labor costs.[11] This is dramatically different from the standard model where wage changes translate directly into cost changes. It follows that discrete changes in wages will give rise to much less than proportional changes in labor costs. This helps to explain how high wage firms endure. From the perspective of this model, the observation that firms in the same industry pay different wages is no more mysterious than the observation that their factor proportions differ. Wage premia in this model contribute directly to productivity, just as do standard factors of production.

Indeed the minimization problem, (1.8), provides an explanation for the union productivity effect discussed by Freeman and Medoff (1984). When wages are increased, firms are able to conserve on resources devoted to monitoring workers and so total productivity is enhanced. As predicted by this line of argument, Brown and Medoff (1978) report some evidence suggesting that the ratio of non-production to production workers is lower in the unionized than the nonunionized sector.

Equation (1.8) also provides a basis for a theory of the firm size effect. Large firms enjoy economies of scale and are able to get reduced prices from suppliers but because of their size have less favorable $\phi(d_2)$ functions. They therefore substitute high wages for supervision. These ideas comport well with a number of Brown and Medoff's (1985) empirical findings: (i) quit rates are lower in large and unionized establishments; (ii) large employers have more pronounced job ladders than do smaller firms and steeper age-wage profiles; (iii) the wage differential between large and small firms decreases with skill

---

[11]The argument here is similar to but different from that of Akerlof and Yellen (1984). They emphasize that inertia may not be very costly for firms. Our point is that a variety of compensation-monitoring strategies will be about equally attractive to firms. The argument here is more general than our model. It will be valid in any setting where firms choose wages. Models based on

level. It is greatest at the bottom of the hierarchy where differences in monitoring technologies are likely to be greatest.

Wage Dynamics

In addition to contributing to an explanation of static wage differentials, the model presented here can contribute to the explanation of aspects of the dynamics of wage differentials that are otherwise puzzling. In a number of high wage industries, in the US and Europe, notably steel, wage differentials increased substantially at the same time that the market conditions facing the industry deteriorated dramatically. In the US for example, steel wages rose by 35 percent relative to those in all manufacturing between 1970 and 1980. No standard competitive or bargaining theory, would predict a dramatic increase in relative wages in the face of declining product demand.[12] One would expect reductions in labor demand to lead to both lower employment and lower wages.

The actual outcome is suggested by the model presented here. An increase in the probability that the employer will initiate a separation will necessitate the granting of a wage increase in order to insure that the no-shirk condition (1.6) is satisfied. Mathematically, a layoff probability would serve the same function in (1.6) as a higher quit rate. Thus failing firms will find themselves forced to raise wages creating the prospect of dynamic instability. This may also partially explain why firms tend to give relatively little notice

---

turnover or labor quality considerations would yield parallel results.

[12]Lawrence and Lawrence (1985) discuss an "endgame" explanation for this phenomenon. Their explanation depends critically on union bargaining strategies while their empirical work suggests that it occurs about equally frequently in the union and non-union sector. This suggests the possible relevance of the explanation presented below.

before plant closings.[13]

Summary

The analysis in this section suggests that in any economy where firms cannot monitor workers perfectly, they will pursue policies which will cause workers to value their jobs. A necessary consequence of workers being led to value their jobs is wage differentials unrelated to skill differentials. The contribution of this factor to wage inequality is difficult to gauge. But some evidence suggests that arbitrary wage differentials not linked to differences in skill may be of substantial importance. Studies consistently find that differences in genes, parental upbringing, years of schooling and IQ all taken together can explain only a very small part of the inequality in wages.[14] As Jencks (1972) and Thurow (1976) argue this suggests the possible importance of luck in wage determination. The formulation presented here shows how luck can affect wages in a competitive market.

A labor market where there are pervasive wage inequalities unrelated to productivity differentials, is profoundly different from a competitive labor market. In the next section we explore the normative properties of our model.

II. Trade and Industrial Policies

The previous section provided a model that was consistent with the

---

[13]On many occasions struggling firms do extract wage reductions from their workers. But the packages frequently include other elements operating directly on the no-shirk condition. These may include job guarantees or profit sharing plans which give workers an incentive to monitor each other.

[14]Results presented in Taubman (1977) estimates suggest that the expected absolute difference in earnings between identical twins is about two-thirds as great as between randomly chosen members of the population. Jencks (1972) reports similar results for a host of other variables, including schooling and IQ.

existence of both large inter-industry and intra-industry wage differentials that are not readily explicable in terms of differences in labor quality. The standard competitive analysis of industrial subsidies and trade policies presumes the absence of these differentials. It is natural to ask how this analysis must be modified in the presence of endogenously generated wage differentials.

American industrial policy advocates such as Robert Reich and Lester Thurow take the position that nations should try to focus economic policies on encouraging high value added sectors. These sectors, it is argued provide "good high wage jobs," in contrast to the low wage jobs found in other sectors of the economy. Industrial policy advocates credit much of the economic success of Japan over the last 20 years to their successful encouragement of high value added production. They are alarmed by what they see as the de-industrialization of America and the consequent loss of high value added employment. While sophisticated industrial policy advocates recognize the economic truth that a nation must have a comparative advantage in something, they find little solace in this. In George Meany's picturesque phrase, "you cannot have a successful economy built on everyone doing everyone else's laundry."

From the perspective of standard neoclassical economic theory the claims of industrial policy advocates are difficult to understand. Competition equalizes the marginal productivities of all equivalent workers. There is no such thing as a good or bad industry. Domestic industrial subsidies distort the composition of economic activity without redeeming benefit. Foreign subsidies should be welcomed by domestic consumers. These conclusions rest on the assumption that competition equalizes wages across sectors. We shall see that

the claims of industrial policy advocates become explicable in the context of a
model like ours with endogenously generated wage differentials.

It is apparent from Figure 2 that a laissez-faire economy will not reach a
first-best outcome in a model like ours. The deadweight loss relative to the
first best is equal to the shaded area in the figure. A subsidy to primary
sector production, which had the effect of shifting the PMC locus to the right
would increase total economic welfare. Indeed, a subsidy large enough to shift
the PMC locus out sufficiently to cause primary sector employment to rise to X',
would achieve the same pareto-optimal outcome that would be achieved under
laissez faire. Note that an equivalent outcome could be attained by subsidizing
primary sector workers and thereby shifting the NSC outwards.

A more subtle question is whether a subsidy to primary sector employment
which was financed by a lump sum tax would represent a pareto improvement.
Workers in the secondary sector would be taxed to subsidize workers in the
primary sector who were already better off than they were. Nonetheless, it
turns out that such a subsidy would represent a pareto improvement. This may be
seen as follows. The NSC implies that:

$$(2.1) \qquad \frac{dPV_1}{ds} - \frac{dPV_2}{ds} = 0$$

where s is the rate of subsidy. A change in the subsidy rate must have an equal
impact on the present value of lifetime welfare for workers in both the
secondary and the primary sector. Equating the present value of the gain in
economic welfare from a change in the subsidy, to the sum of the present value
gains for workers currently in the primary and secondary sectors workers we have:

$$(2.2) \qquad \frac{w_1 - \bar{w}}{r} \frac{dE_1}{ds} = E_1 \frac{dPV_1}{ds} + E_2 \frac{dPV_2}{ds} + (PV_1 - PV_2)\frac{dE_1}{ds}$$
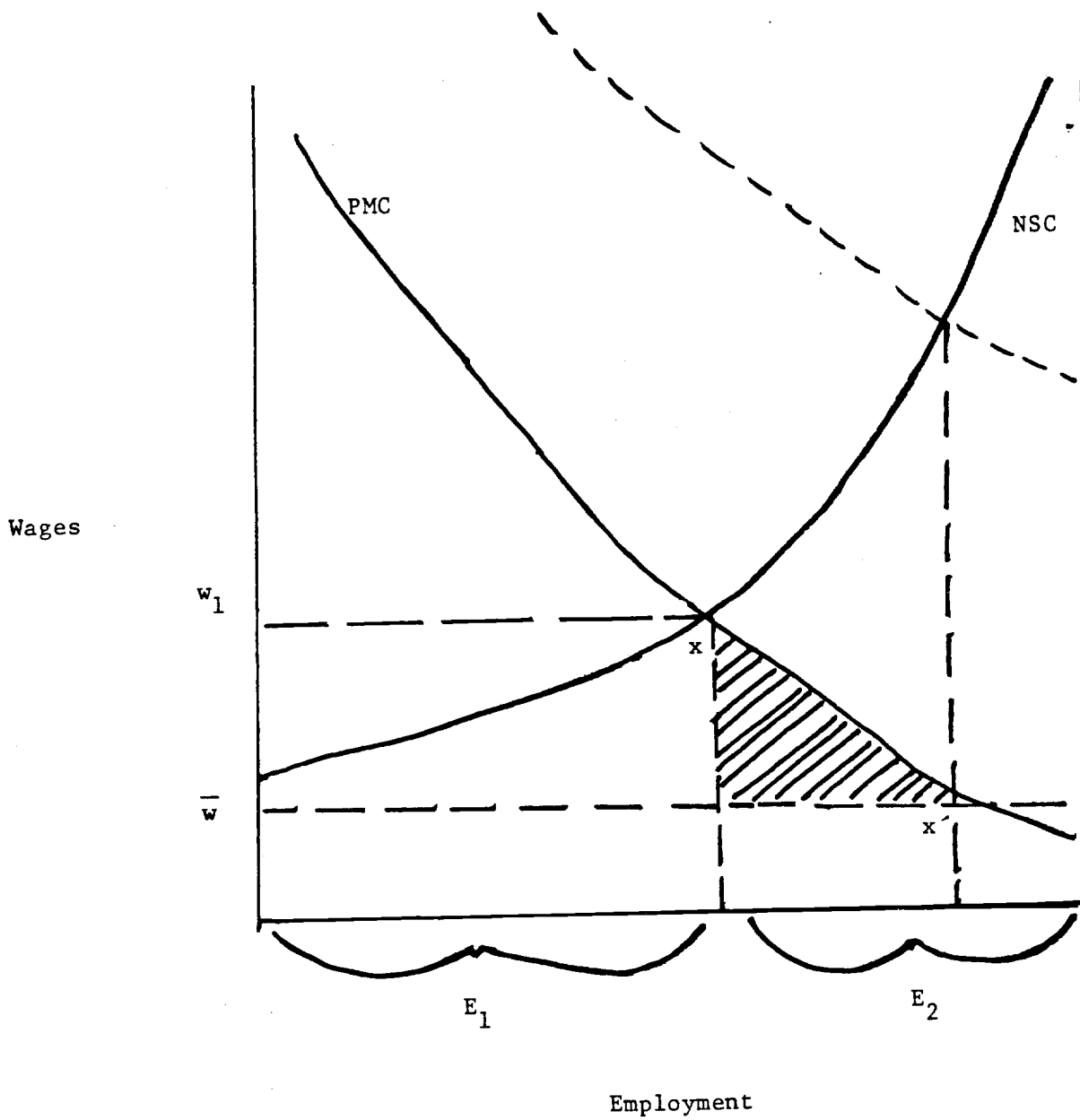
Figure 2

Subsidies and Economic Efficiency

where we have used the fact that $\dfrac{dE_1}{ds} = -\dfrac{dE_2}{ds}$

Since we know that $\dfrac{w_1 - \bar{w}}{r} > PV_1 - PV_2$ starting at a zero subsidy rate, it

follows that $dPV_1/ds = dPV_2/ds > 0$. A small subsidy will represent a pareto-improvement. It is not difficult to show, however, that a subsidy large enough to increase employment to its socially optimal level would reduce the present value of lifetime welfare for workers in the secondary sector. At the utilitarian optimum an increased subsidy to primary sector workers has a second order effect on social welfare but the transfer effect is first order; thus secondary sector workers must lose.[15]

In some respects these results parallel standard results, surveyed in Bhagwati and Srinavasan (1983) suggesting that if wages are artificially high in some part of an economy it is desirable to provide a wage subsidy to counteract the distortion. There are three important differences. First the wage differentials considered here are not arbitrarily fixed as in standard models but instead are determined endogenously and depend on the relative sizes of the two sectors. Increases in the subsidy raise wage differentials but are nonetheless desirable. Second, the distortion considered here is not a result of union power or other institutional forces. Unlike these cases, there is no cause for concern that subsidizing the high wage sector will lead to greater distortion. Third, our model permits us to assess separately the impact of a subsidy on the lifetime welfare of workers in the primary and secondary sector, because the flow between sectors is explicitly modelled.

The analysis so far has concentrated on the case of a closed economy. In

---

[15]These results have implications for the effects of other labor market policies. A proportional or progressive income tax reduces the utility gain from holding a primary rather than a secondary sector job, thus shifting the NSC to the left and reducing economic welfare. Note that this effect is first order

the case of an open economy, the differences between our formulation and the
'standard competitive one become much more dramatic. This is illustrated
heuristically in Figure 3.[16] Imagine that the domestic economy is small and
takes prices as given on world markets. Figure 3 illustrates a situation where
the world price of the primary sector good, $\bar{p}$, is less than the domestic price.
This may occur for any of the conventional "natural" reasons why foreigners
might have a comparative advantage in the primary sector, if foreign workers are
less prone to shirk, if foreign governments subsidize the primary sector, or if
foreign firms have found better contracting devices for inducing workers not to
shirk. Each of these possibilities has some relevance to the current U.S.
policy debate. The region ABDE represents the welfare loss to the economy due
to reductions in rents caused by the reduction in primary sector employment.
The angular area ABC represents the gain from free trade, as the price of
i ported goods is reduced. It is apparent that as long as the domestic primary
sector is "nearly competitive," allowing free trade will reduce national
economic welfare. The lost rents as primary sector employment contracts from OF
to OG represent a greater loss than the second order to gain to consumers from
allowing free trade.

There is a simple intuition for this result. The gains to consumers from

_____

unlike the second order welfare costs normally associated with labor income
taxes. The difference arises because of the pre-existing distortions present in
our model. Similarly, policies directed at compressing wage differentials will
in a model like this one have unfortunate consequences. If primary sector
workers cannot be bonded to work hard by high wages, the market equilibrium will
of necessity involve a large threat of being confined to the secondary sector.
This in turn requires an expansion in the secondary sector at the expense of the
primary sector.

[16]The graphical analysis is only heuristic because of its use of
consumers' surplus as a welfare measure and its neglect of the income effects of
policy changes on the demand for primary sector output.

Wages

PMC

NSC

A

$\overline{P}_1$ $\overline{w}$ B   C

$\overline{w}$

D   E

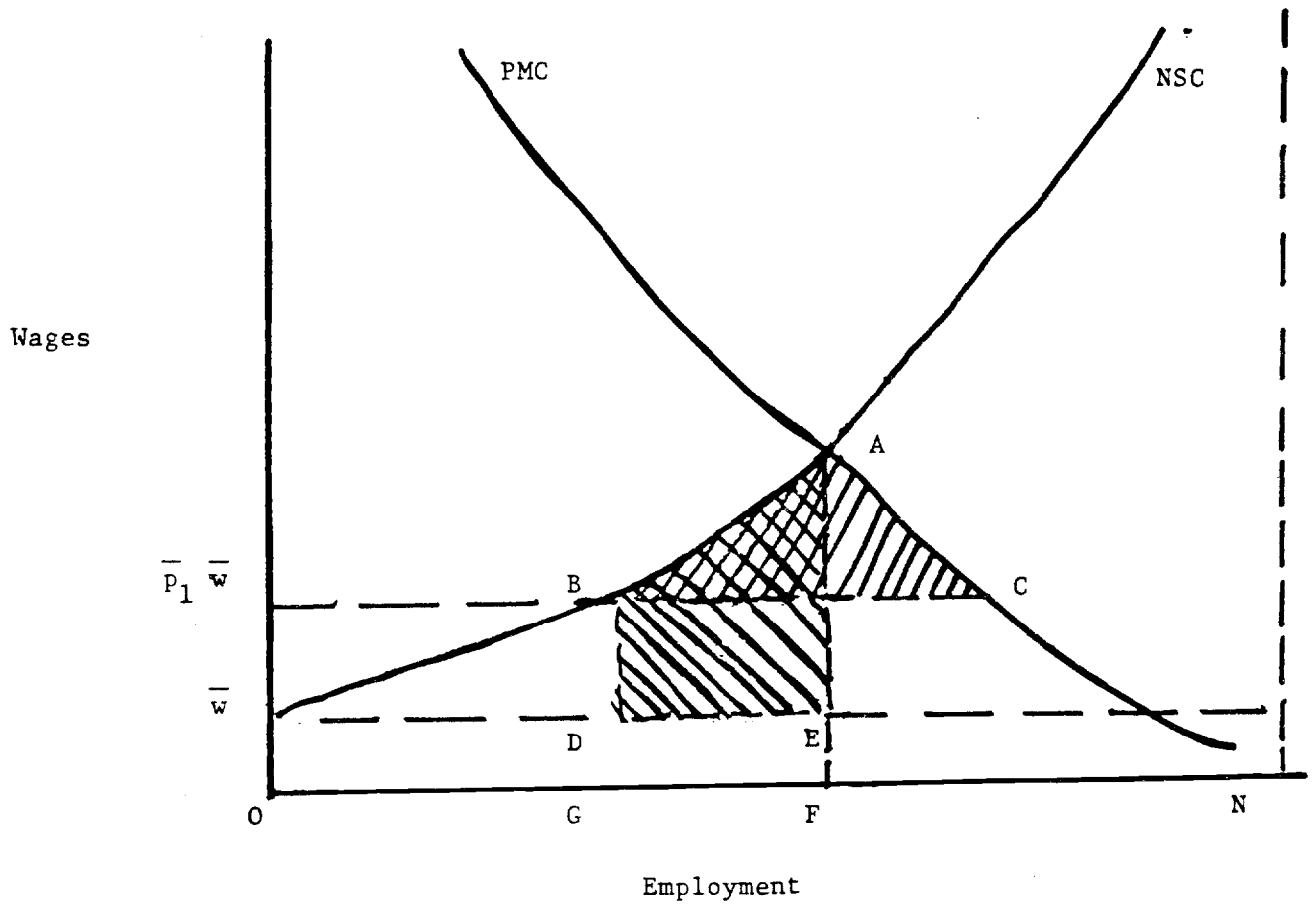O   G   F   N

Employment

Figure 3

International Competition and Welfare

loss from trade   gain from trade

allowing free trade are of second order in the change in the relative price of primary sector output. On the other hand, the gains in primary sector employment from restricting trade represent a first order welfare gain because of the wage differentials between the two sectors.

We can show this result formally by recognizing that in a free trade environment the no shirk condition (1.4) must still be met, but that product market clearing is now given by:

$$(2.3) \qquad w_1 = \bar{p}_1 \bar{w}$$

where $\bar{p}_1$ is the world relative price of the primary good. Other conditions for equilibrium include the trade balance equation

$$(2.4) \qquad \bar{p}_1 \bar{w} E_1 + \bar{w}(N - E_1) = \bar{p}_1 x_1 + x_2$$

where x_i is domestic consumption of good i and

$$(2.5) \qquad \bar{p}_1 = g\left(\frac{x_1}{x_2}\right)$$

which is a result of our assumption of homothetic tastes.

With the system of equations (1.6) and (2.3) to (2.5), we can solve for E as a function of $\bar{p}_1$ to explore the impact of world prices on primary sector employment and welfare. The result is:

$$(2.6) \qquad \frac{dW}{dp} = \frac{\left[(w_1 - \bar{w})\dfrac{dE_1}{d\bar{p}_1} + (\bar{w}E_1 - x_1)\right]}{r}$$

where W is the sum of the equivalent variations associated with a change in $\bar{p}$.

The intuitive explanation of (2.6) is that a reduction in world prices has two effects. The first term on the right side of (2.6) arises because primary

sector size is constrained by a combination of the NSC and world prices; a lower p_i reduces primary sector employment, causing a welfare loss. In a competitive model where wages were equalized, this term would not arise. The second term reflects the welfare gain from the lower foreign prices. It is equal to net imports, $(x_1-\bar{w}E_1)$ times the price drop. It is clear that if the world price is below but sufficiently close to the autarky price so that net imports are small, the loss of primary jobs and their attached rents implies that welfare is reduced by free trade. Furthermore, analysis that is substantively identical to that of (2.1) and (2.2) shows that if world primary sector prices are below but sufficiently close to autarky prices then free trade is pareto-inferior for domestic workers; the gain to secondary workers in lower primary sector prices is outweighed by the reduced opportunities for primary sector employment.

A second conclusion with clear implications for industrial policy debates follows from comparing Figures 4a and 4b. The welfare gains per dollar of revenue cost from a subsidy to the primary sector are greater in an open economy that would have had an autarky price less than or equal to the free trade price than they would be in a closed economy. This is because the demand for primary sector output is perfectly elastic in the case of a small open economy. Therefore, for any subsidy per primary sector employee more people will switch to the primary sector, and marginal worker shifts will provide as much of an income gain as inframarginal movements. Subsidies in a closed economy run into diminishing returns. As primary sector output expands, it comes to be valued less and less. In the open economy, when primary sector output can be exported, there is no similar factor limiting the gains from subsidizing primary sector output.[17]

---

[17]This proposition and the one in the next paragraph may be verified by
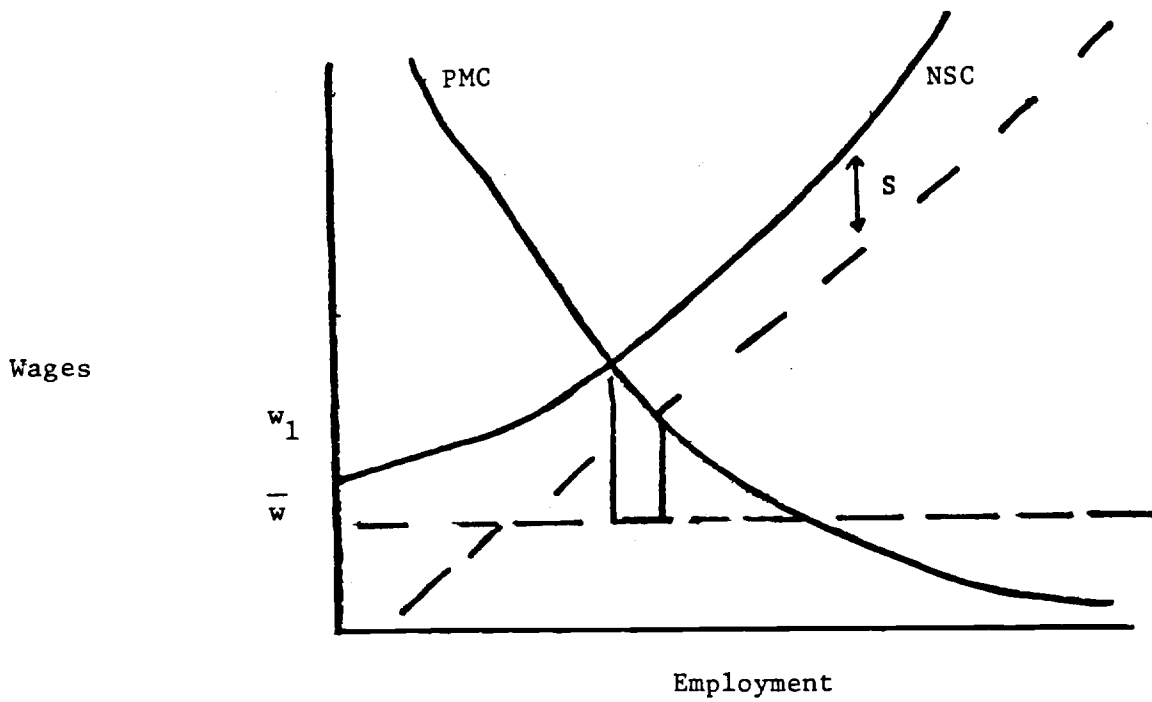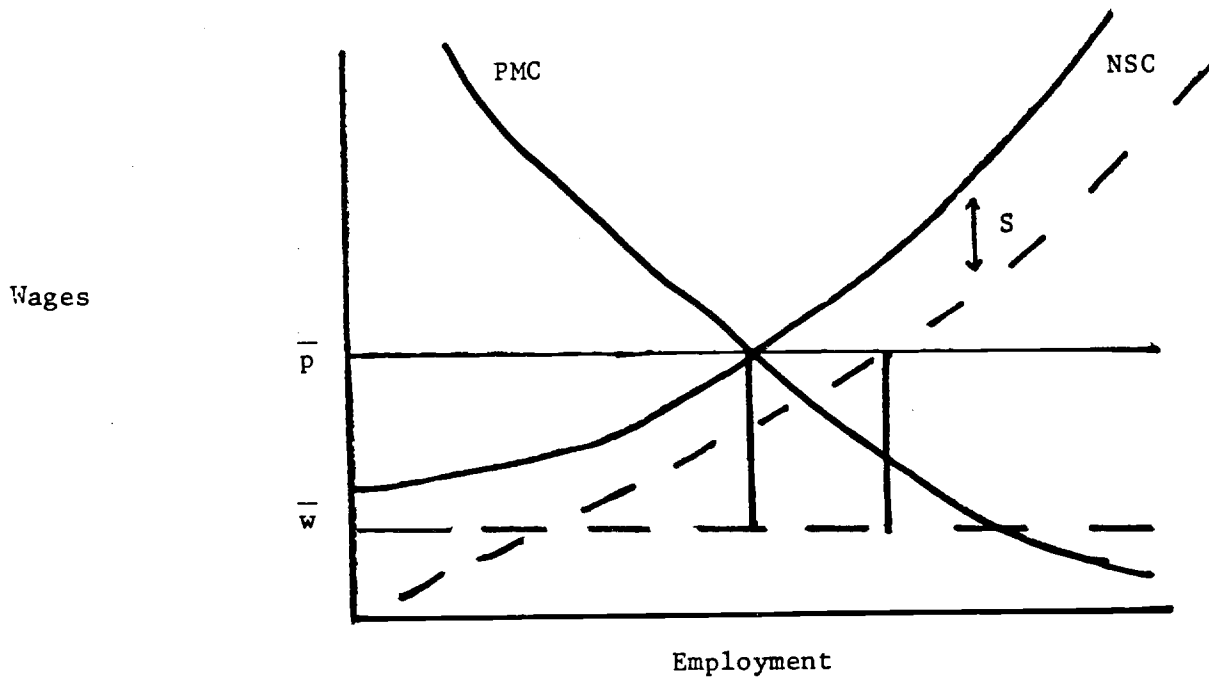
Figure 4

The Effects of Subsidies

a) autarky

b) free trade

An equally important implication of our model, which can be seen from (2.6), is that nations should subsidize "winners" -- primary sector products where they have a comparative advantage. The greater is the world price of primary sector output, the greater is the gain from increasing primary sector employment, $(w_1 - \bar{w})$. There is little advantage to be had from subsidizing primary sector employment if the world price of primary sector output is close to the secondary sector wage $\bar{w}$. In that case, there is little loss from letting workers export secondary sector output, and purchasing the primary sector good from abroad. Where the world price is high, and the home economy has a substantial comparative advantage, a subsidy yields the maximum benefit.

The model presented here captures many of the prominent themes in the industrial policy debate. There are good jobs and bad jobs even in equilibrium. Good jobs have higher wages and higher average productivity. Value-added per worker is higher in these jobs. Protecting sectors with good jobs can raise total economic welfare. International competitiveness raises the stakes and increases the payoff to primary sector subsidies. Previous models endeavoring to illuminate the industrial policy debate have focused on the strategic interactions of imperfectly competitive firms. Policies in some circumstances will help local oligopolistic firms at the expense of foreign oligopolistic firms. These seem less relevant to the issues debated than does our model which focuses on the effects of industrial policies on the number of good jobs

---

calculating dE_1/ds explicitly from equations (1.6) and (1.7) for the closed economy and (1.6) and (2.3) for the open economy.

available for workers.[18]

In models with competitive labor markets public interferences in free trade through protection or export subsidies are undesirable. Yet governments regularly engage in these practices. The model presented here presents a rationale for these actions. Obviously, an analysis of this kind taken alone cannot justify any particular industrial policy. It does however provide a basis for thinking about many of the issues raised by the industrial policy debate.

III. Discrimination in the Dual Labor Market

The persistence of discrimination in competitive markets has been difficult for economic theory to explain. As originally noted by Becker (1957) and emphasized by Arrow (1973), one would expect discrimination to be eliminated by competition. Firms that did not discriminate would have lower costs than firms which did and would therefore grow tending to eliminate discrimination. With perfect capital markets, capital would be put to its most profitable use by ending up in the hands of non-discriminatory entrepreneurs. While enduring discrimination can be rationalized by pointing to the tastes of customers, this explanation seems implausible in many settings where discrimination is thought to occur. Nor is discrimination easily explained by pointing to workers' tastes, since firms comprised only of minority group members would have lower costs than other firms. Furthermore, some evidence, Blau (1982) suggests that men working in heavily female establishments actually receive lower wages

---

[18]The formal model that we have presented here does not include capital as a factor of production. Adding capital as a factor does not change the conclusions reached so far, but leads to another conclusion. Traditional competitive models suggest that it is in the national interest to allow capital to seek the highest return available either at home or abroad. Once wage

ceteris paribus.[19]

Beyond their difficulties in explaining the persistence of discrimination, existing theories have a difficult time accommodating the phenomenon of occupational segregation. Studies of discrimination, particularly sex discrimination find that much of the discrimination takes the form of equal pay for equal work but unequal work (e.g. Lloyd and Niemi (1979)). Doeringer and Piore (1971) in their discussion of the dual labor market emphasize that members of disadvantaged groups are confined to the secondary sector. Even granting the existence of wage differentials arising from employer discrimination, it is hard to see why occupational segregation should result.

The dual labor market model developed here has as its central element wage differentials that are unrelated to productivity differentials. It is natural to suspect therefore that it can provide the basis for a theory of discrimination. Indeed Yellen (1984) suggests that in an environment like the one considered here employers can costlessly discriminate since there is excess demand for primary sector jobs. This is not correct. If the equilibrium allocation of workers in different groups who are otherwise identical differs, firms can lower the wage offered to disadvantaged workers without fear that they will shirk. As noted in Section 1, an increase in the fraction of workers confined to the secondary sector reduces the wage that must be paid to satisfy the no shirk condition. Competition will eliminate discrimination as firms hire

---

differentials are considered, this conclusion is no longer valid. Keeping capital at home, and in the primary sector, may raise welfare by increasing rents created by primary sector jobs.

[19]Phelps (1972) and others have suggested theories of statistical discrimination. As emphasized by Aigner and Cain (1977) these models have the disquieting implication that the difference in average wages between groups exactly equals differences in productivity. Alternatively discrimination can be based on differential measurement ability for different groups as in Lundberg

labor as inexpensively as possible. Our model thus cannot explain discrimination if it is assumed that there are no differences at all between members of different groups.

However, our dual labor market model can rationalize discrimination if it is assumed that there are group differences that are unrelated to productivity. After showing how these differences can lead to occupational segregation and wage discrimination, we investigate the welfare effects of antidiscrimination policies.

Turnover and Discrimination

A major characteristic of groups thought to be disadvantaged in the labor market is that they have very high separation rate. For example, Poterba and Summers (1984) estimate that the rate of labor force withdrawal is about 1 percent per year for males 25-59 but about 19 percent per year for women in the same age bracket.[20] Data in Marston (1976) suggest large age and race differences in separation probabilities as well. These differences have provided the basis for theories of occupational discrimination based on considerations of human capital accumulation as for example in Mincer and Polachek (1974). Here we present an alternative explanation of how higher separation rates can lead to group differences in wages. In our formulation experience has no effect on productivity. Unlike human capital based explanations for wage inequality, our model suggests that anti-discrimination

---

and Startz (1982). This seems a rather weak reed on which to base a theory of discrimination.

[20]Some empirical studies suggest that turnover rates do not differ across groups after controlling for job characteristics. These findings are not inconsistent with the analysis presented here. If as implied by our model, high turnover workers are assigned by the market to certain types of jobs, then

policies may well increase economic efficiency.

Taking account of the fact that there are two identifiable groups in the population the basic equilibrium conditions require that workers be paid their marginal products and that both men and women in the primary sector be induced not to shirk. Thus equilibrium is characterized by the conditions:

$$(3.1) \qquad w_1 = \bar{w}g(E_{1m}+E_{1f}) = p_1(E_{1m}+E_{1f})$$

$$(3.2) \qquad (w_1-\bar{w}) = \frac{\alpha r}{d_2-d_1} + \frac{\alpha r(d_1+q_m)N_m}{(d_2-d_1)E_{2m}}$$

$$(3.2) \qquad (w_1-\bar{w}) = \frac{\alpha r}{d_2-d_1} + \frac{\alpha r(d_1+q_f)N_f}{(d_2-d_1)E_{2f}}$$

where the subscripts m and f are used to denote males and females, for concreteness in the discussion. Note that competition insures that primary sector men and women receive the same wages. However, comparing equations (3.2) and (3.3) it is clear that a higher proportion of women will be confined to the secondary sector if q_f is less than q_m. Since women have a shorter horizon in primary sector jobs, they must receive a greater inducement if they are not to shirk. With equal wages, this can only occur if secondary sector women have a smaller chance of moving to the primary sector than do secondary sector men. These points are illustrated in Figure 5. As long as the NSC for women lies above the NSC for men, there will be partial occupational segregation.

Figure 5 suggests an alternative theory of occupational discrimination. If

---

controlling for job characteristics is inappropriate. Moreover, an analysis depends on workers' expected horizon in the labor force not on the current job. Here the differences between groups are especially pronounced.
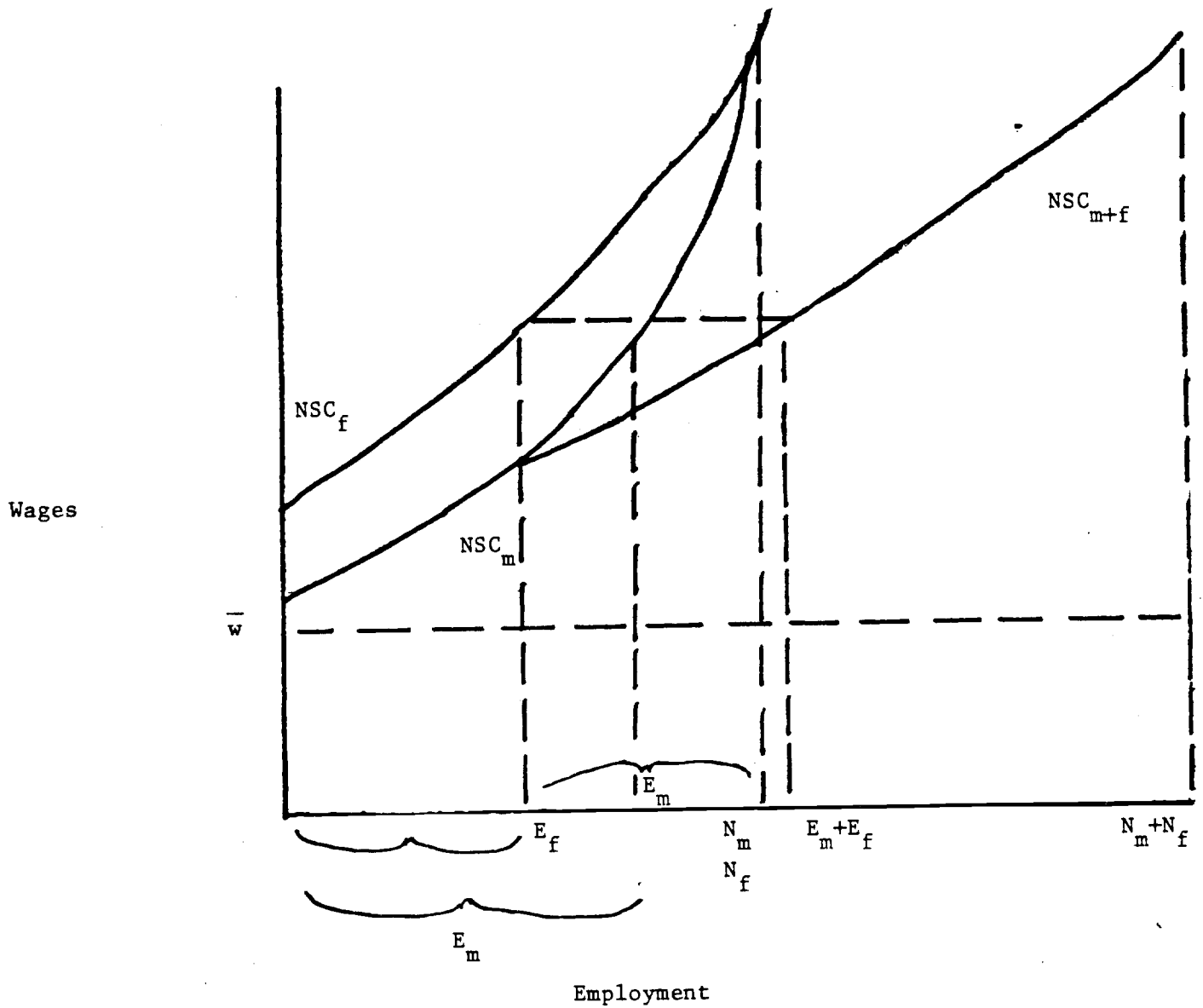
Figure 5

Discrimination in the Dual Labor Market

(drawn assuming $N_f = N_m$)

groups differ in the utility they get from being in the secondary sector, which may be taken to include home production, occupational segregation will result. The model captures a commonly observed aspect of firm behavior. Firms prefer to give jobs to workers who "really need them" than to workers who gain less surplus from holding them.

It is consistent with the observation emphasized by Blau (1982), marriage raises the wages of men but reduces or has no effect on the earnings of women. Marriage reduces male probability of employment separation but increases females' separation probabilities. It is also plausible that marriage raises the cost to men of losing a desirable job, but reduces the costs to women for reasons relating to the analysis of Becker (1985). The differential effects of marriage on men and women mitigates against theories of discrimination based on employer prejudice.

Below, the welfare effects of anti-discrimination policies in this model are considered. Before turning to this analysis it should be noted that our framework suggests a number of other theories of discrimination. One such theory might be based on the differential ability of different groups to enter into efficient labor contracts in the primary sector. Population groups that are liquidity constrained and therefore unable to accept low starting wages and rising age-wage profiles will not be able to get primary sector jobs. The wage they must be paid in order to get them not to shirk may well exceed the going primary sector wage. For example, historically some workers may not have been able to afford apprenticeships.

An argument paralleling theories of statistical discrimination might run as follows. Employers cannot accurately predict workers' productivity. Therefore

they make use of a worker's demographic status in assessing likely productivity. In a perfect labor market, workers in disadvantaged groups could be offered primary sector jobs at low wages reflecting their low expected productivity, and given a chance to prove themselves. In the environment considered here, there would be no incentive for a firm to make such an offer since at low wages all workers would shirk.

Anti-Discrimination policies

It is clear from the discussion in the preceding section that a policy of subsidizing employment in the primary sector would continue to be desirable with two classes of workers. A more interesting question involves a policy of subsidizing the primary sector employment of disadvantaged workers, and taxing the primary sector employment of advantaged workers in a balanced budget manner. Equivalently, what are the normative properties of policies which require that disadvantaged workers and advantaged workers be hired in some fixed proportions in the primary sector? We examine these questions in the context of the model presented above where discrimination arises from differences in separation probabilities, and then consider other possible sources of discrimination.

We consider the effects of a tax at rate t on male employment in the primary sector used finance a subsidy to women's employment. It follows that the subsidy to women's employment is at rate E_{IM}/E_{IF}. In order to show that at least a small tax is desirable, we proceed as follows. Since men and women have the same utility function, a tax increase will be desirable if $d(E_{IM}+E_{IF})/dt > 0$.

Substituting equation (3.1) into (3.2) and (3.3) and introducing the tax into the latter two equations allows us to implictly differentiate (3.2) and (3.3) to find $d(E_{IM}+E_{IF})/dt$. After some tedious algebra this yields

$$(3.4) \quad \frac{\frac{d(E_{1M}+E_{1F})}{dt} = \frac{\alpha r}{d_2-d_1} \cdot \frac{1}{E_{1f}} \left[\frac{(d_1+q_m)N_M E_{1M}}{(N_M-E_{1M})^2} - \frac{(d_1+q_f)N_f E_{1f}}{(N_f-E_{1f})^2} \right] - \frac{t(E_{1M}+E_{1f})}{E_{1f}^2}}{\underset{+}{|\Delta|}}$$

At $t = 0$
$$\frac{(d_1+q_M)N_M}{N_M-E_{1M}} = \frac{(d_1+q_f)N_f}{N_f-E_{1f}}$$

So

$$(3.5) \quad \frac{d(E_{1M}+E_{1f})}{dt} = \frac{\frac{\alpha r}{d_2-d_1} \cdot \frac{1}{E_f} \cdot \frac{(e_1+q_M)N_M}{N_M-E_{1M}} \left[\frac{E_{1M}}{N_M-E_{1M}} - \frac{E_{1f}}{N_f-E_{1f}} \right]}{|\Delta|} > 0$$

Equation (3.5) is clearly positive, because the ratio of male primary participation, $E_{1M}/(N_M-E_{1M})$, exceeds the ratio of female primary sector participation at a tax rate of zero. Its implication is that a small subsidy to women will increase welfare.

However, we can also calculate whether an optimal policy would produce a large enough subsidy to drive the female primary sector participation rate up to the male level. At the point where $E_{1M}/N_M = E_{1f}/N_f$ we find

$$(3.6) \quad \frac{d(E_{1M}+E_{1f})}{dt} = \frac{\frac{\alpha r}{d_2-d_1} \cdot \frac{N_f}{(N_f-E_{1f})^2} (q_M-q_f) - \frac{t(E_{1M}+E_{1f})}{E_{1f}^2}}{|\Delta|}$$

Because $q_f > q_m$, expression (3.6) must be negative. Thus, starting with t positive to the point where the sectoral composition of employment is the same for men and women, a reduction in the subsidy to the hiring of women will raise welfare. A policy intermediate between the laissez faire outcome and the elimination of apparent discrimination is optimal in the sense of maximizing the

sum of utilities.

These results may be justified intuitively. It is clear from Figure 5 that in the absence of any subsidy the NSC for women is less steep than the NSC for men. Hence a policy of taxing each man in the primary sector $1 and subsidizing each woman in the primary sector by E_n/E_f will be a net revenue raiser and will increase total employment. If no revenues were raised, but all the proceeds were used to subsidize the employment of women, the employment gains would be even greater. On the other hand, if the share of men and women in the primary sector is equalized, equations (3.2) and (3.3) imply that the NSC is flatter for men than for women. This is because a small increase in wages raises the present value of a job by more for men than for women because of their longer horizon. In this case, reducing the tax on men and the subsidy to women raises total employment.

It should be clear that equivalent results hold for affirmative action policies. Each possible choice of t is equivalent to a policy mandating that a specific proportion of women be hired in the primary sector. If policy is carried on through such quotas, the wages of women will rise above average primary sector wages, and male wages will fall. A quota policy which reduces male wages by t percent relative to average primary sector wages is exactly equivalent to a t percent tax on male employment with revenues used to finance subsidies to primary sector women.

An alternative policy which might be labelled "comparable worth" would artificially compress the wage differential between the primary and secondary sectors. This would be disastrous for both men and women. With a compressed wage differential between the two sectors, the no shirk condition would require

reduced primary sector employment of both men and women.

These results pertain to a model where apparent discrimination results from differences in turnover probabilities across groups. We have also considered the case where discimination results from one group getting more utility from a secondary sector job than the other. In this case, the effect of a subsidy on total welfare even starting at $t = 0$ is ambiguous. This is because a subsidy in addition to raising total employment which increases welfare causes the substitution of women for men in primary sector jobs. This reduces welfare under the maintained assumption that women value primary sector jobs less than men. We have not yet considered the welfare effects of affirmative action policies in situations where discrimination results from group differences in the ability to post bonds or is statistical. We suspect that it will be easy to construct examples where it will be welfare enhancing.

Although primary sector jobs are rationed in our model, we have so far assumed that no rent seeking behavior takes place. In a sense therefore the allocation of workers to the primary sector is arbitrary. In the remainder of the paper, we relax this assumption and provide a theory of unemployment.

IV. Unemployment in the Dual Labor Market

Most economists accept the idea of involuntary unemployment. Some workers who would like to work at the going wage are unable to find work. Yet as many critics have pointed out, the concept of involuntary unemployment involves an essential ambiguity. What is meant by the going wage? The option of self-employment is always open to the unemployed. And even in times of stagnation, some jobs are available. We propose here a formulation which

resolves these ambiguities. This is done by assuming that primary sector employers only hire workers directly from the pool of unemployed workers. This generates "wait unemployment" as workers queue for primary sector jobs.

Imposing the requirement that workers must be unemployed before they can accept primary sector employment, we have the equilibrium conditions:

$$(4.1) \qquad w_1 = \bar{w} g_1(E_1/E_2)$$

$$(4.2) \qquad \alpha = \frac{(p_1-1)\bar{w}(d_2-d_1)}{d_1 + q + r}$$

$$(4.3) \qquad \bar{w} = \frac{(d_1+q)E_1}{N-E_1-E_2} \cdot \frac{(p_1-1)\bar{w}}{d_1+q+r}$$

recalling from before that the utility from being unemployed (having no income) has been normalized to zero.

These equilibrium conditions are easily interperted by recognizing that in equilibrium, secondary sector workers must be indifferent to becoming unemployed and unemployed workers must be indifferent to accepting secondary sector jobs. Equation (4.1) simply determines the relative price of primary sector output. Note that we can no longer write it as $g(E_1)$ because now some workers are unemployed. Equation (4.2) holds that the utility gain from shirking in the primary sector must equal the present value of holding a job in the primary sector. This is most easily reckoned by considering the alternative of being permanently confined to the secondary sector. Equation (4.3) holds that the difference in instantaneous utility between holding a secondary sector job and remaining unemployed equals the instantaneous probability of obtaining a primary sector job times the present value of such a job. This formulation differs

somewhat from that of Harris and Todaro (1970) who assume that workers must be indifferent between secondary sector employment and locating in the primary sector with proportional chances of being employed and unemployed.

The determination of equilibrium is depicted in Figure 6. Note that equilibrium requires both that the PMC and NSC cross and that secondary sector workers be indifferent to being unemployed. The unemployment in Figure 6 corresponds to" Keynesian unemployment" in three senses. First, it is involuntary as workers are queuing for primary sector jobs, and are identical to primary sector job holders but cannot bid down wages and get primary sector jobs. It is voluntary only in the sense that all unemployment is voluntary -- some undesirable jobs are available.

Second, relative to the first best the level of unemployment is too high and the level of primary sector employment is too low. As we discuss more formally below, subsidies to employment or taxes on unemployment will increase welfare.

Third, if productivity varies, unemployment, wages and aggregate consumption will all move pro-cyclically as long as product demand is not too elastic and/or the primary sector is not too small. This proposition may be examined by substituting (4.1) into (4.2) and (4.3) and implicitly differentiating the latter two equations with respect to $\bar{w}$. We find:

$$(4.4) \quad \frac{d(E_1 + E_2)}{d\bar{w}} = \frac{(p_1 - 1) \left[\frac{(d_1 + q + r)}{(d_1 + q)} \frac{N - E_1 - E_2}{E_1^2} + p_1' \frac{E_1 + E_2}{E_2^2} \right]}{\frac{p_1'(d_1 + q + r)\bar{w}N}{(d_1 + q)E_1 E_2^2}}$$

The denominator of (4.4) is negative (because $p_1' < 0$) so whether total

Wages

w_1

$\overline{w}$

PMC

NSC

E_1
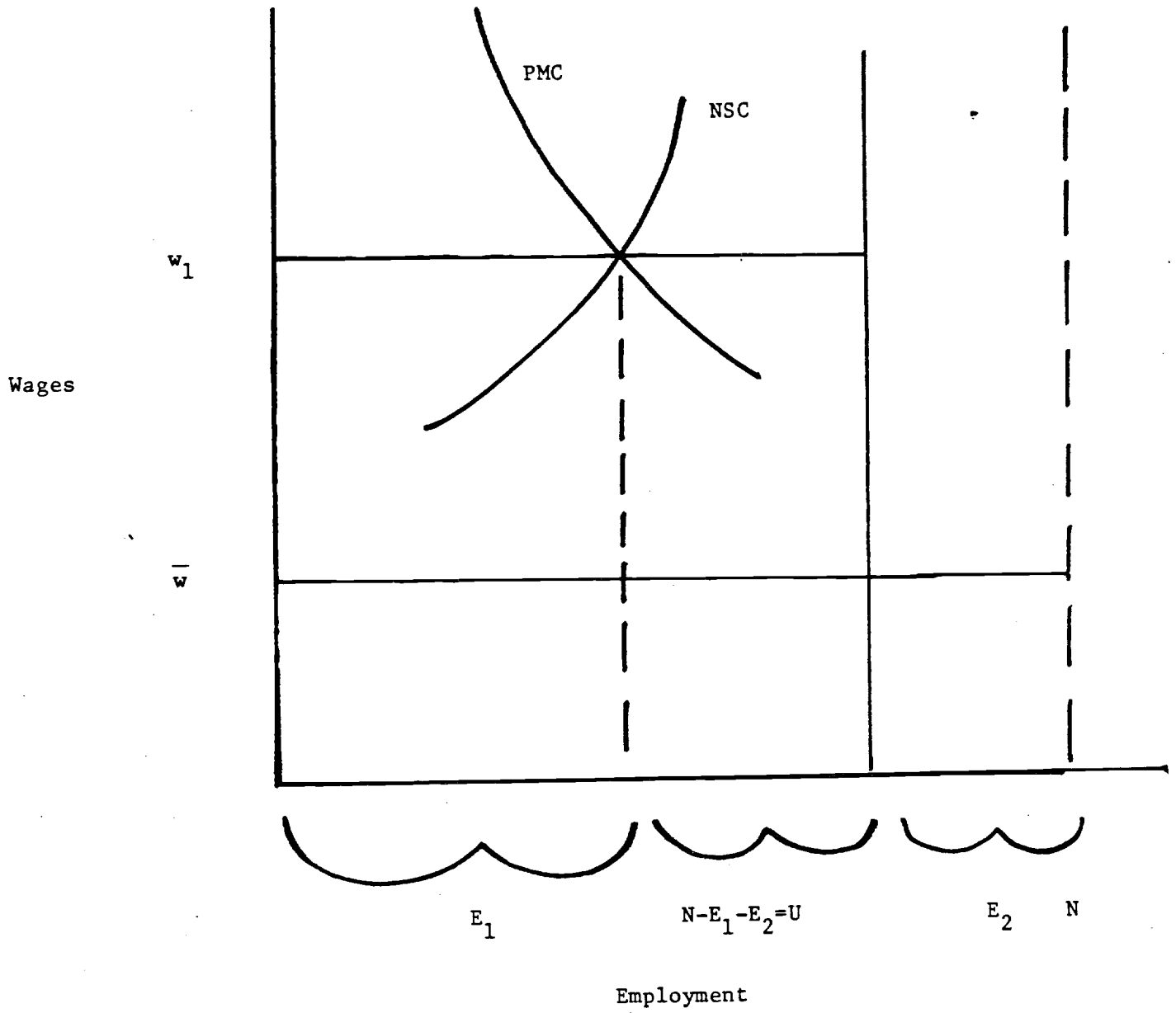
$N-E_1-E_2=U$

E_2 N

Employment

Figure 6

The Determination of Equilibrium Unemployment

employment $E_1 + E_2$ rises depends on whether

$$\frac{(d_1+q+r)}{d_1 + q} \frac{N-E_1-E_2}{E_1^2} + p_1' \frac{(E_1+E_2)}{E_2^2} < 0$$

If $p_1' \to -\infty$ (consumers have very inelastic relative demands for the two goods), then E_1 and E_2 both rise proportionately and $N - E_1 - E_2$ falls. But if $p_1' \to 0$ so that relative prices will not change much, the primary sector will expand in response to a positive productivity shock while constant relative prices will imply the same ratio of $N-E_1-E_2/E_1$ as before the shock also unemployment will be increased.

If the primary sector already dominates the economy (E_1/E_2 large) then an increase in $\bar{w}$ will, by its reduction in the primary sector unemployment rate, raise total employment. If, however, the economy is dominated by the secondary sector, the principle effect of a shock will be to move workers into the primary sector where the unemployment rate is higher.

The Cyclical Behavior of Labor Markets

The response of the economy considered here to productivity shocks corresponds well to the cyclical behavior of actual labor markets as summarized in Okun (1982). First, Okun noted that quits are pro-cyclical while employer initiated separations move counter-cyclically. This is predicted by our model. Positive productivity shocks lead to quits as workers leave the secondary sector, while negative shocks lead to job loss as the primary sector contracts. Second, Okun observed the shrinking of relative wage differentials at cyclical peaks. In the model here, positive productivity shocks lead to equal <u>absolute</u>

increases in wages in the two sectors. This corresponds to a narrowing of relative wage differentials. Third, he emphasized the enduring nature of attachments between workers and firms. As emphasized in Section 1 our formulation provides an explanation for firms use of temporary layoffs and for labor hoarding. Both serve to increase workers' horizon and therefore limit the wage that must be paid to insure that workers do not shirk. A fourth observation most sharply distinguishes our model firms from alternative accounts of employment fluctuations.

The available evidence suggests that the vast majority of cyclical variations in labor input reflect variations in the number of persons working rather than the number of hours worked per person. Worksharing arrangements in the face of downturns in demand are a rarity in the American economy. While it might be argued that technological considerations make short hours infeasible as a worksharing device, it is hard to see why employers who must reduce the size of their work force cannot use rotating layoffs as a worksharing arrangement. As Okun observes, the absence of worksharing antedates any incentives for full layoffs provided by unemployment insurance and any important effects of unionization.

These observations are troubling for many accounts of cyclical fluctuations. It is hard to square discontinuous movements into and out of full time employment with utility maximiation by a representative consumer. Worksharing arrangements are a natural device for performing the insurance function which is central to many implicit contract theories. And formulations which focus on considerations of fairness and equity in determining labor market practices would tend to predict the evolution of worksharing arrangements.

The absence of worksharing arrangements is predicted by the model considered here. Assume that a worker holding a fraction (1-z) of his time to searching for a full time job, and assume that his search efficiency is 1-z times that of a full time worker. Available evidence on job search, Clark and Summers (1979), suggests that in fact part time primary sector workers could probably search about as efficiently as the full time employed. Assuming that dismissal rates are proportional to hours on the job (so that a non-shirking part-timer is less likely to be erroneously caught shirking than a non-shirking full timer) but that quit rates are related to calendar time on the job, the no-shirk wage per unit of time worked for a z-time employee, w_z^*, in an economy where all other primary jobs are full-time is given by:

$$(4.5) \qquad w_z^* - \bar{w} = (w_1 - \bar{w}) \left[\frac{r + e_1(z_t \frac{(1-z)E_1}{N-E_1} + q(1 + \frac{(1-z)E_1}{N-E_1})}{z(r+d_1 q)} \right] \geq w_1 - \bar{w}$$

The crucial insight is that both a full-timer and a part-time worker get the same benefit from shirking for an hour, and each is as likely to be caught. Therefore, to prevent the part-timer from shirking the employer must offer him a job which is equally valued (has the same present value) as a full-time job. Such a constraint necessarily implies that the part-time employees will require higher hourly pay than the full-timers to prevent shirking.

An important feature of actual unemployment is that it is strongly serially correlated. While we have not completed a full dynamic analysis of our model, it is clear that temporary productivity shocks can give rise to serially correlated movements in unemployment. Consider for example an increase in productivity that is known to be temporary. It is apparent that primary sector

employment must move continuously after the productivity shock is observed. If a discontinuous movement were anticipated when productivity reverted to its normal level, the instantaneous separation rate would be arbitrarily high and so workers would shirk unless wages were infinite. It follows that a temporary productivity shock would give rise to an immediate increase in primary sector employment followed by a slow decline. Notice that primary sector wages would rise more and employment less for a temporary productivity shock than for a permanent productivity shock. This is because the knowledge that the shock was temporary would reduce workers' horizons.

Stuctural Unemployment

While our model is able to account for a number of features of cyclical fluctuations, it does not provide a satisfactory eplanation of the driving force behind fluctuations. The work of the real business cycle school notwithstanding, we find it implausible that productivity shocks are the primary cause of cyclical fluctuuations. Aggregate demand, which cannot be defined in a real model like ours with competitive product markets, seems to matter in cyclical fluctuations. We return to this issue in the concluding section of the paper. It may be more appropriate to think of our model as providing insights into the determinants of "natural" or "structural" unemployment. Indeed our model suggests some hypotheses about the dramatic increase in structural unemployment observed in Europe over the last decade.

Consider first the effect of an increase in social insurance benefits. Higher benefits may be thought of as corresponding to an increase in utility from being unemployed a tax on wages. Raising unemployment benefits increases

the incentive of secondary sector employees to become unemployed. This effect is accentuated by the taxes necessary to finance insurance benefits. Reduced secondary sector output leads to a reduced relative price for primary sector output and lower primary sector wages. This in turn causes primary sector output to contract, further increasing unemployment. The analysis parallels the analysis of a negative productivity shock exactly except for the further distortions caused by the taxes necessary to finance unemployment insurance.

Notice that the adverse effect of unemployment insurance on unemployment occurs despite the maintained assumption that it does not affect workers' search decisions and that its financing does not affect employers' decisions about layoffs. Rather, its effect occurs primarily through firms' wage setting behavior. This model provides a possible basis for suspicions about the adverse effects of social insurance programs on macroeocnomic performance that is consistent with the relatively modest effects typically found in microeconomic studies.

A second implication of our analysis is that policies directed at increasing job security may substantially increase unemployment and reduce primary sector employment. Such policies may be thought of as reducing d_2, the likelihood that a shirking worker will be detected and laid off. As we have already seen, policies directed at worksharing may be counterproductive. If as a result of worksharing arrangements, primary sector firms must raise their wages, primary sector employment will decrease and unemployment will rise.

Third, our model suggests that public or private policies directed at compressing wage differentials may have deleterious effects. Equation (4.2), the No Shirk condition, pins down the relationship between primary and secondary

sector wages which must exist if primary workers are not to shirk, as long as secondary sector jobs are freely available. If this wage differential is reduced, primary sector firms will not be able to operate because the No Shirk condition cannot be satisfied.

V.   Conclusions

We believe that the analysis in this paper suggests the applicability of efficiency wage models to a wide variety of problems in labor economics. They can contribute to the understanding of many phenomena that are difficult to account for in other ways. And they have very different normative implications than do other labor market models. Taking account of the effort elicitation problem calls into question the general proposition that private labor market arrangements are efficient. It also complicates the analysis of various labor market policies.

A good example is provided by minimum wage laws. Taking account of the effort elicitation problem introduces several new elements into the analysis of minimum wages. First, minimum wages may interfere with efficient contracting leading to unemployment and allocative distortions. If minimum wage laws constrain firms so that they cannot offer age-wage profiles which are as steep as they would like, they will have to raise the total level of compensation in order to insure that workers do not shirk. Second, minimum wage laws have a potentially beneficial effect which could conceivably offset the distortions they create. Because they will in general create some unemployment they raise the cost to being laid off from the primary sector. This will permit lower primary sector wages and greater primary sector employment. Third, minimum wage

laws may have a second potentially beneficial effect in raising productivity in firms affected by the minimum. Such firms will be able to reduce the resources which they devote to the supervision of workers. If as we argue below competitive equilibrium is characterized by excessive supervision, this may enhance social welfare.

Beyond their application to various labor market issues, there are a number of important directions for future research on efficiency wage models. We consider them in increasing order of importance. First, we and other writers on efficiency wages have made strong assumptions about the nature of compensation arrangements. It is theoretically possible that compensation arrangements can be devised which obviate the need for firms to offer more than the going wage in order to insure that workers do not shirk. It would be valuable to model explicitly the impediments to these arrangements created by the problems of worker liquidity, moral hazard on the part of firms, risk aversion on the part of employees, and inabilities of firms to make certain types of commitments. The observation that many workers are paid more than their opportunity cost requires rationalization.

A related question emphasized by Shapiro and Stiglitz (1984) concerns firms' choice of a supervision technology. It is easy to see that models of the type discussed in this paper suggest that the architecture of economic organizations will be flawed. There will be " too many chiefs and too few indians". Firms will invest in supervision until the marginal effect of a dollar spent on supervision equals the marginal effect of a dollar spent on higher wages in reducing workers' incentive to shirk. Extra wages benefit workers but extra monitoring does not so there will be more supervision than is socially

efficient. Policies which encourage firms to give workers more responsibility and to raise their wages will increase social welfare. Of course they may reduce the welfare of those with human capital which is specialized towards supervisory ability.

A second direction for future research concerns the sources of the linkages between wages and productivity. In keeping with the work of others, we have emphasized the effects of increases in the wages firms pay on the level of effort that they are able to elicit from their workers. The mechanism we have stressed is the role of the fear of job loss in inducing workers not to shirk. This may not in fact be of central importance in wage setting. Textbooks on personnel management, businessmen and union leaders all stress that equitable and fair wage policies are critical to motivating workers. While standard neoclassical economics has no role for income effects in the theory of the firm, the question of how much firms can "afford to pay" plays a prominent role in firm wage setting decisions and in collective bargaining. And the determination of the relative wages paid to workers in different jobs is a major preoccupation of actual wage setters.

Workers effort level may depend more on whether they feel their wage is just than on its absolute level. Depending on how "justice norms" are formed, results similar to those produced by our formulation may result. Even in the face of unemployment or laborers willing to work at wages below current levels firms will be reluctant to cut wages for fear of alienating current employees. Inter-industry wage differentials may persist because they have become established, and the costs of changing them for any one firm may exceed the benefits. It is also clear that justice norms may influence the type of workers

hired by firms.

An important direction for future theoretical work involves explaining cyclical fluctuations in output, that are driven by aggregate demand shocks. One promising avenue towards this objective involves the "menu costs" arguments of Mankiw (1985) and Akerlof and Yellen (1985). These authors show that if there is any cost to changing prices and if as in our model, the economy does not attain the first best level of employment, expansionary policies will have real effects and may raise welfare. Small nominal shocks will have real effects because it will not be worthwhile to change wages or prices.

An alternative direction for elucidating cyclical fluctuations would involve introducing monopolistic competition and increasing returns into the analysis of the primary sector along the lines of Kiyotaki (1985). In such a model, increases in current aggregate demand will have real effects, similar to those of a productivity shock. The demand for primary sector output will increase due to the increase in aggregate demand. Expansion of the primary sector will reduce unemployment, the size of the secondary sector or both.

Finally, our formulation provides a channel through which monetary policy can influence real output. If monetary policy can affect the interest rate, it will effect the wages firms must pay in order to insure that their workers do not shirk. Increases in real interest rates reduce the discounted value of losing a job tending to increase workers' incentive to shirk. This leads to higher wages which in turn reduce primary sector employment. Thus purely monetary changes can affect the level of unemployment as long as they affect interest rates. The empirical importance of this channel of monetary policy seems questionable, however.

Probably the most important direction for future research in this area involves empirical testing of alternative hypotheses about wage differentials. Using recently developed techniques and data sets, it should be possible to determine the extent of variations in occupational and industrial wage structures which cannot be attributed to unmeasured quality differences or compensating differentials. It would also be valuable to relate measures of worker productivity to their wages. At this point we cannot judge whether Ford's experience in instituting the $5 day was aberrant. If not, there is a strong presumption that many firms are or could profit from offering their workers non-competitive wages. Opportunites for empirical work would be greatly enhanced if satisfactory measures of monitoring costs could be constructed. A final direction for empirical tests of these ideas would involve examining the effects of the length of worker horizons on performance. Clearly such an investigation would have to confront the endogeneity of workers horizons on their jobs.

None of these empirical tests are likely to be decisive. Rationalizations can be adduced for most any labor market outcome without resorting to models of the type considered here. Ultimately the case for the utility of models based on the effort elicitation problem, must rest on the common explanation they provide for a very wide range of phenomena. Our investigation convinces us that models based on the incentive effects of non-competitive wages provide the best available formulation for understanding deviations of observed labor markets from the competitive ideal.

References

Aigner, D.J. and G.C. Cain, "Statistical Theories of Discrimination in Labor Markets," _Industrial and Labor Relations Review_ 30, 2 (January 1977), pp. 175-187.

Akerlof, G. and J. Yellen, "The Macroeconomic Consequences of Near Rational Rule of Thumb Behavior," unpublished, 1984.

Arnold, H. L. and F.L. Favarote, FORD METHODS AND THE FORD SHOPS (New York) 1915.

Arrow, K., "The Theory of Discrimination," in O. Ashenfelter and A. Rees, Eds., DISCRIMINATION IN LABOR MARKETS (Princeton, New Jersey: Princeton University Press) 1973 3-33.

Becker, G., THE ECONOMICS OF DISCRIMINATION (Chicago: University of Chicago Press) 1957.

Becker, G., "Human Capital, Effort and the Sexual Division of Labor," _Journal of Labor Economics_, January 1985.

Becker, G. and G. Stigler, "Malfeasance and the Compensation of Enforcers," _Journal of Legal Studies_, 1973.

Bhagwhati, J. and T.N. Srinivasan, LECTURES ON INTERNATIONAL TRADE THEORY (Cambridge, Mass.: MIT Press) 1983.

Blau, F., "Discrimination Against Women: Theory and Evidence," in LABOR ECONOMICS: MODERN VIEWS (Boston: Martinus Nijhoff) 1983.

Bowles, S., "The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian and Marxian Models," _American Economic Review_, March 1985.

Brown, C. and J. Medoff, "The Employer Size-Wage Effect," unpublished, 1984.

Brown, C. and J. Medoff, "Trade Unions in the Production Process," _Journal of Political Economy_, 1978.

Calvo, G., "Quasi-Walrasian Theories of Unemployment," _American Economic Review_, May 1979, pp. 102-107.

Clark, K. and L. Summers, "Labor Market Dynamics and Unemployment: A Reconsideration," _Brookings Papers on Economic Activity_ 1979:1, pp. 14-60.

Doeringer, P.B. and M.J. Piore, INTERNAL LABOR MARKETS AND MANPOWER ANALYSIS (Lexington, Mass.: D. C. Heath and Co.) 1971.

Dunlop, J., "Industrial Relations and Economics: The Common Frontier of Wage Determination," unpublished, 1985.

Freeman, R. and J. Medoff, WHAT DO UNIONS DO? (New York: Basic Books) 1984.

Gintis, H. and T. Ishikawa, "Wages, Work Discipline and Macroeconomic Equilibrium," unpublished (1984).

Hall, R., "The Rigidity of Wages and the Persistence of Unemployment," Brookings Papers on Economic Activity 1975:2, pp. 301-335.

Harris, J. and M. Todaro, "Migration, Unemployment and Development: A Two Sector Analysis," American Economic Review, March 1970, pp. 126-143.

Jencks, C., INEQUALITY, 1972.

Lawrence, C. and R. Lawrence, "Relative Wages in U.S. Manufacturing: An Endgame Interpretation," forthcoming in Brookings Papers on Economic Activity 1985:1.

Lazear, E., "Agency, Earnings Profiles, Productivity, and Hours Restrictions," American Economic Review, September 1981, 606-620.

Lee, J.R., "So Called Profit Sharing System in the Ford Plant," Annals of the American Academy of Political and Social Science, May 1916, pp. 297-310.

Lloyd, C. and B. Niemi, THE ECONOMICS OF SEX DIFFERENTIALS (New York: Columbia University Press, 1979.

Lundberg, S. and R. Startz, "Private Discrimination and Social Intervention in Competitive Labor Markets," American Economic Review, June 1983, pp. 340-347.

Marston, S., "Employment Instability and High Unemployment Rates," Brookings Papers on Economic Activity 1976:1, 169-203.

Medoff, J.L. and K.G. Abraham, "Are Those Paid More Really More Productive? The Case of Experience," Journal of Human Resources 16, 2 (Spring 1981), 186-216.

Mincer, J., "Family Investments in Human Capital: Earnings of Women," Journal of Political Economy (March/April 1974) 82, 2 (Part 2), S76-S108.

Nelson, R., "Perspectives on the Productivity Slowdown," Journal of Economic Literature, September 1981.

Nevins, Alan, FORD, THE TIMES, THE MAN, THE COMPANY (New York: Charles Scribner Sons) 1984.

Okun, A., "Prices and Quantities: A Macro-economic Analysis," Brookings Institution, 1982.

Phelps, E.S., "The Statistical Theory of Racism and Sexism," American Economic Review, 62, 4, September 1972, 659-661

Poterba, J. and L. Summers, "Adjusting the Gross Flows Data: Implications for Labor Market Dynamics," NBER Working Paper 1984.

Rock, Milton (ed.), Handbook of Wage and Salary Administration, 1984.

Shapiro, C. and J. Stiglitz, "Involuntary Unemployment as a Worker Discipline Device," American Economic Review, June 1984.

Taubman, P., KINOMETRICS: DETERMINANTS OF SOCIO-ECONOMIC SUCCESS WITHIN AND BETWEEN FAMILIES (North Holland), 1977

Thurow, L., GENERATING INEQUALITY (New York: Basic Books) 1976

U.S. Department of Health, Education and Welfare, WORKING IN AMERICA, Task Force Report, 1973.

Weitzman, M., THE SHARE ECONOMY (Boston: Harvard University Press) 1984.

Yellen, J., "Efficiency Wage Models of Unemployment," American Economic Review, vol. 74, May 1984, 200-208

</ref_section>