Volume Title: NBER Macroeconomics Annual 1986, Volume 1

Volume Author/Editor: Stanley Fischer, editor

Volume Publisher: MIT Press

Volume ISBN: 0-262-06105-8

Volume URL: http://www.nber.org/books/fisc86-1

Publication Date: 1986

Chapter Title: Efficiency Wage Theories: A Partial Evaluation

Chapter Author: Lawrence F. Katz

Chapter URL: http://www.nber.org/chapters/c4248

Chapter pages in book: (p. 235 - 290)

Lawrence F. Katz
UNIVERSITY OF CALIFORNIA, BERKELEY, AND NBER

Efficiency Wage Theories: A Partial Evaluation

1. *Introduction*

The question of why unemployed workers are unable to bid down the wages of seemingly comparable employed workers and gain jobs has long perplexed economists. A burgeoning literature on efficiency wage theories suggests that the answer may lie in the negative incentive effects of low wages. The basic efficiency wage hypothesis states that workers' productivities depend positively on their wages. If this is the case, firms may find it profitable to pay wages in excess of market clearing. This is possible because the wage that minimizes a firm's labor costs per efficiency unit of labor may not be the wage that clears the labor market. Employers may be quite reluctant to cut wages, even in the presence of an excess supply of labor, since reducing wages may actually lower productivity more than proportionately and increase labor costs. Equilibrium can therefore be consistent with persistent involuntary unemployment in some versions of these models.[1]

A variety of conceptually distinct, although potentially complementary, explanations for the direct relationship between wages and productivity have been analyzed in the literature. These approaches are based on the potential benefits to the firm of higher wages: increased

1. If efficiency wage considerations are equally important in all sectors of the economy, involuntary unemployment can arise—similar workers being treated differently—some employed and others unemployed and the unemployed preferring to be employed. If efficiency wage problems are not important in some sectors, jobs may always be available there. Jobs in the efficiency wage sector will still be rationed and offer a positive utility differential. Equivalent workers are treated differently even if there are always some (typically low-quality) jobs available. Unemployment may result from workers searching and waiting for the better, rationed jobs. See Mookherjee (1984b) for an interesting discussion of alternative concepts of involuntary unemployment.

effort level and reduced shirking by employees; lower turnover costs; a higher-quality labor force; and improved morale, more easily facilitated teamwork, and greater feelings of loyalty by workers to the firm. These economic gains to an employer of high-wage policies have long been stressed by institutional labor economists (Dunlop 1985; Reynolds 1978, chapter 9).

Alternative rationales for the payment of noncompetitive wage premiums relate to the presence of unions or threat of collective action by workers. Firms may find it profitable to pay greater than competitive wages to unionized workers to maintain industrial peace.[2] Industrial relations and human resource specialists and institutional economists have long argued that nonunion firms often pay higher wages than necessary to attract qualified labor for the purpose of avoiding unionization.[3] Dickens (1986) develops a model of the impact of the threat of collective action by workers on wages and employment which closely resembles efficiency wage models.

Efficiency wage models have been advanced in recent literature as providing a coherent explanation of normal unemployment. Some authors have even argued that these theories provide solid microfoundations for Keynesian propositions concerning the importance of wage rigidity and the existence of cyclically varying levels of involuntary unemployment.[4] Efficiency wage considerations also provide a potential explanation for large and persistent "noncompetitive" wage differentials across firms and industries for workers with similar productive characteristics. Bulow and Summers (1986) argue that wage differentials arising from efficiency wage reasons may provide a justification for trade and industrial policies designed to protect and subsidize sectors with high-wage jobs.

In this article, I survey recent developments in the efficiency wage literature and discuss theoretical and practical shortcomings of the models.[5]

2. Freeman and Medoff (1981) and Lewis (1982) provide detailed surveys of empirical studies of union relative wage impacts.
3. Foulkes (1980) presents numerous examples of large nonunion firms that maintain high wages as part of explicit union avoidance strategies. Freeman and Medoff (1984, chapter 10) provide a detailed discussion of the effects of unionization on nonorganized labor in the United States.
4. See, for example, Akerlof and Yellen (1984), Bulow and Summers (1986), Jones (1985), Stiglitz (1984), and Yellen (1984).
5. Stiglitz (1984) compares and contrasts the implicit contract and the efficiency wage theoretical literatures. Akerlof and Yellen (1984), Calvo (1979), and Yellen (1984) present excellent surveys of work on efficiency wage models. This article differs from earlier surveys in that it discusses new developments in the literature, analyzes the similarities of efficiency wage and union threat effect models, and focuses more explicitly on empirical evidence that can help in determining the consistency of the predictions of efficiency wage models with actual labor market behavior.

I review a wide variety of empirical evidence on wage patterns and cyclical properties of labor markets. The consistency of the models with this evidence helps provide a partial evaluation of the usefulness of the efficiency wage approach.

The paper is organized as follows. Section 2 presents a rudimentary efficiency wage model and discusses some of the basic implications of the efficiency wage hypothesis. Alternative models with efficiency wage structures, their empirical predictions, and theoretical shortcomings are analyzed in section 3. Efficiency wage models in which firms are assumed to be able to utilize only simple wage schemes as compensation mechanisms are shown to be capable of generating a number of important labor market phenomena including involuntary unemployment, dual (segmented) labor markets, and wage distributions for workers with identical productive characteristics. The same problems, such as the inability of firms to monitor worker performance costlessly and costly turnover, that give rise to efficiency wage payments above the market clearing level create incentives for the use of alternative incentive devices and the development of internal labor markets and long-term contractual relationships in the labor market. Alternative forms of labor contracts, typically involving the posting of performance bonds, can eliminate the job rationing that arises in versions of the models in which firms are limited to the use of simple wage policies. Practical problems arising from capital market imperfections and moral hazard problems on the part of firms may limit the potential for alternative compensation arrangements to eliminate efficiency wage problems.

A wide variety of evidence on interindustry wage differences is analyzed in section 4. Efficiency wage models make strong predictions concerning the existence of wage differentials arising from differences across industries in the wage-productivity relationship. Important systematic wage differentials across industries do not appear to be easily explained by the standard competitive rationales of differences in labor quality, compensating differentials, or transitory disturbances. Although no single efficiency wage model seems consistent with all the facts, efficiency wage models do appear useful in explaining the observed pattern of wage differentials. The consistency of efficiency wage theories with evidence on the cyclical behavior of labor markets and on labor market discrimination is also discussed in section 4.

Section 5 discusses the mechanisms through which efficiency wage models may help explain wage rigidity and cyclical fluctuations. The models explain why firms may not lose much if they fail to adjust wages to shocks. The addition of small costs of changing prices and wages, as emphasized by Mankiw (1985) and Blanchard and Kiyotaki (1985), or of

near-rational inertial behavior, as analyzed by Akerlof and Yellen (1985a, 1985b), to efficiency wage models leads to a potential model of cyclical fluctuations in response to aggregate demand movements. Concluding remarks concerning the usefulness of the efficiency wage approach are presented in section 6.

2. *The Basic Efficiency Wage Hypothesis*

Some of the primary implications of efficiency wage models can be illuminated in a simple model in which a worker's physical health and therefore productivity is assumed to depend positively on the real wage paid. This formulation was advanced by Leibenstein (1957) to highlight the linkages among wages, nutrition, and health in less-developed countries. Firms, in this context, get healthier, more productive workers if they pay higher wages. Solow (1979) formulates a formally similar model for developed economies in which increased wages improve morale and thus directly affect productivity through an increase in worker effort.

Following Yellen (1984), I consider an economy with identical, perfectly competitive firms, each possessing a short-run production function of the form $Q = aF(e(w)L)$ where e is the effort (or efficiency) level of a worker, L is the number of employees, w is the real wage, a is a productivity shifter, and Q is output. The price of output is taken to be the numeraire. All workers are assumed to have identical wage-productivity relationships of the form $e(w)$ with $e' > 0$, $e(0) \leq 0$, and the elasticity of $e(w)$ with respect to w declining in w.[6]

A profit-maximizing firm, able to hire all the labor it wants at the wage it chooses to offer, solves the following problem:

$$\max_{w,L} aF(e(w)L) - wL \quad (1)$$

The solution to the problem yields

$$e'(w^*)w^*/e(w^*) = 1 \quad (2)$$

and

$$e(w^*)aF'(e(w^*)L) = w^*. \quad (3)$$

6. Akerlof and Yellen (1984) provide economic interpretations of these conditions required on the $e(w)$ function for a sensible solution to the firm's maximization problem.

The optimal wage w^* satisfies the condition that the elasticity of effort with respect to the wage is unity. The wage w^* is known as the efficiency wage since it minimizes wage costs per efficiency unit of labor. Each firm hires labor up to the point where its marginal product equals w^*.

If the aggregate demand for labor falls short of the aggregate labor supply at w^*, equilibrium will entail involuntary unemployment. Unemployed workers will strictly prefer to work at w^* rather than be unemployed, but firms will have no incentive to hire them at that wage or to lower wages. This simple version of the efficiency wage hypothesis can explain equilibrium involuntary unemployment. Real-wage rigidity also arises in this model. Changes in relative price to the firm or productivity shocks (shifts in a) do not affect the efficiency wage w^*, but lead to alterations in the level of employment.[7]

The simple efficiency wage model can easily be extended to provide potential rationales for wage differentials among workers with identical characteristics and the existence of dual labor markets. If the linkages between wages and effort differ across firms, then the optimal wage will differ across firms and a distribution of wages for workers with identical characteristics can arise in equilibrium. These wage differentials are not compensating differences for nonpecuniary aspects of work that directly affect workers' welfare. Dual labor markets of the type described by Doeringer and Piore (1971) can also arise if the wage-productivity relationship is more important in some sectors than in others. High wages and job rationing can arise in the sector where efficiency wage considerations are salient, while the secondary sector, where efficiency wage considerations are less important, acts as a competitive labor market.

The alternative efficiency wage models examined in the next section provide more explicit microeconomic foundations for the wage-productivity relationship in developed economies. A direct derivation of the wage-productivity relationship from assumptions concerning tastes, technology, and information structure is necessary to analyze the welfare implications of unemployment and labor market segmentation in these models.

7. Solow (1979) shows that wage rigidity of this type only arises when the real wage enters the production function in a labor-augmenting way. A general short-run production function of the form $Q = F(w,L)$ need not generate real wage rigidity with respect to these types of shocks.

3. *Variations on the Efficiency Wage Theme*

3.1. THE SHIRKING MODEL

3.1.1. The Basic Approach and Implications Employers typically have only imperfect information concerning the behavior of workers on the job. The supervision and monitoring of worker actions is costly. The punishments for substandard employee performance available to a firm are typically limited by legal constraints and social custom. Firms can suspend, demote, or fire an employee for inadequate performance or misbehavior, but imprisonment, physical torture, direct cash fines, or resort to tort or contract law for redress are simply not available options for many forms of worker malfeasance.

Under these conditions, employers must find mechanisms to elicit adequate effort from their employees. Piece rates and other direct pay-for-performance compensation schemes are often expensive to operate or impracticable since it may be difficult to observe an individual employee's contributions.[8] Firms may find it profitable in this situation to raise wages above the opportunity costs of workers. By increasing wages, firms raise the cost of job loss and encourage workers to put forth adequate effort. When workers are paid wages above their opportunity costs, they value their jobs, and the threat of termination for detected loafing creates an incentive for workers not to shirk. Models in which the need of firms to elicit effort from their workers can lead to the payment of wages in excess of market clearing and generate equilibrium involuntary unemployment have recently been examined by Bowles (1985), Bulow and Summers (1986), Calvo (1979, 1985), Calvo and Wellisz (1978), Eaton and White (1982, 1983), Foster and Wan (1984), Gintis and Ishikawa (1983), Jones (1985), Shapiro and Stiglitz (1984), and Stoft (1982).

In the Shapiro and Stiglitz (1984) version of the model, firms can only imperfectly monitor their workers' job performance, and workers make a discrete choice of whether to work or shirk. Workers and firms are assumed to be homogeneous. If all firms pay the same market clearing wage, there is full employment and no cost to shirking since workers, if fired, can immediately find another job at the same wage. This strong result of no costs of shirking requires no job switching or search costs and no adverse reputational effects on workers in the labor market if they develop a poor employment history. The homogeneous workers' assumption eliminates reputational effects on workers because all workers are assumed to act the same given the same incentives. If effort is costly, all workers shirk under these full-employment conditions. Thus it pays

8. Lazear (1983) and Pencavel (1977a) analyze the major issues arising in the choice of a piece rate as opposed to a salary or time-rate compensation system.

each firm to increase its wage to eliminate shirking. When all firms do this, the average wage rises and employment is reduced. In equilibrium, all firms pay a wage above the market clearing level, creating unemployment. Since jobs are scarce and rationed, the loss of a job can involve a lengthy spell of unemployment. The reserve army of the unemployed acts as a discipline device making shirking costly. Although some unemployment is optimal in this model since it plays a required role in creating work incentives, Shapiro and Stiglitz show that the equilibrium unemployment rate is not Pareto optimal.[9] Equilibrium unemployment is involuntary in this model since identical workers are treated differently and since the unemployed strictly prefer to be employed.

The shirking model postulates a variety of factors that affect the firm's ability to extract effort from workers and consequently yields some potentially testable predictions concerning the nature of wage differentials and unemployment. Firms should pay higher wages to a given quality worker where monitoring is costly or difficult so that the probability of detecting shirking is low. Higher wages may be required for positions in which poor employee performance can cause a great deal of damage. In fact, the job evaluation systems used in the design and maintenance of wage structures in many industries rate positions on a responsibility factor that is directly related to the probable damage that could be caused by improper job performance (Milkovich and Newman 1984). Workers in positions of trust and responsibility should receive wage premiums (Eaton and White 1982). The value to a worker of keeping a job is reduced if the likelihood of a future separation is great. This means firms with monitoring difficulties should avoid hiring workers from groups believed to exhibit high turnover and should attempt to maintain long-term employment relationships, perhaps through the use of work sharing or a layoff-recall process to deal with temporary downturns. Increased variability in labor demand across sectors (greater sectoral shift activity) directly increases unemployment through more separations to facilitate labor reallocation and indirectly raises the structural unemployment rate by requiring firms to pay higher wages to prevent shirking since it induces a greater likelihood of a future separation.[10]

9. The market equilibrium is generally not efficient since firms fail to adequately take into account the impacts of their wage and monitoring levels on the policies other firms must utilize to prevent shirking by employees (Shapiro and Stiglitz 1984, 1985).

10. Bulow and Summers (1986) analyze the impact of sectoral declines on wages and employment in a standard shirking model. Lilien (1982) presents evidence which he interprets as indicating that sectoral shifts are the main contributor to cyclical unemployment fluctuations in the postwar United States. Abraham and Katz (1986) show that his evidence is consistent with standard single-factor (or aggregate-demand-driven) business cycle models. Topel and Weiss (1985) argue that increased sectoral un-

The expected income from unemployment affects the wage needed to induce proper worker behavior. A higher unemployment insurance benefit raises the required wage and reduces employment. A higher unemployment rate and hence longer expected duration of unemployment for a worker fired for shirking reduces the needed wage. Finally, the level of wages offered by other firms affects the prospects of a discharged worker. The model suggests firms should be concerned with their position in the wage hierarchy (Bulow and Summers 1986).

3.1.2. Segmented Labor Markets The shirking model also provides a rationale for dual labor markets with a utility differential for similar workers across the primary and secondary sectors and rationing of primary sector jobs. The dual labor market hypothesis states that the labor market can be roughly divided into a primary sector that offers jobs characterized by high wages and internal labor markets and a secondary sector that offers low-paying, menial jobs with little room for advancement (Doeringer and Piore 1971).

Dickens and Lang (1985a, 1985b) find in two different micro data sets that the estimation of a switching model of wage determination with unknown regimes yields two distinct wage equations. The two equations closely resemble the predictions of dual market theory for the characteristics of earning functions in the primary and secondary sectors. The equation with which most workers are associated yields significant returns to experience and education. The other equation indicates little or no return from human capital variables. The estimation technique allows the simultaneous determination of the probability of each worker's attachment to each sector and each sector's earnings equation. The procedure allows a hypothesis test that can be interpreted as a test of the rationing of primary sector jobs. Their results indicate the presence of some job rationing (particularly for minority workers).

The basic objection to the dual labor market approach is based on the argument that if secondary workers envy primary workers and are as productive, then primary sector wages should be bid down to clear the market. One possibility is that wage differences across the sectors reflect unmeasured worker quality differences. Alternatively, the shirking model provides a coherent explanation for dual markets with job rationing of "good" primary jobs even in an economy populated by homogeneous workers.

---

certainty can explain the increase in the average (or natural) rate of unemployment in the United States since the mid-1970s. Altonji and Ham (1985) survey recent empirical work on sectoral shifts and unemployment.

Bulow and Summers (1986) and Jones (1985) analyze versions of the shirking model in which the detection of shirkers is difficult in one sector of the economy (the primary sector) and monitoring is costless in the other sector (the secondary sector). This roughly fits the empirical observation that the typical primary job entails a fair degree of responsibility and independent action on the part of the employee, while most secondary jobs involve assignments that are more easily supervised. Efficiency wages above market clearing arise in the primary sector, creating a utility differential between primary and secondary jobs that creates a cost to loss of a primary sector job.[11]

Wait unemployment (as in Hall 1975) can be generated if it is easier to get a primary sector job out of unemployment than out of the secondary sector and workers line up for primary sector jobs. This is likely if a history of secondary sector employment is a bad signal to primary sector employers. Workers with values of leisure greater than the secondary sector wage but lower than the primary sector wage may also enter unemployment to line up for primary sector vacancies. Wage differentials arising from differences in monitoring difficulties across firms create incentives for search unemployment.

Internal labor markets with internal promotion ladders are likely to arise in the primary sector to maintain long-term employment relationships. The development of internal labor markets and deferred payment schemes to induce effort in the primary sector may obviate the need for the use of efficiency wages with banishment to the secondary sector or unemployment as incentive devices. This type of alternative means to motivate workers is a basic difficulty with the efficiency wage models.

3.1.3. Objections to the Shirking Model: The Bonding Critique The predictions of the shirking model concerning job rationing and involuntary unemployment arise from the dual economic functions performed by the wage. The wage serves both to allocate labor and to provide incentives for adequate employee performance (Shapiro and Stiglitz 1984). The primary objection to the shirking model is that firms have other methods to enforce employee discipline in a more efficient manner than the use of a high wage plus threat of dismissal.

A variety of labor market bonding mechanisms can potentially eliminate the need for unemployment as a worker discipline device. One direct method is for workers to post performance bonds at the time of hiring that would be forfeited if they were caught shirking. Alternatively,

11. Bulow and Summers (1986) discuss the implications for industrial policy, trade policy, and antidiscrimination policy of the noncompetitive wage differentials arising in this shirking model.

firms can pay efficiency wages but charge workers an employment (or "entrance") fee (Becker and Stigler 1974). If firms use high wages to reduce incentives for workers to shirk or steal, unemployed workers should be willing to pay entrance fees or post bonds to gain employment at these firms. The threat of forfeiting a bond or paying a new employment fee to gain employment can create work incentives and enable the market for jobs to clear, thus eliminating involuntary unemployment.

Instances of workers posting direct performance bonds or purchasing their jobs are rare, although not entirely unknown. Employment arrangements that may implicitly perform bonding functions such as upward sloping age-earnings profiles, pensions and other deferred compensation schemes, and internal promotion ladders are observed in a large segment (the primary sector) of the labor market. These mechanisms appear in large establishments where monitoring problems are likely to be important. Lazear (1979, 1981) demonstrates that seniority wage systems in which workers post a bond against cheating by accepting wages below their marginal product initially and have it returned in the form of wages above the value of their marginal product later in their careers or in the form of a pension upon retirement can solve the effort elicitation problem. If workers are unable to post upfront bonds upon taking a job (pay an employment fee), optimal deferred payment schemes are likely to involve an efficiency wage premium above the market clearing level (Akerlof and Katz 1986).

Practical objections arise to the use of complete bonding schemes in the labor market. In the first place, workers, particularly early in their working lives, face capital market constraints and lack the liquidity required to post large bonds. If the probability of detecting shirking is low, the required bond or employment fee may be substantial. Carmichael (1985) argues that even if capital markets are imperfect, firms can charge a fee sufficient to make the expected utility of the job offer equal to the value of workers' reservation wages. Although this may eliminate the directly involuntary nature of unemployment, it does not lead to an efficient level of unemployment (Shapiro and Stiglitz 1985).

Since independent verification of detected shirking is difficult, firms have an incentive to say falsely that workers are shirking and claim the bonds; firms may collect employment fees and then dismiss workers. The firm's concern for its reputation as an employer may be able to overcome this problem (Lazear 1979, 1981). The difficulty of potential workers in verifying the honesty of a firm's behavior means that the reputation mechanism is quite fragile and may be a far-from-perfect enforcement mechanism (Kreps 1984). The likelihood of firm default on the bond can be reduced if the firm does not expect to gain from falsely claiming that

the worker is shirking. For example, if the firm claims that a worker has shirked and discharges him or her, it is possible for the bond to be paid to a third party instead of to the firm. The worker is disciplined and the firm does not gain from cheating. We do not see the direct use of third-party systems like this in practice. Carmichael (1983) argues that seniority promotion rules with a fixed wage hierarchy can play this role. Additionally, tournament schemes (Bhattacharya 1983, Malcolmson 1984) may permit the firm to commit itself to a wage plan that creates the proper incentives for workers.

Reputational considerations are likely to be important precisely for the large, visible employers, such as IBM or General Motors, that provide high-paying, primary sector jobs. These large firms offer exactly the type of jobs that the shirking model indicates should pay efficiency wages and be rationed. Smaller, less visible secondary firms that may not stay in business long are unlikely to be able to get workers who will trust them not to renege on agreements concerning deferred compensation. Thus, the implication of considering possible bonding mechanisms and where they may be effective is that the secondary sector needs to pay efficiency wages while the primary sector can utilize deferred payments to reduce the required efficiency wage premiums. This unrealistic prediction of the model suggests that further work must be done on why full bonding schemes are not practicable for visible employers quite likely to value their reputations for keeping (implicit) promises to their employees.

If capital markets were perfect and third-party verification of shirking always possible, firms would spend next to nothing on raising the probability of detecting shirking and demand large bonds since monitoring is costly and the posting of bonds would be costless under these circumstances (Becker and Stigler 1974; Dickens, Katz, and Lang 1986). The empirical observation that firms devote substantial resources to monitoring workers suggests that a full bonding solution to the shirking problem is unattainable.[12] Thus, firms must be choosing one or both of two second-best alternatives to bonding: monitoring workers intensively or paying efficiency wages. The likely outcome is that firms utilize bonds to the ex-

12. Dickens, Katz, and Lang (1986) analyze these issues in far more detail. A caveat to this argument is that firms may monitor workers for reasons other than preventing shirking. Monitoring and supervision may be a way for firms to prevent costly, but honest, mistakes by workers. Monitoring may also help firms sort heterogeneous workers into tasks for which they are best suited. If firms cannot prevent mistakes and sometimes fire workers who were not actually cheating (make type II errors) and if workers are risk-averse, the firm may find it optimal to expend resources on monitoring to reduce the required bond and the wage differential needed to compensate workers for mistaken appropriations of their bonds. This is really a variation on the theme that third-party verification of detected shirking is not possible.

tent possible, and then choose the optimal combination of efficiency wages and monitoring required to prevent shirking in the presence of limited bonding ability.

Lazear (1979) argues that the existence and observed pattern of mandatory retirement provisions indicates that some bonding must be utilized in the labor market. Adverse selection problems provide an alternative rationale for the use of mandatory retirement policies. If individual performance is hard to observe, wage cuts for older workers may lead to the better workers moving to other jobs and the "lemons" remaining. Additionally, Medoff and Abraham (1980) provide evidence that experience-earnings and tenure-earnings profiles cannot be fully accounted for by rising productivity. The typical finding in cross-section wage equation estimates of a positive effect of years of tenure at current job (seniority) on earnings is often pointed to as evidence of these types of bonding arrangements, especially taken in conjunction with the Medoff and Abraham evidence. These estimates merely show that workers who have been on a given job for a longer period of time earn higher wages. This may reflect returns to seniority beyond those to general labor market experience as in the bonding stories, or it may reflect the fact that workers in good jobs or good matches earn higher wages throughout their job tenure and are less likely to quit these valuable jobs (Abraham and Farber 1986, Altonji and Shakotko 1985). Alternatively, better workers may earn more throughout their careers and have greater job tenure in any given cross-section since they may tend to be more stable. Nevertheless, pensions may be the primary labor force bonding mechanisms. Ippolito (1985) presents evidence indicating the importance of bonding through pensions.

Practical limitations on the use of alternative incentive mechanisms suggest that high wages and involuntary unemployment may be a profitable discipline device. However, the limitations on bonding devices appear least important in exactly the type of jobs the model predicts should pay efficiency wages (jobs in large, primary sector firms).

3.2. THE LABOR TURNOVER MODEL

Workers are likely to be more reluctant to quit a job the higher the (relative) wage paid by the current firm and the worse the prospects in the external labor market (e.g., the higher the aggregate unemployment rate). If firms must bear part of the costs of turnover and if quit rates are a decreasing function of wages paid, firms have an incentive to pay high wages to reduce costly labor turnover. Salop (1979) and Stiglitz (1974, 1985) formally analyze models based on these features. The formal struc-

ture of the labor turnover model is quite similar to that of the shirking model. Firms' attempts to pay high relative wages to minimize turnover costs can lead to an equilibrium with wages in excess of market clearing and involuntary unemployment serving to reduce quit rates.

The market failure in this model arises, as in the shirking model, because of the dual role played by the wage. The same wage is unable simultaneously to clear the market for new hires and the market for trained workers (Salop 1979). A seniority wage system in which new workers accept initial lower wages below their marginal product to pay for their training and hiring costs can solve the problem and eliminate involuntary unemployment. If training and hiring costs are large and are concentrated in a short initial period, employment or application fees need to be levied on new workers. Firms do not have an incentive to induce workers to quit if training must actually be provided or hiring costs entailed. Thus, the moral hazard problem on the firm's side is not as serious as in the shirking model. Capital market imperfections may make the payment of large fees impractical. Salop (1979) and Stiglitz (1984) point out that risk-averse workers are unlikely to be willing to post bonds and bear the risk of being unsuited to a job. These considerations indicate that it is realistic to assume that firms must bear part of the costs of turnover.

The model predicts that high wages will be found where hiring and training costs are formidable. These wage premiums should be associated with lower turnover rates. Stiglitz (1985) shows that the model provides an explanation for wage distributions within an industry for similar workers. Equally profitable high wage–low turnover and low wage–high turnover strategies can coexist for identical firms and workers for certain types of quit functions.[13]

3.3. THE ADVERSE SELECTION MODEL

Imperfect information by firms about the abilities of workers may provide a selection rationale for efficiency wage payments. If workers are heterogeneous in ability and if ability and reservation wages are positively correlated, firms that offer higher wages will attract higher-quality job applicants. If firms cannot observe applicant quality and lack devices to induce workers to reveal their true abilities, random hiring from the applicant pool must be done. A higher wage increases the expected ability of a worker hired randomly from the applicant pool. A wage above the market clearing level may minimize costs per efficiency unit of labor

13. See Stiglitz (1985) for a formal discussion.

under these circumstances (Stiglitz 1976, Weiss 1980). Institutional, legal, or sociological constraints preventing firms from differentiating wages across workers with different productive characteristics can lead to similar results.[14]

A basic objection to the model is that firms are likely to eventually learn a worker's ability. In this case, performance bonds can solve the adverse selection problem. The same moral hazard problem on the part of firms and capital market imperfections that limit the use of bonding for shirking problems apply in this context as well. If firms can measure performance on the job, pay-for-performance schemes eliminate the problem. Employment contracts with self-selection incentives (nonlinear wage-employment contracts) can also potentially ease adverse selection problems (Mookherjee 1984a).

3.4. SOCIOLOGICAL MODELS

Workers' effort levels may significantly depend on the extent to which they feel they are being treated fairly by their employers. The perceived justness of the wage may affect worker productivity if effort levels are linked to worker morale and feelings of loyalty to the firm.[15] Akerlof (1982, 1984) and Solow (1979) argue that wage rigidity in the face of unemployment may be due to the importance of social wage norms and other behavior not well captured by traditional individualistic utility functions. Akerlof (1984) discusses evidence from sociological studies indicating that a worker's effort level depends on the norms of his or her work group (peer pressure) and posits a number of sociologically based models with efficiency wage implications. Akerlof (1982) develops a model in which firms can raise group work norms by offering wages above the level necessary to attract a labor force. The firm's "gift" of high wages is rewarded by the "gift" of improved work norms and increased individual effort. Wages in excess of market clearing may be the outcome when wages play a dual role of both allocating labor across firms and of satisfying interpersonal and intertemporal wage norms that matter for worker performance.

Most firms pay careful attention to the perceived fairness and consistency of their internal wage structures. Doeringer and Piore (1971) find that firms devote far more resources to and place more weight in

14. The standard rate wage policies favored by many unions provide an example. Brown (1985) presents evidence on the prevalence and analyzes the implications of such policies.

15. Pencavel (1977b) presents an interesting analysis of the causes and effects of worker morale with an empirical application to British coal mining.

their wage policies on job evaluation programs designed to maintain or justify their internal wage structures than to market wage surveys used to keep wages in line with those of product or labor-market competitors. Richard Wing (1984, pp. 9–18), a former compensation director at Eastman Kodak, notes that "there is always some degree of conflict between internal and external pay equity. The position taken by most salary administrators is that internal relationships should be given first priority, and external pay relationships for certain jobs must be compromised on occasion." If certain wage relationships matter to a firm's work force, it is in the firm's interest to take those considerations into account.

The sociological models indicate that efficiency wages are likely to arise where work groups and teamwork are important. These models may also explain the direct impact of product market factors and the firm's "ability-to-pay" on wages. Worker morale and loyalty and consequently productivity may depend on the extent to which the firm shares its rents with its employees.

3.5. THE UNION THREAT MODEL

Firms are likely to face important diseconomies of scale in the hiring and training of workers. If there are costs to job search or relocation, a firm may have to increase its wage offer or reduce the quality of accepted applicants to replace a large number of workers quickly. If only a few workers leave a firm, their coworkers are likely to have overlapping firm- and job-specific knowledge and be able to train replacements. When many workers quit more or less at once, more valuable knowledge is lost per worker and no one may be left who is capable of training replacement employees. Since the costs of turnover to the firm rise rapidly as the number of workers needing to be replaced in a given interval increases, collective action can provide workers with more bargaining power than they have acting individually. Social norms, such as the willingness of customers and suppliers to boycott a struck firm, may in many settings prevent firms from doing business even if they can replace workers who attempt to act collectively. Collective action by workers may enable them to shut a business down. This bargaining power can potentially be used by workers to claim for themselves some part of the monopoly rents earned by an enterprise and (in the short run) a share of the returns on the fixed assets.

Dickens (1986) analyzes the effect of the threat of collective action by workers on wages and employment on firms that attempt to avoid collective bargaining with their employees (i.e., attempt to keep a union out). A firm can avoid unionization in the model by choosing wages and em-

ployment so that no coalition greater than or equal to a fixed fraction of the work force (in U.S. labor law a majority) can be formed around a feasible union contract. The organization of a union is assumed (realistically) to be costly to the work force. A firm can prevent collective action by paying its workers a wage as high as they would receive under a collective bargaining agreement minus the cost to the workers of organizing. Foulkes (1980) provides many examples of nonunion firms that try to keep wages close to the union wage level for comparable jobs. Dickens shows that the threat of collective action can explain why unemployed workers cannot bid down a firm's wages.[16] Unemployed workers who are hired and paid a lower wage will attempt to organize collectively. Firms may find it profitable to pay wages above the market clearing level to try to prevent unionization. Thus, the model can lead to job rationing and unemployment in a manner similar to efficiency wage models. The evidence presented by Ruback and Zimmerman (1984) that union organization drives reduce a firm's stock price and Freeman (1983) that unions are associated with lower profitability provides a strong rationale for firms to develop labor relations policies that help maintain a nonunion environment.

Firms may be able to avoid unionization in manners other than paying high wages. One possibility is the posting of a bond that is forfeited if a worker is involved in union activity. Jacoby (1983) notes that some firms in the late nineteenth century withheld some of workers' earnings. These withheld earnings were forfeited if workers went on strike. If a union were formed, the firm might extract the return of the bonds. Firms could also require workers to sign contracts barring them from engaging in collective action as a condition of employment. So-called yellow-dog contracts are not now enforceable in the United States.

The union threat model predicts that wage premiums should arise where the costs of organization are low for workers and where the potential gains from unionization are high. Product market power (larger monopoly rents per worker) and high capital/labor ratios increase the potential gains from collective action and should be associated with higher union wages and higher wages of nonunion workers with a credible threat to organize. Dickens (1986) provides a number of arguments within the framework of the model for the stability of real or nominal wages over the business cycle. International differences in labor law and the potential threat of unionization provide important identifying infor-

16. The "insider-outsider" theories of unemployment developed by Lindbeck and Snower (1984) and Solow (1985) also generate job rationing outcomes from the bargaining power of "insiders" (incumbent employees).

mation for the possible importance of collective action threats as a cause of unemployment and wage rigidity.

3.6. SUMMARY

There are several plausible explanations why firms may find it in their interest to pay wages in excess of market clearing. The mechanisms underlying these alternative efficiency wage theories are summarized in table 1. These models appear capable of explaining persistent involuntary unemployment, segmented labor markets, and wage differentials for similar workers that are not equalizing differences. The use of deferred payment schemes and internal promotion ladders within long-term employment relationships may be able to solve some efficiency wage problems without resort to job rationing. The empirical relevance of efficiency wage theories is examined in more detail in the next section.

Table 1 A SYNOPSIS OF ALTERNATIVE EFFICIENCY WAGE THEORIES

Theory	*Problems leading to efficiency wage payments*	*Benefits to firm of high wages*
Shirking	Imperfect observability of worker effort level and performance; monitoring is costly	Raise cost of job loss encouraging good performance; economize on monitoring costs
Turnover	Firms must bear part of turnover costs (hiring and training costs)	High wages reduce turnover costs if quit rate is decreasing function of wages
Adverse selection	Imperfect observability of worker quality and performance	Attract higher quality pool of applicants if more productive workers have better outside opportunities
Sociological	Morale and worker feelings of loyalty to firm depend on perceived fairness of wages	Improved work norms, morale, feelings of loyalty to firms which raise productivity
Union threat	Costs of replacing existing workforce gives employees bargaining power	Maintain industrial peace or prevent unionization

4. *Some Empirical Evidence Relating to the Usefulness of the Approach*

4.1. INTERINDUSTRY WAGE DIFFERENCES

4.1.1. Some Implications of Efficiency Wage Models for Wage Differentials. It has long been noted that there are large differences in wages across industries for apparently similar work. Substantial industry wage differentials remain even after controlling for union status and observed worker and job characteristics (Bloch and Kuskin 1978, Dickens and Katz 1986, Krueger and Summers 1986, among others). In fact, large wage differences for essentially the same type of work in the same locality have invariably been uncovered by area wage surveys. Slichter (1950) observes that hiring rates paid for common labor by 85 plants in Cleveland in February 1947 ranged from $0.50 to $1.09 an hour. He notes that this spread persisted despite the fact that rates had been compiled and distributed to firms throughout the city by the Cleveland Chamber of Commerce for over twenty years. John Dunlop (1985, p. 18) summarizes the typical finding of studies of local wage variation: "It is a well-established fact that wage rates or average hourly earnings for a defined job classification, such as maintenance electrician or key punch operator, show very wide variation in a locality, particularly in a community with a variety of industries. The top wage rates for the same job classification are often two or three times the low ones. Differences in fringe benefit programs enlarge these differences." These wage differentials provide an empirical challenge to alternative labor market models. The ability of competitive and efficiency wage models of the labor market to meet this challenge is the focus of this section, which draws heavily upon the review and analysis presented in Dickens and Katz (1986).

A standard competitive labor market model implies that persistent industry wage premiums require industry-related differences in labor quality (skill) or nonwage dimensions of work requiring equalizing differences. Reder (1962, p. 276) summarizes these predictions as follows: "In the long run, under competitive conditions, any industry will pay the same price for a given grade of labor as any other industry hiring in the same location. This remark must be qualified for differences in nonpecuniary attractions of different industries and locations. . . Therefore, in the long run, the real wage differentials among industries will reflect differences in the skill mix." Alternatively, industry wage differences at any given time for similar work may reflect transitory differentials related to shifts in labor demand across sectors and imperfect short-run labor mobility.

On the other hand, a basic implication of efficiency wage models is

that if the conditions necessitating efficiency wage payments differ across industries, then the optimal wage will differ among industries. This means that workers with identical productive characteristics are paid differently depending on their industry affiliation. These wage differences for similar workers reflect industry characteristics that do not directly affect the utility of workers and thus do not require compensating differentials. Intra-industry wage distributions for similar workers may arise from differences in the wage-productivity relationship across firms in an industry.

Each variant of the efficiency wage hypothesis potentially predicts that particular industry and firm characteristics should be associated with industry wage premiums. The shirking model leads to the prediction that wages should be high where monitoring costs are large. In these circumstances, high wages are likely to be substituted for intensive monitoring activities. Wage differentials may also be required where the costs of worker malfeasance are high. Oi (1983) suggests that higher wages are required in large establishments since monitoring is typically more difficult. The cost of foul-ups is likely to be large in industries with expensive equipment (possibly proxied by high capital/labor ratios) and for workers in positions where poor performance may affect many other workers' performances (e.g., workers in coordinating positions and workers involved in integrated production processes).

The turnover model implies that wage premiums should arise where turnover and training costs are large and that wage premiums should yield the benefit of lower quit rates. The adverse selection model predicts higher wages, after controlling for observables, where it is difficult to evaluate labor quality. The sociological (or normative) models are less specific but suggest that the importance of teamwork and ability to pay may be relevant. The importance of relative wage comparisons in some sociological models provides a rationale for long-term stability of wage differentials and for linkages in wage differentials across occupations within a firm or industry. Finally, the union threat model suggests that wage premiums arise where the costs of unionization are low to workers and where the firm has rents derived from market power or has large fixed capital investments. This means that product market power should be directly reflected in wages. Differences in industry wage premiums across occupations with important union threats (blue-collar occupations) and those with smaller threats or no possibility of unionization (managers and professional workers) provides further information on the importance of union-based models.

The primary point is that efficiency wage models predict that there should be important wage differentials not explained by compensating

differentials, labor quality differences, or shifts in labor demand across sectors. An important approach to the assessment of the empirical relevance of wage models involves the determination of the importance of noncompetitive wage differences. Further relevant information can be gained by examining the nature of these wage premiums for consistency with the predictions of individual efficiency wage alternatives.

Some initial evidence on the nature and importance of industry wage differentials is provided through cross-section estimates of industry effects utilizing individual level data. Table 2 presents estimated wage differentials for broadly defined industries based on the results of a regression of log hourly earnings on industry dummies with human capital, demographic, and locational controls for a large sample of private-sector workers from the combined 1983 Current Population Sur-

Table 2 ESTIMATED OLS LOG WAGE DIFFERENTIALS FOR ONE-DIGIT INDUSTRIES AND UNION STATUS 1983 CPS—NONAGRICULTURAL PRIVATE-SECTOR WORKERS

Variable	*Coefficient (Standard Errors)*
Mining	.289
	(.009)
Construction	.127
	(.005)
Nondurables	.050
	(.004)
Durables	.098
	(.004)
Transportation and public utilities	.154
	(.005)
Wholesale trade	.042
	(.005)
Retail trade	−.161
	(.004)
FIRE	.052
	(.005)
Services	−.064
	(.003)
Union	.192
	(.003)
Sample size	134,928
$\bar{R}^2$	.546

Controls included are education (years of schooling) and its square; experience (age-education-5) and its square; married; sex; race; part-time work; SMSA; interaction terms for both experience and its square with married, race, sex, education, part-time work and SMSA; 11 occupation dummies; and 50 state dummies.

vey (CPS).[17] The employment-weighted average of the coefficients of the industry dummy variables from the regression was calculated with the omitted industry dummy treated as having a zero effect on wages.[18] The estimated differentials presented in table 2 are the differences between the actual industry coefficients and the weighted average. These differentials indicate the proportional difference in wages between an employee in a given industry and the average employee in the sample after controlling for individual characteristics, SMSA status, and state of residence. The industry variables have a sizable impact on wages. For example, workers in mining, transportation, and public utilities earn approximately 45 and 32 percent more than (observationally) equivalent workers in retail trade. The industry wage effects are comparable in magnitude to the effect of union status.

Krueger and Summers (1986) provide evidence on industry wage differentials for a more disaggregated industry breakdown. They report industry wage premiums that range from 38 percent above the average, for the petroleum industry, to 37 percent below the average, for private household workers. Their estimates are based on the May 1984 Current Population Survey that includes a wide variety of controls for individual characteristics, union status, and occupation. Krueger and Summers find that the employment-weighted standard deviation of industry wage premiums for two-digit industries ranges from 10 to 15 percent for different years of the CPS from 1974 to 1984.

Dickens and Katz (1986) find that the industry wage differentials persist when union and nonunion workers are analyzed separately. The patterns of industry wage premiums are extremely similar for union and nonunion employees. The raw correlation of three-digit industry wage premiums for union and nonunion workers in the combined 1983 CPS sample is 0.65.[19]

4.1.2. Competitive Explanations for Wage Differentials A number of explanations consistent with standard competitive labor market models are possible for the large impact of industry affiliation on wages even after

17. The data set is described in detail in Dickens and Katz (1986). The sample consists of private sector, nonagricultural employees, 16 years of age and older. The sample combines information on the outgoing rotation groups from all twelve months of the 1983 CPS. Workers with wages less than $1 an hour and greater than $250 an hour were treated as outliers and eliminated from the sample. The results do not change qualitatively when these observations are left in the data set.
18. In other words, the weights are the number of workers (observations) in an industry in the sample.
19. This raw correlation is a biased estimate of the true correlation since it is not corrected for the fact that the wage differentials are estimated rather than known.

controlling for measured human capital variables. The first explanation posits that differences in technology across industries make it profitable to hire higher-quality workers (and hence pay higher wages) in some industries. Estimated industry wage premiums in a cross-section may primarily reflect individual-specific components of earnings capacity that are unobservable (to the econometrician) and correlated with industry status. If unmeasured ability is highly correlated with observed labor quality variables, such as years of schooling and labor-market experience, then unmeasured labor quality cannot provide an explanation for the large estimated industry wage effects. Dickens and Katz (1986) show that the size of industry wage effects is not much altered if wage equations are first estimated without industry variables and then the residuals are used to determine the industry impacts. This approach credits observed quality variables with the impacts of unobserved variables correlated with both measured quality variables and industry status.

Longitudinal data provide a potential vehicle to control for time invariant, unmeasured labor quality. If high-wage industries simply have workers of high unobserved ability and if workers of a given quality are paid equally in different industries, wage changes should not systematically be linked to changes in industry status. Longitudinal data allow one to examine the wages of a given individual as he or she switches industries. First-difference (or fixed-effects) estimation allows one to eliminate the impacts of unchanging unobserved ability components (that are rewarded equally in all industries) on the industry wage effects estimates. Krueger and Summers (1986) estimate large effects of industry switches (for broadly defined industries) on wages in first-differenced regressions using a pooled sample of matched CPS May data for 1974–75, 1977–78, and 1979–80. The estimated industry effects from the first-differenced regression are similar in direction and magnitude to pooled regression estimates. Thus, workers moving from high- to low-wage industries appear to experience wage declines and workers moving from low- to high-wage industries appear to experience wage gains.[20] Vroman (1978) reports similar results for industry switchers in the 1964–71 period, using Social Security continuous wage-history data on individuals' annual earnings. Murphy and Topel (1986) find in matched CPS data for

20. These longitudinal results (industry switch effects) are potentially consistent with models in which worker quality is heterogeneous (multidimensional) and match quality varies. If match quality is not fully revealed at the start of a match, one could generate systematic relations of industry switches and wage changes of the type found by Krueger and Summers (1986). A matching model with costly renegotiation and uncertain match quality, such as the model analyzed in Antel (1985), may also be consistent with these results if the switches from high- to low-wage industries are primarily layoffs or discharges and the moves from low to high are primarily quits.

1977–1984 that industry switchers receive only 30 percent of the wage gains that would have been predicted from the industry differentials observed in cross-sectional data. The Krueger and Summers and Murphy and Topel studies differ substantially in their sample selection rules, procedures for the correction of measurement error in industry transitions, and other control variables included in the first-differenced regressions. The existing longitudinal evidence suggests that at least a moderate portion of the industry differentials found in cross-sectional analyses reflect "true" differentials rather than just unmeasured ability.

A second possible competitive explanation is that the industry wage differentials are compensating differentials for nonwage job attributes that directly affect the utility of workers.[21] In fact, this is often the justification for the inclusion of industry dummies in estimated wage equations with individual cross-section data. Krueger and Summers (1986) find that the inclusion of 10 working conditions variables in a standard wage equation barely affects the estimated industry wage premiums.[22] Many important nonpecuniary job attributes are unlikely to be captured by their control variables. Freeman (1981) and Krueger and Summers find that the fringe benefit differentials tend to expand wage differences. Murphy and Topel (1986) find that differences in unemployment risk across industries can account for only quite a small fraction of industry wage differentials.

If industry wage premiums reflect equalizing differences, then they do not reflect rents that make jobs especially valuable to workers. The implication is that the wage premiums should not be systematically related to quit rates. Industry and individual level studies both indicate that wage premiums are strongly associated with lower quit rates (Pencavel 1970, Freeman 1980, and Krueger and Summers 1986). This suggests that industry wage premiums reflect rents to good jobs or good matches and are not merely compensating differences.[23]

21. Rosen (1985) provides a comprehensive treatment of the theory of equalizing differences in the labor market and a review of empirical studies of compensating differentials.
22. Krueger and Summers (1986) use a sample derived from the 1977 Quality of Employment Survey. The working conditions variables included are weekly hours, variables indicating dangerous or unhealthy conditions on the job, commuting time, workshift dummies, dummies indicating extent of choice of overtime, and variables indicating whether working conditions are pleasant.
23. This interpretation is clean if workers have homogeneous tastes concerning nonpecuniary aspects of work. If workers have heterogeneous preferences, then it is possible to imagine distributions of worker preferences with respect to nonwage aspects of work in which wage differentials that reflect compensating wage differentials for marginal workers may be negatively correlated with average quit rates in an industry. This means that quit rates do not depend on wage differences for marginal workers but do for inframarginal workers. A particular contrived example is the case of one disamenity that some workers mind and others do not. If enough workers care about the disamenity, a

An additional competitive explanation for industry wage premiums observed at any particular time is that they largely reflect transitory differentials created by shifts in labor demand across sectors and maintained by incomplete labor mobility in the short run. The strong stability of industry differentials over time appears to rule out transitory factors as a major component of the explanation. Cullen (1956) presents data showing remarkable stability in the industry wage structure in the United States from 1899 to 1950. He finds the rank correlation of average annual earnings for 76 manufacturing industries for the years 1899 and 1950 to be .66. Cullen finds for a group of 84 manufacturing industries that 14 of the 21 industries in the highest-wage quarter in 1899 were still in the highest-wage quarter in 1947. Also, 15 of the 21 lowest-wage industries in 1899 remained in the lowest-wage quarter in 1947. Cullen furthermore provides evidence of stability in the extent of wage dispersion over the long term. The degree of dispersion across industries was approximately the same in 1950 as it was in 1899.

This long-term stability in industry average wages may reflect stability in skill mix differences. Industry differentials for any given grade of labor could reflect responses to sectoral labor demand or supply shifts. Limited evidence available from this time period suggest that industry wage differences for particular grades of labor were fairly stable. Slichter (1950) finds the rank correlation of males' unskilled average hourly earnings for 20 manufacturing industries between 1923 and 1946 to be .73.[24]

Strong stability in interindustry wage rankings is also evident for the postwar United States. Montgomery and Stockton (1985) report that the rank correlation of mean hourly wages for 20 two-digit manufacturing industries between 1951 and 1981 was .675. Bell and Freeman (1985) find strong stability in the rankings for a group of 53 industries (both manufacturing and nonmanufacturing) from 1948 to 1982. Both Bell and

---

compensating differential may arise to compensate the marginal worker for this disamenity. Workers who do not care about the disamenity take jobs at the high-wage firms with the disamenity and earn rents. These workers have lower quit rates and reduce the average quit rate in high-wage firms. Low-wage firms without the disamenity have no workers earning rents. In this example, more inframarginal workers are at the high-wage firm and average quit rates are negatively correlated with wages. One could also construct examples going in the other direction. Thus, if workers have heterogeneous preferences, the relationship among quit rate and wage differentials may be difficult to relate to the importance of equalizing differences in the labor market.

24. Slichter uses data from the National Industrial Conference Board surveys of establishments. The unskilled wage rate applies to jobs for which no previous job training is required. Similar stability is apparent in the industry rankings of the male skilled and semiskilled wage rate for this period. This may reflect stability in skill differences given the heterogeneity of the category. Katz (1986) provides a more detailed analysis of the stability and determinants of interindustry wage structure in the pre-1950 period.

Freeman and Montgomery and Stockton note that the dispersion (log standard deviation) of industry mean wages increased substantially during the 1970s. On the other hand, Krueger and Summers (1986) find that the estimated two-digit wage premiums using individual data from the CPS and controlling for individual characteristics did not appear to have grown from 1974 to 1984. Krueger and Summers find the correlation of the estimated industry wage premiums between 1974 and 1984 to be .970. These results indicate that the rising dispersion in average industry wages since the early 1970s may largely reflect changes in the composition of labor forces across industries and possibly also changes in union wage impacts.

A further possibility is that industry wage differences arise from differences in patterns of human capital accumulation across industries. Krueger and Summers (1986) find that industry wage differentials are approximately equal in magnitude and highly correlated for young (ages 20 to 35) and older (ages 50 to 65) workers. Furthermore, the 1979 Current Population Survey contains information on job tenure (years with current employer). I used this data set to see whether industry wage differences vary with job tenure. Estimates of log earnings equations for separate tenure groups for nonunion, private sector workers indicated that industry differentials are quite similar for different tenure groups. For example, the standard deviation of one-digit industry wage differentials for workers with less than one year of tenure and for workers with more than ten years of tenure was 0.11 and 0.12 respectively.[25] The correlation (corrected for sampling error) for the differentials of the two tenure groups was .74. Large industry wage effects were apparent for entry-level workers that are close in size and highly correlated with those of long-term employees. One exception was that the differentials of all other industries versus retail trade appeared to be substantially larger for workers with long job tenure than for entry-level workers.

4.1.3. The Occupational Structure of Industry Wage Premiums Although most explanations for wage differentials provide reasons why one would expect particular occupational groups to be highly paid in some indus-

25. These are standard deviations of the estimated industry differentials from separate regressions for each tenure group of log earnings on the same set of control variables as those listed in table 2 and one-digit industry dummy variables. The standard deviations listed are unbiased standard deviations corrected for sampling error in the estimates of the industry dummy variable coefficients. The sample is private sector, nonagricultural, nonunion workers, 16 years of age and older from the May 1979 CPS. Workers without tenure data and with earnings less than $1 an hour and greater than $250 an hour were deleted from the sample. The sample size for the less-than-one-year-of-tenure group is 2770 and for the ten-years-or-more is 1912.

tries relative to others, they do not lead one to expect the pattern of wage premiums to be the same across industries for diverse occupational groups. For example, a standard competitive model suggests that an industry with dangerous production jobs may pay its blue-collar workers high wages to compensate them for the risks their jobs entail, but it does not suggest that secretaries in this industry should earn a pay premium. Working conditions, skill requirements, and monitoring problems are quite likely to differ across occupations in a firm or industry.

Dickens and Katz (1986) estimate industry wage differentials by occupation (for a 12-occupation and three-digit Census of Population industry breakdown) for nonunion, private sector workers from the combined 1983 CPS sample. The effects of human capital variables, demographic characteristics, and locational variables were constrained to be the same across occupational groups. Industry wage effects were allowed to vary by occupation. This involves the estimation of an earnings function of the following form:

$$W_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \alpha_j + \varepsilon_{ij}, \quad (4)$$

where W_{ij} is log(hourly wage) of individual i in industry-occupation cell j, $\mathbf{X}_{ij}$ is a vector of individual and locational variables for individual i, $\boldsymbol{\beta}$ is a vector of parameters, α_j is a fixed effect (or differential) for industry-occupation cell j, and ε_{ij} is an error term. This is equivalent to a wage equation with industry dummies, occupation dummies, and a full set of interaction terms between the industry and occupation dummies. The large number of industry-occupational cells implies that the feasible approach to estimating the industry differentials for each occupation is first to run a de-meaned regression in which the industry-occupation cell means are subtracted off for the dependent variable and all the independent variables:

$$W_{ij} - \bar{W}_j = (X_{ij} - \bar{\mathbf{X}}_j)\beta + u_{ij}, \quad (5)$$

where $\bar{W}_j$ is the mean of the log of hourly earnings for workers in cell j, $\bar{\mathbf{X}}_j$ is the vector of the means of the individual and locational variables for workers in cell j, and u_{ij} is a regression error. This regression, assuming that the ε_{ij} in equation (4) are uncorrelated with the X_{ij}, yields a consistent estimate $\hat{\beta}$ of β. The mean residual for each cell j is then a consistent estimate of the industry-occupation j fixed effect:

$$\hat{\alpha}_j = \bar{W}_j - \bar{\mathbf{X}}_j\hat{\beta}. \quad (6)$$

Table 3 CORRELATIONS OF INDUSTRY WAGE DIFFERENTIALS ACROSS OCCUPATIONAL GROUPS—FIXED EFFECTS 1983 CPS PRIVATE-SECTOR NONUNION WORKERS

Position	*Manage*	*Prof.*	*Tech.*	*Super.*	*Sales*	*Cleric*	*Service*	*Crafts*	*Operate*	*Semi-skilled*	*Trans. oper.*	*Labor*
Managers	1.00	.99	.60	.85	.84	.99	.84	.90	.70	.67	.80	.81
Professionals		1.00	.72	.94	.74	.90	.73	.86	.82	.77	.74	.77
Technicians			1.00	.78	.61	.76	.70	.80	.63	.37	.45	.56
Supervisors				1.00	.88	.91	.97	.99	.78	.94	.79	.90
Sales					1.00	.79	.73	.78	.41	.46	.53	.69
Clerical						1.00	.86	.90	.76	.70	.78	.88
Service							1.00	1.00*	.98	.75	.74	.89
Crafts								1.00	.93	.95	.79	.78
Operatives									1.00	1.00*	.76	1.00*
Semi-skilled										1.00	.91	.90
Trans. equip. oper.											1.00	.91
Laborers												1.00

*The correction for sampling error in industry-occupation cell means causes some correlations to be greater than 1.00.

The regression from which the fixed effects were calculated contained education (years of schooling) and its square; experience (age-education-5) and its square; 50 state dummy variables; dummy variables for marital status, race, sex, part-time work, and whether or not an individual lives in an SMSA; and interaction terms for both experience and experience squared with all the other variables except the state dummies and education squared. The correlation coefficients are corrected for within cell sampling error. A detailed discussion of the correction procedure is available in Dickens and Katz (1986).

The estimated fixed effects were then grouped by occupation to analyze the relationships among industry impacts on wages in different occupations.

Correlations of these estimated industry wage differentials across occupations are presented in table 3. The correlations are corrected for sampling error.[26] Table 3 indicates that even after controlling for a wide variety of individual and geographic variables there are quite large correlations (most in the range .7 to 1) between average wages for workers in any two occupations within an industry. If one occupational group in an industry is highly paid, all categories of workers tend to be highly paid. This finding is hard to reconcile with views that the industry wage differences reflect unobserved ability or compensating differentials since skill requirements and working conditions are unlikely to be common to all occupations in an industry.

The results seem to support the union threat model since product market power or profitability is likely to raise worker bargaining power across occupations in an industry. Furthermore, the occupational structure of industry wage effects is quite similar for union and nonunion workers (Dickens and Katz 1986). The union threat model does not explain why groups of workers that do not pose a threat of collective action (e.g., managers) also share in the industry pay premiums.

The high correlation in industry differentials among different occupations is consistent with sociological models in which wage norms are linked in a firm or industry. The great efforts taken by firms to maintain their internal wage structure is quite consistent with these findings.

4.1.4. *Industry Characteristics and Industry Wage Patterns* An understanding of the empirical relevance of alternative theories of wage determination requires knowledge of the industry characteristics associated with high wages and low wages after controlling for worker characteristics. A considerable amount of empirical research has focused on the relationship between wages and industry structure. These studies (partially surveyed by Long and Link (1983) and Kwoka (1983)) have focused on the

26. Dickens and Katz (1986) provide further details on the estimation technique and describe the procedure for correcting the correlation coefficients for within cell sampling error. Since the industry differentials are estimated and since some of the industry-occupation cells are small, sampling error can lead to an upward bias in the standard deviations of industry wage effects within an occupation and a downward bias in the correlations of industry premiums across occupations. The raw correlations (uncorrected for sampling error) of the industry wage differentials across occupations lead to similar results, all quite positive. Dickens and Katz get results almost identical to those reported in table 3 when a full correction procedure for both within cell sampling error and the potential correlation in the estimation errors in the fixed effects across cells is utilized under the assumption of homoscedastic errors in the fixed effects regression.

influences of product market power, firm (or plant) size, and extent of unionization on wages. The industry characteristics that affect wage levels and the extent to which these variables matter appear to be quite sensitive to the specification (e.g., other control variables included) and to the particular sample analyzed (e.g., time period and use of only manufacturing vs. wider variety of industries). This suggests the effects are not uniform across industries and that multicollinearity is important for many industry variables. Nevertheless, some patterns emerge from these studies.

Industry wage levels are strongly positively correlated with industry concentration when no labor quality variables are utilized (L. Weiss 1966). The relationship is far more ambiguous when detailed labor quality controls are included. Pugel (1980) and Hodson and England (1985) find strong positive effects of industry profitability measures on average industry wages even with controls for average worker characteristics, extent of unionization, and other industry variables, including the rate of employment growth. Dickens and Katz (1986) find that profits as a percentage of sales are strongly positively related to industry wage premiums for nonunion workers. Kwoka (1983), Long and Link (1983), and Mellow (1982) find a positive and significant effect of industry concentration on wages, using individual-level data on earnings and worker characteristics combined with other industry level variables. This contrasts with L. Weiss's (1966) finding that concentration does not matter once individual worker controls are taken into account. Overall, industry wage differences appear to be related to product market power (ability to pay) although measurement problems in variables such as concentration and accounting profits mean these conclusions should be viewed as somewhat tentative.

The proportion of workers in an industry in plants of large or average size have typically been found to be positively related to industry wage levels even in the presence of detailed control variables (Kwoka 1983, Long and Link 1983, Pugel 1980, and others). Although establishment size and firm size appear to have quite important effects on wages within industries, they cannot explain much of interindustry wage differentials. The May 1979 CPS contains a special survey including questions on establishment and firm size. Krueger and Summers (1986) find in analyzing this data that the inclusion of plant size and firm size controls barely affects the estimates of industry wage differentials. They find the employment weighted standard deviation of two-digit industry log wage differentials falls only from 0.104 to 0.99 when plant and firm size controls are added to a log earnings equation with controls for occupation, region, union status, and individual characteristics. The raw correlation

of the estimated industry differentials with and without employer size controls is .96. In regressions (not reported here but available upon request) using the 1979 CPS, I find that estimated industry differentials are only slightly affected by the inclusion of plant size and firm size dummies when nonunion workers are analyzed in isolation. These results correspond to the conclusion of Brown and Medoff (1985) that most of the employer size effect on wages occurs within detailed industries.

The percentage of workers covered by collective bargaining in an industry has a strong positive effect on average industry wages.[27] Dickens and Katz (1986) find that the extent of industry unionization has a strong positive effect on both union and nonunion wages. They also find that regional union density has a strong positive effect on nonunion wages and a much weaker impact on union wages. Dickens (1986) argues that this is the pattern of union density impacts that arises from an important role of a union threat in wage determination.

The impacts of industry variables on wages remain a bit of a puzzle. The findings of most studies are fairly consistent with some role for union threat effects since product market power and extent of unionization seem to explain a fair portion of interindustry wage differentials for nonunion workers. Sociological models of the Akerlof (1984) variety also seem to have some support. The findings of Hodson and England (1986) and Lawrence and Lawrence (1985) that capital intensity (capital to labor ratio) has a positive effect on industry wages provides some support for the shirking model since the cost of worker malfeasance is likely to be greater in capital-intensive industries. Capital–skilled labor complementarity suggests that the capital-labor ratio finding may simply proxy for unmeasured labor quality. The strong linkages of wages to product market variables even after controlling for a large number of individual and locational variables appear difficult to reconcile with a strict unobserved ability interpretation of industry wage differentials.

4.1.5. Direct Evidence on the Benefits to Firms of High Wages Efficiency wage models postulate that firms pay wages above the market clearing level because there are cost-reducing or productivity-enhancing reasons to do so. Some limited empirical evidence exists on the benefits to firms of higher wages. As noted previously, wage premiums are associated with lower quit rates. Thus, high wages help to economize on turnover costs. The direct cost savings from lower turnover do not appear to be large enough to justify the magnitude of observed wage differentials. For example, Freeman and Medoff (1984, p. 109) estimate that the cost

27. Lewis (1983) provides a comprehensive survey of estimates of the extent of unionization on wages in industry-level studies.

savings associated with lower quit rates from the presence of a union is 1 to 2 percent of labor costs. They also find that the impact of unionism on quit rates to be equivalent to the impact of a 40 percent wage differential. The indirect gains of enhanced teamwork from continuity in work relationships may be the more important element of the benefits of lower turnover.

Hammermesh (1977) finds that the deviation of a worker's wage from the wage predicted by the worker's personal and job characteristics is positively correlated with various measures of job satisfaction. High wages appear to raise morale, as suggested by the sociological models and many personnel professionals. A further unanswered question is whether job satisfaction measures have much to do with productivity. Freeman and Medoff (1984) conclude from surveying a larger number of studies on worker attitudes and unionization drives that worker dissatisfaction is strongly correlated with increased desire for unionization as expressed by greater union organizing activity and a higher likelihood of votes in favor of unionization in representation elections. High wages, by raising worker satisfaction, reduce the likelihood of union organization as predicted by the union threat model.

Krueger and Summers (1986) provide some further evidence linking wage premiums to worker behavior consistent with some of the supposed benefits to a firm of efficiency wage payments. Industry wage premiums are found to be negatively related to absenteeism (due to weather) and positively related to employee self-evaluations of work effort. The positive correlation of wage differentials and employee views of work effort may simply indicate that high wages are acting as a compensating differential for greater effort required on the job or a fast pace of work. Allen (1984) finds consistently in an analysis of several data sets that positive wage differentials are associated with reduced absenteeism. Although absenteeism is something that can easily be observed by a firm, the reasons for absenteeism are not easily monitored. High wages combined with the threat of job loss for too much absenteeism might be an effective personnel policy.

Bulow and Summers (1986) discuss the introduction of the five-dollar-a-day pay system at Ford in 1914. They note that historical observers found that the higher wages led to large increases in productivity and reductions in absenteeism and turnover. This case provides some support for the implications of the shirking and turnover model.

Industry wage differences not captured by observed worker characteristics, working conditions variables, and locational variables are large and persistent. These differentials are not well explained by compensating differentials or transitory rents arising from shifts in labor demand

across sectors. More work using longitudinal data is required to determine the extent to which unmeasured ability may account for the estimated differentials. The union threat model appears quite consistent with industry differentials for nonunion production workers. The long-term stability of industry wage differences through periods with substantial differences in the extent of union organization indicate only a more limited role for union threats and direct union impacts. Large wage differentials across industries for occupations with little threat of collective action suggests that other factors must also be important. Industry differentials are strongly correlated across occupational groups. Sociological models in which industry wage contours or wage norms gain normative significance appear consistent with these similarities in the industry wage patterns across occupations. Economic efficiency wage rationales arising from monitoring, selection, or turnover problems can provide reasons why certain jobs in an industry require wage premiums. The concerns of firms with the perceived fairness of their internal wage structure may mean that these differentials come to permeate the entire wage structure in the industry. A combination of the economic efficiency wage models with the sociological (normative) efficiency wage models provides a fairly consistent, though far from elegant, account of the observed pattern of wage differentials. Much more empirical work is needed to uncover direct evidence of the gains to firms of high-wage policies. Better measures of the factors that economic efficiency wage models indicate should be important for wage differences (e.g., monitoring costs) are required to better determine the relevance of these models for understanding the apparent large impact of industry and firm affiliation on earnings.

4.2. LABOR MARKET DISCRIMINATION[28]

Efficiency wage models provide several reasons for persistent discrimination by race and sex in a competitive labor market. Group differences unrelated to productivity can potentially generate wage differences and occupational segregation. If two identifiable labor market groups differ in their rates of turnover or labor force withdrawal, the group with the higher turnover propensity will have a shorter horizon on a job and is likely to require greater inducement not to shirk. Bulow and Summers (1986) show that in the dual labor market version of the shirking model the higher turnover group must be overrepresented in the secondary sector since if the wage is the same for both groups in the primary sector,

28. This section is intended as a brief discussion of some of the implications of efficiency wage models for the analysis of labor market discrimination. I make no attempt to survey the vast empirical and theoretical literature on discrimination in the labor market.

the shorter horizon group is more likely to shirk unless the chances of being able to get a primary job in the future are lower than for the long horizon group. Since Poterba and Summers (1984) estimate a much higher rate of labor force withdrawal for women aged 25 to 59 than for men in the same age group, this yields a prediction of occupational and industrial segregation by sex with women tending to be found in lower-paying jobs that are easier to monitor. Bulow and Summers (1986) demonstrate that if labor market discrimination arises from a result of differences in separation probabilities by groups, then antidiscrimination policies such as affirmative action can increase welfare under a utilitarian welfare criterion.

Johnson and Solon (1984) estimate that the earnings of both males and females are negatively related to the proportion of females in the occupation even when a wide variety of individual control variables are included. Additionally, Johnson and Solon show that much of the difference in male and female earnings after controlling for differences in individual characteristics is related to differences in the industrial distribution of employment by sex. If industry wage differentials arise from efficiency wage considerations, then this is strong evidence that these factors play a major role in differences in earnings by sex.

Goldin (1986) presents evidence on widespread sex segregation across jobs within manufacturing requiring similar training and ability in 1890. She also finds that 50 percent of female operatives were paid piece rates as opposed to 13 percent of male operatives. Monitoring costs are found to be cheaper for piece rates than for time rates in female-dominated industries; the opposite is true for male-dominated industries. Goldin argues that deferred payment systems, such as discussed in Lazear (1979, 1981), conserved on monitoring costs and were feasible for males but not for females because of their shorter work horizons. Females were confined to jobs utilizing more expensive piece rates and thereby received lower wages than males. The same argument would apply to the use of efficiency wages rather than deferred payments for male-dominated jobs if full bonding were not feasible. Goldin also presents evidence showing that the feminization of the clerical occupations occurred only with standardization and division of tasks that made monitoring easier.

Occupational and industrial segregation arises in the shirking model because of the inability of high-turnover groups to post performance bonds. This suggests that groups facing capital market imperfections are more likely to be affected. Furthermore, even if groups do not differ in turnover propensities, if disadvantaged groups are liquidity constrained and unable to post bonds (accept deferred payment schemes), they will be less able to get primary sector jobs. Dickens and Lang (1985a, 1985b)

find that nonwhites are overrepresented relative to their observed characteristics in secondary sector jobs.

4.3. CYCLICAL BEHAVIOR OF LABOR MARKETS

The behavior of labor markets in which efficiency wage payments are required in some sectors in response to demand or productivity shocks corresponds well with stylized facts concerning the cyclical behavior of actual labor markets.[29] Since in a downturn the pool of the unemployed increases and the duration of unemployment goes up, the cost of quitting increases in an efficiency wage model. This implies procyclical quit behavior. Additionally, dual market versions yield predictions of cyclical upgrading as the primary sector expands in response to a positive shock (Bulow and Summers 1986, Jones 1985). The quit rate increases in upturns as workers quit secondary jobs to take available primary sector jobs. In response to negative shocks, workers are displaced from primary sector jobs.

This cyclical upgrading and downgrading corresponds with empirical findings that most of the employment growth in an upturn is associated with jobs in high-wage sectors such as durable goods manufacturing, construction, transportation, and public utilities (Okun 1973, Abraham and Katz 1986). Furthermore, Bils (1985) finds in an analysis of panel data that the real wages of workers who switch employers in a given year are strongly procyclical. He finds that the real wages of workers remaining on the same job are only slightly procyclical. Vroman (1978) also finds that workers switching jobs over the cycle have lower wages on average than job stayers and far more procyclical real wages. Vroman's results indicate that industry switchers have the most procyclical wages of any group.

This process of worker upgrading also appears to be important in understanding movements in productivity over the business cycle. Calculations based on figures presented in Okun (1973, p. 215) suggest that about 40 to 50 percent of the increase in labor productivity associated with an expansion in business activity is attributable to shifts in employment from low- to high-productivity sectors. Bernanke and Powell (1984) find that productivity is most procyclical in durable goods manufacturing. Efficiency wage models in which worker effort is variable provide a plausible partial explanation. The simple shirking model predicts countercyclical productivity within a sector in response to aggregate-

29. Okun (1973) discusses in detail empirical regularities observed in the cyclical behavior of labor markets. Bernanke and Powell (1984) empirically analyze the differences in similarities in the cyclical behavior of industrial labor markets in the prewar and postwar United States.

demand-driven business cycles since the greater cost of job loss in recessions spurs greater work effort. Yet, efficiency wage considerations create incentives for long-term relationships in the labor market. In this case, effort should vary with the amount of work needed to be done. Labor hoarding and cyclical productivity shifts appear most important in sectors where efficiency wage considerations are likely to be most important (capital-intensive industries).

Most of the cyclical variation in labor input takes place in fluctuations in employment rather than in hours worked per employee. Bulow and Summers (1986) argue that in the shirking model the value of a job is greater to a full-time worker than to a part-time worker. Thus the use of work sharing in response to a downturn reduces the value of jobs to workers and means higher wages are required to prevent shirking. If layoffs are utilized, the remaining workers have stronger work incentives. This makes sense for permanent declines in demand, but may not be reasonable for temporary downturns. Firms should be interested in increasing long-term horizons by maintaining long-term relations with workers. Workers' shirking decisions should be based on the long-term value of the job. The adverse selection model of A. Weiss (1980) provides an alternative reason for the use of layoffs rather than wage cuts or work sharing in response to declines in demand. If workers differ in unobservables and more productive workers have better outside opportunities, wage cuts and work sharing reducing the value of the job mean that the better workers are the ones most likely to leave for alternative jobs or self-employment.

5. *Cyclical Fluctuations and Efficiency Wages*

Efficiency wage models provide several mechanisms through which cyclical fluctuations in output can be generated by aggregate-demand shocks. In the first place, a basic property of these models is that wages are set by firms to maximize profits as the interior solution to a maximization problem.[30] In this case, the failure of firms to adjust wages to small shocks leads to only second-order losses. This differs sharply from a competitive labor market model where firms face large losses from failing to pay the competitive wage. Akerlof and Yellen (1985b) demonstrate that if firms are efficiency-wage setters in the labor market and monopolistic competitors in the product market, then inertial wage and price behavior in response to small nominal shocks leads to only second-order

30. This follows directly from the structure of the model presented in section 2. Stiglitz (1984) and Akerlof and Yellen (1985a, 1985b) develop this point in detail.

losses to the firms that follow such behavior even though this behavior generates a macroeconomic response with first-order welfare consequences.[31] Imperfect competition in the product market combined with efficiency wages in the labor market can potentially yield a model of cyclical fluctuations in response to aggregate-demand disturbances.

Strong incentives for the creation of long-term firm-worker relationships arise from efficiency wage considerations. The emergence of long-term employment relationships means that wage payments may reflect installment payments on long-term obligations (Hall 1980). In this case, employment decisions are not completely guided by current wages. This may reduce the importance of sticky wages from near-rationality and menu costs as a potential rationale for real impacts of nominal shocks.

The importance of relative wage concerns in efficiency wage settings provides a further rationale for inertial wage policies in response to real and nominal shocks. This point has been illustrated for the turnover model by Stiglitz (1984, 1985). A drop in the money supply requiring a reduction in nominal wages to maintain the existing level of unemployment may lead to unchanged nominal wages with decentralized wage setting. Any individual firm reducing its wage will tend to experience a higher quit rate and lower profits. Frictions in wage setting, such as staggered contracts, can exacerbate the difficulties in adjusting to nominal shocks when relative wages matter to firms and workers (Taylor 1982). If wages at all firms could be adjusted to shocks in a coordinated manner, these difficulties would not arise.

Monetary policy can also affect real output and the unemployment rate in the shirking model if it can affect real interest rates (Bulow and Summers 1986). An increase in the real interest rate reduces the discounted value of keeping a primary sector job and thereby increases the incentive of workers to shirk. This can reduce employment by requiring higher wages to prevent shirking. The quantitative importance of this mechanism for monetary policy is likely to be quite circumscribed.

6. *Conclusions*

Efficiency wage theories suggest that firms may find it profitable to pay workers' wages above the market clearing level since such wage premiums can help reduce turnover, prevent worker malfeasance and collective action, attract higher-quality employees, and facilitate the elicitation of

31. Adjustment costs may be greater for output and employment than for prices and wages. The menu costs or "near-rationality" argument appears equally consistent with inertial output and employment policies with fluctuating wages and prices in response to small shocks.

effort by creating feelings of equitable treatment among employees. Simple versions of efficiency wage models can explain involuntary unemployment, segmented labor markets with queues for primary sector jobs, and large differences among firms in the wages paid for what appear to be similar workers. Suitably modified, these models can parsimoniously explain many of the major stylized facts concerning the behavior of labor markets over the business cycle.

The primary criticism of efficiency wage models is that bonding mechanisms can solve effort elicitation, turnover, and adverse selection problems in an efficient manner. Additionally, the primary rationales for the use of efficiency wages are all arguments for the emergence of long-term contracts and long-term employer-employee attachments. Such long-term relationships appear quite important in the primary sector of the labor market where efficiency wage considerations are typically viewed as most important. These long-term attachments help facilitate bonding through the use of deferred payment mechanisms. It is an open empirical question whether seniority wage systems and pensions provide full bonding or whether they are only partial solutions leaving room for efficiency wages. Future theoretical work is required to analyze efficiency wage problems in an explicitly contractual setting.[32] The reasons why the contract market fails to clear need to be more fully explicated. The presence of more than one efficiency wage consideration may mean that bonding mechanisms solving one efficiency wage problem exacerbate others. For example, Shapiro and Stiglitz (1985) argue that employment fees to clear the market in the presence of the shirking problem may exacerbate adverse selection problems. Complicated contracts required to perform implicit bonding functions may create misunderstandings and lead to feelings of inequity that harm morale and productivity.

Evidence on industry wage differences indicates that large differentials remain that are quite difficult to explain in terms of differences in labor quality or differences in important nonpecuniary aspects of work requiring compensating differentials. The persistence of industry wage premiums for long time periods implies that they are not just transitory differentials arising to facilitate the sectoral reallocation of labor in a dynamic market economy. Large, persistent wage differentials for similar workers and types of jobs provide strong evidence in favor of the importance of some type of efficiency wage behavior by many firms. The complex pattern of differentials is difficult to reconcile with individual variants of the efficiency wage argument. Further empirical research is required to iso-

32. Mookherjee (1984b) and Bester (1985) provide some interesting initial attempts to link efficiency wage and implicit contract theories. Stiglitz (1984) discusses future directions for research in this area.

late the primary benefits to firms of high wages, the determinants of the uses of alternative compensation systems (piece rate vs. time rate, etc.), and the industry characteristics associated with large wage premiums.

I am indebted to William Dickens for numerous discussions and for comments on previous drafts of this paper; a significant part of this paper grew out of our joint work. I thank George Akerlof, Joe Altonji, Charles O'Reilly, Kevin Lang, Lawrence Summers, and Janet Yellen for helpful discussions. I am grateful to Phil Bokovoy and Elizabeth Bishop for expert research assistance. The Institute of Industrial Relations at U.C. Berkeley provided research support. All remaining errors are my own.

REFERENCES

Abraham, K. G., and H. S. Farber. 1986. Job duration, seniority, and earnings. MIT Department of Economics Working Paper no. 407 (January).

Abraham, K. G., and L. F. Katz. 1986. Cyclical unemployment: Sectoral shifts or aggregate disturbances? *Journal of Political Economy*. 94 (June): 507–22.

Akerlof, G. A. 1982. Labor contracts as partial gift exchange. *Quarterly Journal of Economics* 87 (November): 543–69.

———. 1984. Gift exchange and efficiency wages: Four views. *American Economic Review* 74 (May): 79–83.

Akerlof, G. A., and L. F. Katz. 1986. Do Deferred Wages Dominate Involuntary Unemployment As A Worker Discipline Device? University of California at Berkeley. Mimeo.

Akerlof, G. A., and J. Yellen. 1984. Efficiency wage models of the labor market: Introduction. University of California at Berkeley. Mimeo.

———. 1985a. A near rational model of the business cycle, with wage and price inertia. *Quarterly Journal of Economics* 100 (August): 823–38.

———. 1985b. Can small deviations from rationality make a significant difference to economic equilibria? *American Economic Review* 75 (September): 708–20.

Allen, S. 1984. Trade unions, absenteeism and exit-voice. *Industrial and Labor Relations Review* 37 (April): 331–45.

Altonji, J., and J. Ham. 1985. The collective impact of sectoral shocks on aggregate employment fluctuations. Columbia University. Mimeo.

Altonji, J., and R. Shakotko. 1985. Do wages rise with job seniority? NBER Working Paper 1616 (May).

Antel, J. J. 1985. Costly labor contract negotiation and the labor mobility of young men. *American Economic Review* 75: 976–91.

Becker, G. S., and G. J. Stigler. 1974. Law enforcement, malfeasance, and the compensation of enforcers. *Journal of Legal Studies* 3 (January): 1–18.

Bell, L., and R. B. Freeman. 1985. Does a flexible industry wage structure increase employment? The U.S. experience. NBER Working Paper 1604 (April).

Bernanke, B., and J. Powell. 1984. The cyclical behavior of industrial labor markets: A comparison of pre-war and post-war eras. Stanford University. Mimeo.

Bester, H. 1985. Long-term wage contracts and dual labor markets. University of Bonn (February). Mimeo.

Bhattacharya, S. 1983. Tournaments and incentives: Heterogeneity and essentiality. University of California at Berkeley (March). Mimeo.

Bils, M. J. 1985. Real wages over the business cycle: Evidence from panel data. *Journal of Political Economy* 93 (August): 666–89.

Blanchard, O. J., and N. Kiyotaki. 1985. Monopolistic competition, aggregate demand externalities, and the real effects of nominal money. NBER Working Paper 1770 (December).
Bloch, F., and M. Kuskin. 1978. Wage determination in the union and nonunion sectors. *Industrial and Labor Relations Review* 31 (January): 183–92.
Bowles, S. 1985. The production process in a competitive economy: Walrasian, Neo-Hobbesian and Marxian models. *American Economic Review* 75 (March): 16–36.
Brown, C. 1985. Standard-rate wage setting, labor quality, and unions. NBER Working Paper 1717 (October).
Brown, C., and J. Medoff. 1985. The employer size wage effect. Harvard University. Mimeo.
Bulow, J. I., and Summers, L. H. 1986. A theory of dual labor markets with application to industrial policy, discrimination and Keynesian unemployment. *Journal of Labor Economics*. Forthcoming.
Calvo, G. 1979. Quasi-Walrasian theories of unemployment. *American Economic Review* 69 (May): 102–7.
———. 1985. The inefficiency of unemployment: The supervision perspective. *Quarterly Journal of Economics* 100 (May): 373–87.
Calvo, G., and S. Wellisz. 1978. Supervision, loss of control, and the optimum size of the firm. *Journal of Political Economy* 86 (October): 943–52.
Carmichael, L. 1983. Firm-specific human capital and promotion ladders. *Bell Journal of Economics* 14 (Spring): 251–58.
———. 1985. Can unemployment be involuntary?: Comment. *American Economic Review* 75 (December): 1213–14.
Cullen, D. 1956. The interindustry wage structure, 1899–1950. *American Economic Review* 46 (June): 353–69.
Dickens, W. T. 1986. Wages, employment and the threat of collective action by workers. University of California at Berkeley. Mimeo.
Dickens, W. T., and L. F. Katz. 1986. Industry wage patterns and theories of wage determination. University of California at Berkeley. Mimeo.
Dickens, W. T., L. F. Katz, and K. Lang. 1986. Are efficiency wages efficient? NBER Working Paper 1935.
Dickens, W. T., and K. Lang. 1985a. A test of dual labor market theory. *American Economic Review* 75 (September): 792–805.
———. 1985b. Testing dual labor market theory: A reconsideration of the evidence. NBER Working Paper 1670.
Doeringer, P. B., and M. J. Piore. 1971. *Internal labor markets and manpower analysis*. Lexington, Ma.: D. C. Heath.
Dunlop, J. 1985. Industrial relations and economics: The common frontier of wage determination. *IRRA Proceedings 1984*.
Eaton, C., and W. D. White. 1982. Agent compensation and the limits of bonding. *Economic Inquiry* 20 (July): 330–43.
———. 1983. The economy of high wages: An agency problem. *Economica* 50 (April): 175–81.
Foster, J., and H. Wan. 1984. Involuntary unemployment as a principal-agent equilibrium. *American Economic Review* 74 (June): 476–84.
Foulkes, F. 1980. *Personnel policies in large nonunion companies*. Englewood Cliffs, N.J.: Prentice-Hall.
Freeman, R. B. 1980. The exit-voice tradeoff in the labor market, unionism, job

tenure, quits, and separations. *Quarterly Journal of Economics* 94 (June): 643–73.
———. 1981. The effect of trade unionism on fringe benefits. *Industrial and Labor Relations Review* 34 (July): 489–509.
———. 1983. Unionism, price-cost margins and the return to capital. Harvard University (January). Mimeo.
Freeman, R. B., and J. Medoff. 1981. The impact of collective bargaining: Illusion or reality? in *U.S. industrial relations 1950–1980: A critical assessment*, ed. J. Stieber, R. B. McKersie, and D. Q. Mills. Madison: IRRA.
———. 1984. *What do unions do?* New York: Basic Books.
Gintis, H., and T. Ishikawa. 1983. Wages, work discipline and macroeconomic equilibrium. University of Massachusetts at Amherst. Mimeo.
Goldin, C. 1986. Monitoring costs and occupational segregation by sex: A historical analysis. *Journal of Labor Economics* 4 (January): 1–27.
Hall, R. E. 1975. The rigidity of wages and the persistence of unemployment. *Brookings Papers on Economic Activity* 2: 301–35.
———. 1980. Employment fluctuations and wage rigidity. *Brookings Papers on Economic Activity* 1: 91–132.
Hammermesh, D. 1977. Economic aspects of job satisfaction. In *Essays in Labor Market Analysis*, ed. O. Ashenfelter and W. Oates. New York: Halstead.
Hodson, R., and P. England. 1986. Industrial structure and sex differences in earnings. *Industrial Relations* 25 (Winter): 16–32.
Ippolito, R. A. 1985. The labor contract and true economic pension liabilities. *American Economic Review* 75 (December): 1031–1043.
Jacoby, S. 1983. Industrial mobility in historical perspective. *Industrial Relations* 22 (Spring): 261–82.
Johnson, G., and G. Solon. 1984. Pay differences between women's and men's jobs: The empirical foundations of comparable worth legislation. NBER Working Paper 1472 (September).
Jones, S. R. G. 1985. Dual labor markets, productivity and unemployment. University of British Columbia (April). Mimeo.
Katz, L. F. 1986. Inter-industry wage structure in historical perspective. University of California at Berkeley (January). Mimeo.
Kreps, D. 1984. Corporate culture and economic theory. Stanford Graduate School of Business (May). Mimeo.
Krueger, A., and L. H. Summers. 1986. Efficiency wages and the inter-industry wage structure. Harvard University (February). Mimeo.
Kwoka, J. 1983. Monopoly, plant, and union effects on worker wages. *Industrial and Labor Relations Review* 36 (January): 251–57.
Lawrence, C., and R. Lawrence. 1985. Relative wages in U.S. manufacturing: An endgame interpretation. *Brookings Papers on Economic Activity* 1: 47–106.
Lazear, E. 1979. Why is there mandatory retirement? *Journal of Political Economy* 87 (December): 261–84.
———. 1981. Agency, earnings profiles, productivity, and hours restrictions. *American Economic Review* 71 (September): 606–20.
———. 1983. Salaries and piece rates. NORC Working Paper No. 83–15, June.
Leibenstein, H. 1957. The theory of underdevelopment in densely populated backward areas. In *Economic Backwardness and Economic Growth*, ed. H. Leibenstein. New York: Wiley.
Lewis, H. G. 1982. Union relative wage effects: A survey. Duke University (April). Mimeo.

———. 1983. Union relative wage effects: A survey of macro estimates. *Journal of Labor Economics* 1 (January): 1–27.
Lilien, D. 1982. Sectoral shifts and cyclical unemployment. *Journal of Political Economy* 90 (August): 777–93.
Lindbeck, A., and D. J. Snower. 1984. Involuntary unemployment as an insider-outsider dilemma. Seminar Paper No. 282, Institute for International Economic Studies, Stockholm, Sweden.
Long, J. E., and Link, A. N. 1983. The impact of market structure on wages, fringe benefits, and turnover. *Industrial and Labor Relations Review* 36 (January): 239–50.
Malcolmson, J. 1984. Work incentives, hierarchy, and internal labor markets. *Journal of Political Economy* 92 (June): 486–507.
Mankiw, N. G. 1985. Small menu costs and large business cycles: A macroeconomic model of monopoly. *Quarterly Journal of Economics* 100 (May): 529–37.
Medoff, J., and K. G. Abraham. 1980. Experience, performance and earnings. *Quarterly Journal of Economics* 95 (December): 703–36.
Mellow, W. 1982. Employer size and wages. *Review of Economics and Statistics* 64 (August): 495–501.
Milkovich, G., and J. Newman. 1984. *Compensation*. Plano, Tx.: Business Publications Inc.
Montgomery, E., and D. Stockton. 1985. Evidence on causes of the rising dispersion of relative wages. (April). Carnegie-Mellon University. Mimeo.
Mookherjee, D. 1984a. Involuntary unemployment and worker self-selection. Research Paper 741, Graduate School of Business, Stanford University (April).
———. 1984b. Involuntary unemployment and worker moral hazard. Research Paper 745, Graduate School of Business, Stanford University (June).
Murphy, K. M., and R. Topel. 1986. Unemployment, risk, and earnings: Testing for equalizing wage differences in the labor market. UCLA (March). Mimeo.
Oi, W. 1983. Heterogeneous firms and the organization of production. *Economic Inquiry* 21 (April): 147–71.
Okun, A. 1973. Upward mobility in a high-pressure economy. *Brookings Papers on Economic Activity* 1: 207–52.
Pencavel, J. 1970. *An analysis of the quit rate in American manufacturing industry*. Princeton, N.J.: Industrial Relations Section, Princeton.
———. 1977a. Work effort, on-the-job screening and alternative methods of remuneration. In *Research in Labor Economics*, ed. R. Ehrenberg. Greenwich, Ct.: JAI Press.
———. 1977b. Industrial morale. In *Essays in Labor Market Analysis*, ed. O. Ashenfelter and W. Oates. New York: Halstead.
Poterba, J., and L. H. Summers. 1984. Adjusting the gross changes data: Implications for labor market dynamics. NBER Working Paper 1436 (August).
Pugel, T. 1980. Profitability, concentration and the interindustry variation in wages. *Review of Economics and Statistics* 62 (May): 248–53.
Reder, M. 1962. Wage differentials: Theory and measurement. In *Aspects of Labor Economics*. Princeton, N.J.: Universities-NBER, Princeton U. Press.
Reynolds, L. 1978. *Labor economics and labor relations*, 7th ed. Englewood Cliffs, N.J.: Prentice-Hall.
Rosen, S. 1985. The theory of equalizing differences. NORC Discussion Paper No. 85–3 (January).

Ruback, R., and M. Zimmerman. 1984. Unionization and profitability: Evidence from the capital market. *Journal of Political Economy* 92 (December): 1134–57.
Salop, S. 1979. A model of the natural rate of unemployment. *American Economic Review* 69 (March): 117–25.
Shapiro, C., and J. E. Stiglitz. 1984. Equilibrium unemployment as a worker discipline device. *American Economic Review* 74 (June): 433–44.
———. 1985. Can unemployment be involuntary?: Reply. *American Economic Review* 75 (December): 1215–17.
Slichter, S. 1950. Notes on the structure of wages. *Review of Economics and Statistics* 32 (February): 80–91.
Solow, R. 1979. Another possible source of wage stickiness. *Journal of Macroeconomics* 1: 79–82.
———. 1985. Insiders and outsiders in wage determination. *Scandinavian Journal of Economics* 87 (2): 411–28.
Stiglitz, J. E. 1974. Wage determination and unemployment in L.D.C.'s: The labor turnover model. *Quarterly Journal of Economics* 88 (May): 194–227.
———. 1976. Prices and queues as screening devices in competitive markets. IMSS Technical Report No. 212, Stanford University (August).
———. 1984. Theories of wage rigidity. NBER Working Paper 1442 (September).
———. 1985. Equilibrium wage distributions. *Economic Journal* 95 (September): 595–618.
Stoft, S. 1982. Cheat threat theory: An explanation of involuntary unemployment. Boston University (May). Mimeo.
Taylor, J. 1982. The role of expectations in the choice of monetary policy. NBER Working Paper 1044 (December).
Topel, R., and L. Weiss. 1985. Sectoral uncertainty and unemployment. UCLA Department of Economics Working Paper 384 (September).
Vroman, W. 1978. Cyclical earnings changes of low wage workers. In *Research in Labor Economics*, vol. 2, ed. R. Ehrenberg. Greenwich, Ct.: JAI Press.
Weiss, A. 1980. Job queues and layoffs in labor markets with flexible wages. *Journal of Political Economy* 88 (June): 526–38.
Weiss, L. 1966. Concentration and labor earnings. *American Economic Review* 56 (March): 96–117.
Wing, R. 1984. Achieving internal equity through job measurement. In *Handbook of Wage and Salary Administration*, ed. M. Rock. New York: McGraw-Hill.
Yellen, J. 1984. Efficiency wage models of unemployment. *American Economic Review Proceedings* 74 (May): 200–205.

Comment

JOSEPH G. ALTONJI
Columbia University and NBER

Larry Katz has made a valiant attempt to evaluate recent theoretical and empirical work on efficiency wage models—models in which wages have a direct influence on labor productivity and turnover costs. It

would take a monograph to provide a complete survey of the many variations on the efficiency wage theme. Furthermore, a large empirical and theoretical literature in labor economics on the structure of wages, turnover behavior, firm and worker search behavior, and the composition of unemployment is potentially relevant for an assessment of efficiency wage models.

My comments fall into four areas. First, a few remarks about efficiency wage theory. Second, I discuss some general issues concerning the empirical assessment of efficiency wage models. Third, I consider the evidence reviewed in the article. In closing, I briefly comment on the relevance of efficiency wage models for business cycle fluctuations.

1. *Comments on the Theoretical Discussion*

Katz provides a brief summary of the links from wages to labor productivity that have received the most attention in the theoretical literature. The first mechanism is the shirking model, in which high wages induce high effort levels. The second is the labor turnover model, in which high wages reduce quits and thus lower turnover costs. The third is the adverse selection model in which firms use high wages to sort out high-quality workers from lower-quality workers when quality is difficult to observe directly. The fourth are sociological models in which worker perceptions of the fairness of the wages offered by the firm affects their effort level and perhaps their turnover behavior. In addition, Katz devotes considerable attention to the union threat model. In this model profit-maximizing nonunion firms choose to pay wages above competitive levels to avoid unionization. The union threat effect model has received considerable attention in discussions of the link between unionism and the structure of wages, although the model is not usually included in lists of efficiency wage mechanisms.

The exposition of the theories is generally evenhanded and clear. A few points deserve special emphasis. The first is that the importance of the shirking and adverse selection models depends critically upon the extent to which workers acquire reputations. Most men spend more than thirty-five years as full-time workers, and many women also spend long periods of time in the labor force. Although it may be difficult for employers accurately to evaluate performance in a given period, the cumulative effect of such evaluations may have a key influence on a worker's prospects in and outside the firm. Evidence from panel data suggests that wages evolve over time for a given individual. (See, for example, MaCurdy (1982)). A large proportion of wage changes have per-

manent effects on future wages. In an expected value sense, the rewards to good performance may be substantial.

A note on the sociological theories. There is a long tradition in industrial relations and personnel literature stressing the importance of interpersonal comparisons and notions of fairness in the determination of wage rates. I have no doubt that these factors play a role. However, the sociological theories are vague about how such standards are determined or what the appropriate reference group is. Do workers compare themselves to other workers in the same industry, in the same firm, in the same occupation, or in the same position within a firm? Are these standards of comparison constant over time?

Evidence from panel data sets suggests that wages of individuals are subject to substantial variation from year to year, even after the substantial measurement error in such data is accounted for. It would be very interesting to perform similar calculations using data from workers in the same firm. This could be done with information from firm personnel files. If the results were to show little dispersion in wage growth for employees within a given firm, they would be consistent with the notion that interpersonal comparisons have an important influence on the wage structure, and that the wage structure within firms is fairly rigid. My guess is that there is in fact a substantial amount of dispersion in wage growth within a given firm over an interval of two or three years. The importance of interpersonal comparisons and notions of fairness may slow down relative wage adjustments. However, I doubt if these factors can explain differentials across firms, occupations, or industries that persist for many years.

I wish to highlight a point made in the article concerning the possibility that firms may use entrance fees, deferred wage payments, and fringe benefits tied to seniority as a way to induce workers not to shirk, quit, etc. I think that it would be possible to devise a market clearing optimal compensation package that deals with the shirking problem, or that generates optimal turnover decisions, given training costs. However, it may be very costly to provide such a package if several efficiency wage mechanisms operate simultaneously, and if wages are also influenced by the risk-sharing considerations emphasized in the implicit contracts literature. It may be even more difficult to do so in a stochastic environment in which wages must be renegotiated, for many of the same reasons discussed in the literature on the feasibility of efficient contracts (see Hall and Lazear (1984) and Hart (1983)). Consequently, I find the view that in many situations firms will resort to raising the wage profile above market clearing levels to be fairly persuasive on a priori grounds.

In discussing the role of bonding mechanisms in efficiency wages, a number of economists have argued that demographic and skill groups who suffer from liquidity constraints and thus cannot post performance bonds are most likely to be affected by unemployment. However, this argument ignores the fact that a spell of unemployment is likely to be most costly for those who are liquidity constrained. Consequently, it is not clear that limits on the feasibility of bonding arrangements arising from imperfect capital markets can provide an explanation for age and skill differentials in unemployment.

2. *General Issues in Testing Efficiency Wage Models*

The term "efficiency wage model" is given a very broad interpretation in Katz's article. The following simple framework may help to sharpen the distinction between these models and neoclassical alternatives and to show the main issues involved in testing the efficiency wage models. The null hypothesis against which efficiency wage models are evaluated is a Reder or Rosen type hedonic model. In this competitive model, wages and employment levels adjust over time to equate the demand and supply for labor and jobs of various types. More formally, let the demand and supply equations be

$$D_{ij} = D(w_{ij} \; ; \; \mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_{ij}, \mathbf{w}) \quad (1)$$

$$S_{ij} = S(w_{ij} \; ; \; \mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_{ij}, \mathbf{w}, \mathbf{U}), \quad (2)$$

where S_{ij} is the supply of workers of type i to jobs of type j, D_{ij} is the demand for workers of type i in jobs of type j, w_{ij} is the current wage for workers of type i in job j, $\mathbf{Z}_i$ is a vector of personal characteristics variables which affect the supply or the demand for workers in all types of jobs, $\mathbf{Z}_j$ is a vector of job characteristics which affect the supply or the demand for workers of all types for job j, $\mathbf{Z}_{ij}$ is a vector of interaction variables which reflect variation across jobs in the productivity of workers of type i or variation in their evaluation of the characteristics of the job, $\mathbf{w}$ is a vector of wages for all other worker/job combinations, and $\mathbf{U}$ is a vector of the balance between the supply of workers and the number of jobs for each of the various worker/job combinations. It should be noted that $\mathbf{Z}_j$ and $\mathbf{Z}_{ij}$ include variables describing the wages and fringe benefits which worker i may expect to receive in future periods. (The human capital model implies differences across firms in wage profiles which arise from differences in the levels of specific and general training that

are provided.) The form of the demand and supply functions depends on the distribution of $\mathbf{Z}_i$, $\mathbf{Z}_j$, and $\mathbf{Z}_{ij}$ over the economy. Conventional labor supply/search models imply that employment prospects affect the supply of labor to a particular type of job. For this reason, the vector $\mathbf{U}$ of excess demands for matches of various types is included in the equation.

The key assumption of the hedonic model is that the market clears in a stochastic sense for all types of workers and all types of jobs, with

$$D(w_{ij} \; ; \; \mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_{ij}, \mathbf{w}) = S(w_{ij} \; ; \; \mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_{ij}, \mathbf{w}, \mathbf{U}). \quad (3)$$

Thus, the excess demand vector $\mathbf{U}$ is 0 in equilibrium. The solution implies an equilibrium relationship between the wage and employment level for a given ij combination. These functions are

$$w_{ij} = w(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_{ij}) \quad (4)$$

$$E_{ij} = E(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_{ij}), \quad (5)$$

where E_{ij} is the employment of workers of type i in jobs of type j. Although the excess demand vector is 0, a stochastic version of the hedonic market story would add that product demand variations across firms, changes in labor supply preference, and other factors generate labor turnover and movements in and out of the labor market. The associated frictional unemployment rates may vary with the type of job and the individual.

Efficiency wage models replace the hedonic market equilibrium condition, equation (3), with a wage setting equation that is based upon a model of how wages affect labor productivity and nonwage labor costs.

$$w_{ij}^e = f(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_{ij} \; ; \; \mathbf{w}, \mathbf{U}) \quad (6)$$

In equation (6) w_{ij}^e is the optimal wage chosen by the firm. The equation makes explicit the fact that the efficiency wage may depend upon a vector $\mathbf{w}$ of wages in all other ij matches as well as the vector $\mathbf{U}$ summarizing the differences ($S_{i'j'} - D_{i'j'}$) between supply and demand for matches of various other $i'j'$ types. This dependence arises because efficiency wage theories emphasize that alternative wages and unemployment rates serve as a discipline on worker behavior.

Associated with equation (6) is a labor demand function, equation (7). This function also accounts for the fact that in the efficiency wage models alternative wages and the tightness of the labor market affect worker productivity and thus labor demand.

$$D_{ij} = D^e(w_{ij}\,;\, \mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_{ij}, \mathbf{w}, \mathbf{U}) \quad (7)$$

Equations (6) and (7) determine employment. Subtraction of equation (7) from equation (2) yields the excess demand U_{ij} for job matches of type ij.

$$U_{ij} = S_{ij} - E_{ij} \quad (8)$$

Simultaneous solution across ij combinations in conjunction with the supply equation (2) yields the wage vector **w** and the excess demand vector **U**. In equilibrium, there may be excess demand for some types of matches and excess supply for other types. The solution to the model implies an equilibrium relationship of w_{ij}, E_{ij}, and U_{ij} to characteristics of i, j, and the match ij:

$$w_{ij} = w^e(Z_i, Z_j, Z_{ij}) \quad (9)$$

$$E_{ij} = E^e(Z_i, Z_j, Z_{ij}) \quad (10)$$

$$U_{ij} = U^e(Z_i, Z_j, Z_{ij}). \quad (11)$$

How can one distinguish empirically between the hedonic model and the efficiency wage model? There are likely to be few exclusion restrictions from the wage and employment loci for the two models. One approach is to estimate the wage locus and consider whether the relationship is more consistent with the hedonic model or with the various efficiency wage models. Most of the empirical evidence discussed in the article involves this approach. This may be done in two levels. The first is to see whether wages vary across types of firms or across personal characteristics in ways that, given what we know about labor supply and labor demand, are hard to square with the hedonic model. If the results seem unfavorable to the hedonic model, then the second level of the analysis is to see if the results may be explained with other models, such as the various efficiency wage models. The problem, however, is that given available data it is very difficult to control for the many variables that may be correlated with supply and demand for particular job matches.

A second strategy is to examine the distribution of employment. This approach receives less attention in Katz's paper. To give an example, the discussion of dual labor markets in Bulow and Summers (1986) focuses on the distribution of various types of workers across job types. Again, there are often several alternative explanations for such distributions. For example, some economists argue that differences between men and

women in an industry and the occupational distribution of employment arise from sex differences in schooling choice, labor force attachment, and preferences for various job attributes. Others interpret the differences as evidence of discrimination in the labor market. (See Cain's (1984) survey for many references.)

The third strategy is to look for evidence of imbalance in the supply and demand for various worker/job combinations. In my view, a key implication of the efficiency wage models is that such imbalances will persist over time. One can only get rough indicators of such imbalances by examining industry/occupation unemployment rates, patterns of job mobility broken down by quit/layoff, and vacancy data. Nevertheless, I think a careful survey of the literature on the structure of unemployment and on mobility patterns would be helpful, particularly since "stylized facts" on this issue seem to have motivated much of the work on efficiency wages.

3. *Comments on the Empirical Work*

Most of the empirical evidence discussed in the article concerns the wage structure. First, Katz notes that area wage surveys find large differences across firms in wages for a given job classification within a geographical area.

Second, he cites evidence from several studies on industry wage differentials, including some of his own work. These differentials are substantial, stable over time, and remain after one controls for a set of observed individual and job characteristics. I am impressed by Krueger and Summers's (1986) and Vroman's (1978) finding that large industry effects are found even if one uses a first-difference estimator. Such a procedure is less sensitive to the criticism that the wages premiums reflect unmeasured differences in the quality of workers employed in the various industries.[1]

Many labor economists have argued that industries provide different levels of specific and general training, and that these underlie the indus-

1. If anything, measurement error in industry classifications and selection bias are likely to result in a downward bias in the first difference estimates. Krueger and Summers's first-difference estimates contain a rough correction for the effects of measurement error. Without this correction their first-difference estimates of industry wage differentials are considerably smaller than the estimates based on level equations. More research on this will be needed, especially in light of the recent study by Kevin Murphy and Robert Topel (1986). As Katz notes, Murphy and Topel have obtained estimates of industry differentials that are well below those of Krueger and Summers, despite the fact that Murphy and Topel use similar data and work with first differences. They use an instrumental variables procedure to deal with the measurement error problems.

try differentials. I found Katz's evidence suggesting that industry differentials have a fairly high correlation across experience levels to be an interesting challenge to this view. On the other hand, I am not sure if Katz's finding of .74 correlation among the industry differentials for workers with less than a year of tenure should be considered large or small. A much more detailed analysis of these issues seems warranted.

The article discusses the possibility that compensating differentials for other job attributes, such as work hours, work pace, safety, and unemployment risk can reconcile the industry differentials with a hedonic model of the labor market. I think this is very much an open question. Abowd and Ashenfelter, Topel, and others suggest that unemployment risk does play a role in industry wage differentials, as does safety. Furthermore, the fact that industry differentials survive an attempt to control for a set of job characteristics is not decisive evidence against the compensating differentials view, since micro data on job characteristics are not only incomplete but noisy. If the wage premiums do not reflect compensation for training opportunities and other job characteristics, we need to know why workers move from high-wage industries to low-wage industries. Some tabulations on the relationship between the industry wage differentials and interindustry job changes broken down by quit and layoff, with and without intervening spells of unemployment, would be very interesting. If the movements from high-wage to low-wage industries are dominated by layoffs, this would be further evidence that the industry differentials are not compensating. This type of evidence relates to the central issue of whether markets clear for various worker/job combinations.

The article also discusses recent work by the author (in collaboration with William Dickens) indicating that workers in all occupations within an industry are affected more or less equally by the industry wage differentials, although they also find that the dispersion of the industry effects is larger for managerial, professional, and supervisory workers than for other groups. As Katz points out, these findings do not support the hedonic model, since the job characteristics are presumably highly occupation-specific. But as he also recognizes, they are not very favorable to the different variants of the efficiency wage hypothesis either, with the possible exception of the sociologically based models.

There does not seem to be much evidence for the hypothesis that employers derive benefits from paying higher wages. The finding that quits are a negative function of the wage is predicted by virtually any model of the labor market. Unfortunately, there is little information on the costs to firms of turnover. Since absenteeism can easily be monitored and penalized, the fact that higher wages reduce absenteeism would seem to be

more consistent with the hedonic model in which workers trade off the right to be absent for higher wages. Indeed, Katz mentions a number of facts that do not sit well with the various efficiency wage models.

4. *Efficiency Wages and Business Cycle Behavior*

My final comments concern the efficiency wage models and business cycle behavior. Katz does not devote much space to these issues. He cites evidence from several studies that low-skill workers are more likely to be employed in high-wage sectors such as durable goods manufacturing, construction, transportation, and public utilities during booms than during recessions. This in itself is not surprising, since labor demand in these sectors is the most sensitive to the business cycle. However, competitive theories of the labor market do not provide an adequate explanation for why low-skill workers who switch industries during expansions experience wage gains that exceed the cyclical wage adjustment for workers who were already employed in the high-wage sectors.[2] More work on the effects of labor demand shifts on wage adjustments, promotion patterns within firms, and mobility patterns of workers of various skill levels would be useful.

Finally, the article only briefly discusses the macroeconomic implications of efficiency wage models. I think that efficiency wage theories will ultimately prove useful in explaining the wage and employment decisions of firms over the cycle. The papers of Phelps (1970) and Mortensen (1970), which stress labor turnover and recruiting costs and the short-run monopsony power of firms, are early prototypes of such models. I think that we must explicitly account for the wage decisions of firms if we are to understand wage and employment dynamics, rather than simply treat wages as determined in a passive way by the balance of supply and demand. However, a well-worked-out theory with supporting empirical evidence is still not available despite much interesting research.

REFERENCES

Altonji, J. G., J. Mincer, and R. Shakotko. 1984. *Employer/employee attachments*, Final report to the Office of the Assistant Secretary for Policy and Evaluation, U.S. Department of Labor, Washington, D.C.

Bulow, J. I., and L. H. Summers. 1986. A theory of dual labor markets with appli-

2. Although Katz cites two studies that find large cyclical wage changes for job changers, my own work with the Panel Study of Income Dynamics indicates that the response of wages to changes in aggregate unemployment, state employment growth rates, and county unemployment rates is about the same for workers who have been on the job for less than a year as for workers with higher levels of job seniority. See Altonji, Mincer, and Shakotko (1984, chapter 4).

cations to industrial policy, discrimination and Keynesian unemployment. *Journal of Labor Economics*. Forthcoming.
Cain, G. G. 1984. The economic analysis of labor market discrimination: A survey. Institute for Research on Poverty Special Report 37 (August) forthcoming in *Handbook of Labor Economics*, ed. O. Ashenfelter and R. Layard.
Hall, R. E., and E. Lazear. 1984. The excess sensitivity of layoffs and quits to demand. *Journal of Labor Economics* 2 (2): 243–58.
Hart, O. D. 1983. Optimal labour contracts under asymmetric information: An introduction. *Review of Economic Studies* 50 (1) (January): 3–35.
Krueger, A., and L. H. Summers. 1985. Efficiency wages and the inter-industry wage structure. Harvard University, Department of Economics (October). Mimeo.
MaCurdy, T. E. 1982. The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics* 18: 83–114.
Mortensen, D. T. 1970. A theory of wage and employment dynamics. In *Microeconomic foundations of employment and inflation theory*, ed. E. S. Phelps et al. New York: Norton.
Murphy, K. M., and R. H. Topel. 1986. Unemployment, risk, and earnings: Testing for equalizing wage differences in the labor market. (March). Mimeo.
Phelps, E. S. 1970. Money wage dynamics and labor market equilibrium. In *Microeconomic foundations of employment and inflation theory*, ed. E. S. Phelps et al. New York: Norton.
Vroman, W. 1978. Cyclical earnings changes of low wage workers. In *Research in Labor Economics*, vol. 2. ed. R. Ehrenberg. Greenwich, Ct.: JAI Press.

Comment

LAURENCE WEISS
Goldman Sachs & Co.

Any paper entitled "Efficiency Wages—A Partial Evaluation" is not likely to be too far wrong. Larry Katz's tone is somewhat agnostic and steadfastly fair and impartial. He plays his cards pretty close to his vest and I can't really tell if he's rooting for these theories or thinks they miss the mark.

The title of this paper could have been "Is Life Fair?" The theories examined here have the common theme that identical individuals wind up with very different levels of utility in economic equilibrium. These theories purport to "explain" why there are wide wage differentials across various industries not explained by observable worker characteristics. In a business cycle context, the theories seek to find an explanation for that incubus of economic thought termed "involuntary unemployment," which has been interpreted here to be a situation in which one jobless

person envies an employed person, even though they are considered by everyone else to be identical. After reading Katz's paper and the paper by Krueger and Summers referenced in it, I am skeptical about both the prevalence of these sorts of phenomena and the usefulness of these theories for explaining them.

I have asked some of my colleagues what they thought about the proposition that employees work harder when they're paid more, a key observation of the "shirking" model of efficiency wages. Two of them told me (independently) that they pay their baby-sitters more than is really necessary. Their reasoning is broadly consistent with the underlying logic of the model; higher wages would induce more responsible behavior. What features of baby-sitting make it particularly relevant for the shirking model? The answer is obvious: baby-sitting is a dead-end job without much hope for advancement. This I take to be an important (and unstated) assumption of the model; not only must there be a lack of punitive measures against employees for dereliction of duties, but there cannot be any rewards for superlative performance. The relevance of this sort of model would appear to be confined to the rather small minority of people with menial jobs. Casual empiricism suggests that there is no shortage of carrots out there and the fear of getting fired is low on the list of people's motivation for working hard.

I then asked my wife if people work harder when they're paid more. "Ridiculous," was her unabashed answer. "People work as hard as the people around them. There's a lot of peer pressure." This latter effect is, upon reflection, very important and noticeably missing from the formal development of shirking models. Employees tend to be able to monitor the performance of similar workers fairly accurately. In situations where the group's output as a whole is observable, it is the job of management to elicit and use peer review to compensate and promote employees. In this context, an employee's perception of management's "fairness" is not merely a nonpecuniary aspect of the employee's compensation (as some of these theories suggest), it is a vital managerial attribute necessary for the hiring and retention of employees.

The evidence concerning the industry effects on individual wages is impressive, but I do not think it suggests a radical departure from the neoclassical paradigm. What this evidence tells me is that there are enormous unobserved quality differences among individuals and enormous unobserved job-attribute differences across industries. I do not find the fact that these industry effects hold across different occupations within an industry especially hard to explain—I would expect to find the most productive secretary working with the most productive executive.

The "sociological" theories in this article appear vacuous. I have no

doubt that any successful enterprise must have shared goals, ideals, and tastes that transcend mere money-making. I also have no doubt that the best way to make profits is to convince employees to share management's fidelity to these higher concerns. Thus I see no power of these alternative theories to refute the conventional view.

As for the ability of these theories to enhance our understanding of cyclical unemployment, I think they are worse than useless. I can think of no substantive issue for which they are even slightly helpful. Examples of what I take to be substantive questions are: Why was unemployment higher in 1982 than in 1965? Why has black teen-age unemployment doubled? Why has the return to going to college in terms of decreased unemployment gone up since 1973? Why has the traditional male-female pattern of unemployment changed recently?

These models may have something to say about why reductions in aggregate labor input fall disproportionately on that small fraction of the labor force that reports itself as unemployed. However, this is hardly the biggest puzzle in macroeconomics and there are other alternative explanations for the relative lack of work sharing. The big question continues to be the sources of fluctuations in aggregate demand for labor. These efficiency wage theories do not suggest any new candidates for aggregate disturbances. Their relevance in macroeconomic dynamics appears confined to perhaps explaining some propagation mechanism of cycles and possibly explaining their persistence, but the model that formally develops this intuition has yet to be written.

Discussion

In responding to the discussants, Lawrence Katz pointed out that Weiss's emphasis on group pressure rather than the efficiency wage has been developed by Akerlof in his gift exchange model. He argued that it is difficult to find a competitive model that explains wage differentials across industries, and that efficiency wage explanations are very likely to account for much of the difference.

Maurice Obstfeld commented that peer pressure may work well in eliciting effort in small groups, but that the efficiency wage argument may be more important in large groups where individual effort is more difficult to identify.

The ability of efficiency wage theory to explain fluctuations of the aggregate level of employment was questioned by Robert Barro. The main problem in business cycle theory is to identify possible sources of large and frequent disturbances, whereas efficiency wage theory appears

to explain the average level of output and unemployment rather than fluctuations.

Martin Weitzman doubted the empirical significance of efficiency wages as a mechanism for producing output fluctuations, despite the theoretical appeal of models with small menu costs of changing wages and prices in which prices become relatively rigid and output flexible. The difficulty with this argument is that it is entirely symmetric. We could as well argue that the costs of changing output are high and the costs of changing price low—then the result would be price flexibility with relatively stable output. There is a serious need for research on adjustment costs to establish which approach is more accurate.

Robert Hall argued that the primary issue in the efficiency wage literature is whether potential workers line up for higher-wage jobs. There is some evidence that at the lower end of the job market, people do stand in line for jobs. On the other hand, a variety of sophisticated effort-eliciting mechanisms—such as tournaments and rising wage profiles—exists at the higher end of the job market. Consideration of these alternative mechanisms is missing from the Katz article. Hall also argued that evidence that wages change when workers change jobs is not convincing support for efficiency wage theory. The wage changes may simply reflect successful selection by firms. Hall also doubted that there was a connection between efficiency wage theory and cyclical fluctuations. Efficiency wages suggest long-term employment relationships, which would surely mitigate employment fluctuations.

On the question of the relation between efficiency wages and economic fluctuations, Lawrence Summers noted that many theories show that wage rigidity is related to fluctuations: efficiency wage theory, which states it is profitable for firms to fix the real wage, is a promising explanation of real wage rigidity. The evidence that firms with different wages can survive in the product market suggests that cost of production is not proportional to the wage, providing support for the theory.

John Taylor stressed the role of relative wages in wage and price dynamics. Theories in which the efficiency of labor is a function of a firm's wage relative to other wages may therefore provide an important element in explaining business cycles.

Olivier Blanchard was worried about the relation between efficiency wage theory and real wage rigidity. One way to put the problem is to consider two models with competitive product markets, one of which has an efficiency wage labor market and the other a competitive classical labor market. Comparison of the degree of wage rigidity demonstrated in the responses of these two economies to a technological shock is completely ambiguous. The only efficiency wage model that gives unam-

biguously more rigidity is the social norm model of real wages, but that model leaves the norm itself unexplained.

Katz concluded the discussion, maintaining that efficiency wage theories are indeed promising for explaining real wage differentials. He accepted the existence of other mechanisms for eliciting effort, particularly in the high-quality job market, but believed there was still room for efficiency wages. He also stressed the importance of wage rigidity in business cycle theories through possibly temporary wage rigidity, although it is true, as Blanchard commented, that efficiency wage theory does not generate wage rigidity in a general equilibrium model.