The Economic Effects of Employment Regulation: What are the Limits?

John T. Addison
Department of Economics
University of South Carolina
Columbia, SC

and

Barry T. Hirsch
Department of Economics
Florida State University
Tallahassee, FL 32306

We have benefited from discussion with or comments from Kin Blackburn, John Chilton, Sam Estreicher, Dave Macpherson, and Paula Voos.
Introduction

In this paper we investigate the controversial topic of government mandates and regulation in the labor market, with emphasis given to empirical evaluation and evidence. We say "controversial" for several reasons. First, at the level of theory, it is not clear that any single framework can accommodate adequately the plethora of labor market interventions in place today, despite the attempt in the literature to categorize and justify mandates according to broad types of market failure. In the second place, there is the critically important issue of fixing the level of a mandate in those circumstances where market failure is identified. Third, it is rare that we have precise estimates of the effects of proposed or adopted policies. And finally, the adoption, interpretation, and implementation of labor market policies are products of the political process, a process that is informed by but not ultimately determined by economic arguments.

In the political and sometimes scholarly debate regarding mandates, the case made by either side is often less than compelling. The more enthusiastic supporters of mandates attempt to justify them with rationale from what are incomplete models, ignore what are unintended although predictable secondary effects of employment regulation, exaggerate benefits and understate costs, and demonstrate little appreciation for market alternatives. Because mandates typically set standards without sufficient knowledge of their effects, proponents sometimes advocate what might be termed a "try it and see" approach -- that is, they call for revision of the mandate as required in the light of experience. This is a rather disingenuous response to concerns about lack of knowledge, given the difficulty of subsequently observing and measuring the consequences of mandates and the endogeneity of public policy. Once in place, mandates may be difficult to modify given the constituencies ranged against change.

Excess is no stranger to opponents of mandates either. Benefits are downplayed, costs are overstated, and recognition of the possibility of market failure is often conspicuous in its absence. Such opponents would be better advised to examine the logic of individual mandates, rather than behave as if even qualified support for any one intervention might transform into unqualified support for all interventions. And even in those instances where the case for opposition to mandates is a strong one, opponents should be cognizant of the political economy surrounding policy formation. The alternative to a

---

1 There is a logical inconsistency here. Were markets free of distortion, then parties could negotiate around mandates, mitigating negative effects. This argument assumes that mandates do not bestow inalienable rights.
given mandate may be an even less attractive policy, thus introducing the need to confront what are essentially second-best arguments for particular policies.

We have in these opening remarks rather over-accentuated the drama surrounding mandates. Mandates may be less far reaching than was the intention of their proponents or the fear of opponents. Mitigating factors that might obtain here include the development of market escape routes, modification of costly measures during the passage of enabling legislation, adjustments in enforcement and interpretation of the law in response to actual experience, and, in some cases, noncompliance. Nevertheless, it is commonly asserted that the U.S. is today subject to a regulatory morass that may ultimately if not immediately bring about death by a thousand possibly small cuts. At the same time, there has been sufficient political support to largely preserve and in some areas expand government's role in the labor market. These seemingly conflicting beliefs underscore the need for an analysis of employment regulation via mandates, an analysis that is necessarily exploratory rather than definitive, and that demonstrates what is still a rudimentary understanding of its effects.

In what follows, we will argue that broad allegations of market failure do not in general provide a convincing case for mandates, that some sources of market failure may not be amenable to correction, that the imprecision of the political marketplace augurs ill for the design of efficient policy instruments, and that methodological problems and data limitations cloud measurement of the effects of employment regulations. This is all rather negative, but the positive case for mandates is implicit in our critique. Although a general case for mandates cannot be made, we do not discount the benefits of individual mandates. And where specific mandates may be dominated by politically less feasible mechanisms, second-best arguments for mandates may prevail. Although allocative inefficiencies accompanying employment regulations abound, empirical evidence, albeit hazy, suggests that a number of mandates may pass at least a crude benefit-cost test. The broad scope of the paper leads us to provide what will be a rather brief treatment of individual mandates, while ignoring altogether some topics addressed elsewhere in this volume (e.g., discrimination).

**Employment Regulation: The Theoretical Framework**

The standard competitive model views labor markets as not fundamentally different in kind from product or other markets. To be sure, standard models account to some degree for labor market
complexities. In particular, the wage is viewed as just one component of the compensation bundle that also includes fringes and a variety of characteristics that accompany the job. These characteristics include effort, job security, job hazards and other working conditions, opportunities for advancement, and so on, with such characteristics resulting in market-determined compensating differentials. The costs of varying the components of the payment bundle differ across firms so that there will be variation in the job characteristics offered in the marketplace. Equally important, workers will not value a given mix of job characteristics in the same way. Given the heterogeneity of firm costs and worker tastes a sorting process takes place. Thus, for example, risk averse workers will gravitate toward firms that can provide stable jobs most cheaply. In these circumstances, wages will still vary with, say, job security, but the gradient will be less steep than would exist absent the matching process that extracts the maximum of worker utility from each unit of labor cost.

If we project on to this distortion-free full-information scenario a mandate that decrees some fixed or minimum level of a given benefit, it follows unambiguously that welfare -- defined as the sum of the joint employer-employee surplus -- cannot be increased, given that firms fulfill an arbitraging function for which they are rewarded by lower costs. The simple case can be seen using a demand-supply diagram, as shown in Figure 1. The mandate shifts downward both the labor demand and labor supply curves -- the demand decrease owing to firms' cost in complying with the mandate and the supply increase owing to employees' valuation of the mandated benefits. Absent distortions or imperfect information, employers will have already provided (i.e., "sold" to workers) all workplace benefits whose value exceeds costs. Hence, a mandated benefit will shift the demand curve downward by more than the supply curve, leading to a decrease in employment and a welfare loss.² Were government to mandate a workplace regulation or nonwage benefit that the market fails to produce, but that workers value by more than the costs, there may be an efficiency gain (i.e., a larger shift in supply than demand). This is possible if adverse selection/asymmetric information lead to underproduction (see below). Likewise, if benefits accrue to

² The loss in surplus is seen by the decrease in area of the triangles (the sum of the employer and employee surplus) moving from $W_1 - E_1$ to $W_2 - E_2$. With linear demand and supply curves and parallel shifts, this area decreases as long as employment decreases. More generally, the magnitude of the allocative losses depend on the elasticities of demand and supply. See Summers (1989) for a more comprehensive presentation. If wages are inflexible downward, by reason of a minimum wage or other constraint, then the allocative costs are even higher.
nonworkers the labor supply curve will not shift downward, yet economy-wide efficiency may nonetheless increase. This is the externality case, also discussed below.

**Market Failure**

The conclusion that mandates produce welfare losses need not follow if there exists market failure. Categories of market failures that might permit an efficiency-enhancing mandate are externalities (including public goods), adverse selection/information asymmetries, and imperfectly competitive labor markets.

Although economic efficiency will provide the primary basis for much of the analysis that follows, it is not the sole criterion for policy evaluation. Redistribution is an important rationale, as seen, for example, by the income maintenance basis for minimum wages and unemployment insurance. Institutionalists (see Kaufman in this volume) have emphasized that workers should have basic rights in the workplace that parallel those in the broader society -- due process, fair treatment, certain freedoms of speech, and the like. And in practice, policy is determined in a political "marketplace" in no small part through the relative strength and efforts of alternative constituencies.

Externalities are spillover (i.e., third-party) benefits or costs not properly accounted for by decision makers. Unlike other market failure arguments, inefficiency resulting from externalities is not normally based on imperfect information, nor is it a product of any irrationality among parties to the employment contract. Rather, imperfectly defined resource rights and high transaction costs result in decision makers basing decisions on socially incorrect prices. When such externalities are present (at the margin), social benefits or costs diverge from private benefits or costs. Allocative inefficiency caused by a mandate, while imposing losses on workers and firms, might pass a social benefit-cost test owing to third-party benefits or reductions in costs.

Examples of externalities often cited in the literature include health insurance, the public goods aspects of certain working conditions, advance notice, severance payments, and parental leave. Taking each in turn, the case for mandatory health insurance may have a basis in the fact that health care benefits are in practice extended by society to indigents so that an uninsured individual who obtains a job with health
insurance reduces the costs to others. Similar reasoning can be applied to mandatory pensions and employment protection. Relatedly, the public goods aspects of employment conditions can create underprovision of valued services since individuals may be expected to underinvest in making their preferences known or in investigating, say, whether the use of certain chemicals in the workplace poses a threat that may be remedied by changes in conditions or the establishment of a compensating differential (Krueger 1994: 302). Advance notice for its part has been advocated on the grounds that plant closings and mass layoffs impose costs on communities in which the plant is located. The argument often also includes reference to imperfect experience rating so that layoffs at one firm raise the costs of other firms. By interfering with private layoff decisions through advance notice, \textit{inter alia}, so the argument runs, the distortions introduced through another measure may be counteracted. Exactly the same second-best argument can be applied to severance payments. Finally, parental leave has been advocated on externality grounds, it being argued that the care of children of working mothers leads to healthier and more productive adults, thereby placing lesser strain on government support systems.

The prisoner's dilemma case is a special case of externalities and pertains to a situation in which individually rational behavior is nonetheless inefficient because it generates an outcome that is less preferred by all the parties than a cooperative (but unstable) outcome. This has led a number of observers to argue that it is possible for a government mandate to shepherd them to a preferred solution. The argument has most recently been applied to worker participation (Levine and Tyson 1990; Freeman and Lazear 1995), which we comment on subsequently.

Another source of market failure is adverse selection, typically associated with asymmetric information or an inability to distinguish among heterogeneous parties. Here, private contracting does not maximize the surplus because of the risks associated with worker/firm heterogeneity. Thus, for example, a firm that voluntarily adopts a just cause dismissals policy may be expected to attract a disproportionate share

\footnote{For a good discussion of the rationale for health care policy, see Cutler (1996).}

\footnote{It is interesting to note that Pigou (1920, p. 187) identified as the "crowning illustration" of negative external effects the work done by women in factories since it threatened injury to the health of their children particularly during the periods immediately preceding and succeeding confinement. Pigou's solution, however was simply to prohibit such work!}
of workers who will supply low effort or shirk, but be difficult to dismiss with cause (Levine 1991). Or if private insurance companies sold unemployment insurance to firms, they would attract as clients those firms most prone to unemployment. As with externalities, the identification of adverse selection or other potential examples of market failure does not establish the case for mandates, nor suggest the complete absence of markets. For example, private insurance companies do provide medical insurance and firms do provide just cause in nonunion regimes. Given the presence of adverse selection, however, it need not follow that provision levels and mix in private markets are optimal.

In recent years the literature on mandates, in particular, has emphasized the role of asymmetric information (Summers 1989). It is argued that where workers or firms have private information that they may be unwilling or unable to disclose, mandates can facilitate an improvement in efficiency. Addison, Barrett, and Siebert (1995) offer an evaluation of this important argument. The informed party in this case is the worker side. It is confirmed that insurance-type components of the labor contract (e.g. maternity benefits and medical insurance) may be underprovided compared to the full-information contract. In these circumstances, it is shown that mandates that increase the level of insurance offered may permit potential Pareto improvements, irrespective of whether or not the pre-mandate contracts differentiated between high- and low-risk workers (see also Aghion and Hermalin 1990). (Differentiation means a separating equilibrium, whereas non-differentiation implies uniform contracts or a pooling equilibrium.) In the model, the improvement comes as a result of gains to high-risk workers exceeding the losses of the low-risk workers. This redistribution in favor of high-risk workers accords well with equity considerations in the case of unhealthy workers, although this is not true in general. For example, it is not so easy to justify redistribution from high-effort to low-effort workers implied by an employment protection mandate.

Apart from these equity considerations, there is a fly in the theoretical ointment. The model assumes homogeneous firms. If a mandate implies higher costs for some firms than others -- maternity leave may be more disruptive to small than to large firms -- the conclusions change. If pre-mandate contracts do in fact differentiate between worker types then the switch in regime (from a pre-mandate separating equilibrium to the pooling equilibrium of the mandate) can be shown to reduce output and no longer guarantee a potential Pareto improvement. The burden of this result is that the designers of mandates
have to be concerned with differences between large and small firms and indeed other sources of firm heterogeneity that might lead to misallocation. In other words, mandates may need to be carefully targeted rather than uniform in reach. Yet targeted policies introduce their own set of informational, legislative, and administrative requirements.

Needless to say, not all mandates conform to the structure imposed by this model. A case in point is advance notice, which can be modeled as forcing firms to reveal their type -- either temporary or permanent employers. Kuhn (1992) argues that a notice mandate can yield actual Pareto improvements. His model assumes that notice contracts are prohibitively costly to write and enforce. Where employers cannot use the wage to signal their type, it is shown that the pre-mandate situation conforms to a pooling equilibrium. Absent valued notice, there is obvious scope for efficiency gains through the better informed quit behavior of workers, abstracting from the thorny problem of the length of notice to be fixed. Of course if one argues to the contrary that firms can commit to give notice, then, even in the presence of market failure on the other side (i.e., the fundamental inability of workers to alienate their right to quit), the case for a mandate can be shown to evaporate (Addison and Chilton 1997).

Labor market distortions arising from market structure, in particular the possibility of monopsonistic power among employers, can provide a rationale for at least some forms of regulation (most notably, a minimum wage). Although pure monopsony (i.e., a single employer and the absence of mobility across labor markets) is rare, a violation of the competitive assumption of a perfectly elastic long run labor supply curve is not. Most firms face upward sloping labor supply curves in all but the very long run, owing to costly mobility among workers and training that is nontransferable across employers, among other reasons. Although upward sloping labor supply curves can indicate a potential for the exercise of monopsony power (i.e., through lower employment and wages), evidence for such outcomes is meager. Even in textbook examples of monopsony, such as the employment of registered nurses (RNs) by hospitals, the presence of upward sloping labor supply curves (Sullivan 1989) does not imply monopsonistic outcomes. For example, Hirsch and Schumacher (1995) find that the wages of RNs relative to similar workers within 252 labor markets vary neither with the number or density of hospitals nor with market size. In a comprehensive survey of the theoretical and empirical literature on monopsony, Boal and Ransom (forthcoming) conclude
that there is little evidence of long-run monopsony power being exercised in the labor market.

More broadly, the fact that most labor markets are at variance with the textbook characterization of a distortion-free competitive market in and of itself provides no rationale for labor market regulation and mandates. As we have seen, development of such a rationale, as has been done for selected types of mandates in the case of externalities and adverse selection, requires careful analysis. That labor markets need not operate "by the book" simply makes good analysis all the more difficult. Indeed, imperfections elsewhere in the system (e.g. progressive income taxation and imperfectly-rated unemployment insurance) in some instances provide a cogent case for mandates if reform of these preexisting distortionary influences is deemed politically infeasible.

Although our focus has been on economic efficiency, we return to the point that equity may figure more largely than efficiency considerations in the design of public policy. Two major questions are posed here. In the first place, measures that may notionally increase earnings equality (e.g. minimum wages) may not do the same for incomes in the presence of disemployment effects. Here the fundamental question is whether the focus should be upon employment mandates that benefit only those in employment. The second question concerns the degree of equality of incomes to be sought. The problem arises in the situation where there is a tradeoff between average income and equality of incomes. Here, efficiency considerations are inextricably linked with redistribution. In this context, we should like to find policies that improve or at least little affect efficiency while redistributing toward the disadvantaged. Yet the disadvantaged are sometimes the victims of measures seeking to place floors under working conditions. Given these problems the search should perhaps be toward alternative measures for improving the prospects of unskilled and disadvantaged workers.

We would be remiss if we failed to emphasize that policy is determined through the political process, and not primarily on the basis of efficiency arguments or benefit-cost calculations. Politicians and, to a lesser extent, regulatory agencies and the courts, respond to lobbying by interest groups. Efficiency considerations do matter, to the extent that the benefits and costs associated with employment regulations are transmitted through the political process. In those cases where a particular workplace arrangement provides large social benefits relative to costs, we should see it evolve voluntarily unless market failure from
externalities, asymmetric information, or the like are important. If market failure is present, but benefits of a policy are high relative to costs, political opposition should not be strong, although it need not follow there will be an organized constituency lobbying for beneficial policies. In cases where there is not a compelling economic case for a workplace mandate or regulation, its passage and implementation face constraints. Lobbying by business groups (or others) expected to be hurt by legislation lowers the probability of passage, or alters legislation in a way that mitigates the costs. Likewise, agency and court interpretation and implementation of legislation are likely to display some sensitivity to costs. For those policies where total benefits and costs move in tandem (e.g., a family leave policy in which few workers participate might entail low costs and benefits, and vice-versa), it follows that opposition (and costs) should be substantial in those very cases where benefits are most significant. Legislation that overcomes what is relatively weak opposition may well provide small gains.

An interesting facet of the political debate surrounding employment regulation is that much emphasis appears to be given to effects that economists would expect to be short run. Economic analysis suggests that in circumstances where workers value mandated benefits, a mandate should in the long run result in a substantial wage offset yet have little effect on employment. Even where workers place little value on the benefits, the costs of a universal mandate (i.e., one covering all workplaces) will be borne largely by workers and have few aggregate employment effects, given the low elasticity of aggregate labor supply (but there will be allocative costs, since even universal mandates will not affect all sectors identically). Yet political lobbying by business emphasizes its added costs and the expected deleterious effects on employment, while organized labor and other worker interest groups rarely express concern for the downward wage pressure associated with mandates. The seeming incongruity between the political debate and long-run economic effects may result from an extreme short-run emphasis in policy making or from general economic illiteracy by the public, politicians, and interest groups. Alternatively, standard economic models may fail to explain the effects of workplace mandates. In what follows, we give little credence to the latter view.

Finally, we should mention two opposing views of the political process that impact on the debate over employment mandates. The "foot in the door" view, heard frequently from regulation opponents,
asserts that adoption of any new mandate makes future regulation more likely. Hence, even though a particular policy proposal may be unobjectionable or entail minor costs, business groups might nonetheless fight it vigorously to reduce the likelihood of future policies with substantial costs. A very different view, what we term the "try it and see" approach, is that we typically have poor knowledge about the benefits and costs of policies until after they are adopted. Once adopted, policies whose costs are high relative to benefits can be modified or abolished. Neither view is altogether without plausibility. By the same token, neither should be adopted as a general rule.

**Evaluation: Measuring the Effects of Workplace Regulations**

A principal argument of our paper is that the evaluation of employment regulations and workplace mandates should be informed by empirical evidence. In this section, we outline in general terms the methodological framework by which employment regulations are assessed, using language popularized in social science literature characterized by quasi-experimental methods (Meyer 1994). This framework is then used in the next section of the paper when we review evidence on individual mandates and regulations.

Suppose we are interested in the effect of policy Z on outcome Y, for example, the effect of a mandated family leave policy on employment and/or wage outcomes. The prototypical experiment would randomly assign the treatment in some markets but not in others. One might then measure the treatment effect by taking the simple difference between outcomes in those markets with and without the policy. That is,

\[
(1) \quad E[Y|Z=1] - E[Y|Z=0],
\]

where the first term is the mean value of outcome Y for the group receiving treatment Z and the second term is the mean outcome for those not receiving treatment. Alternatively, equation (1) can correspond to differencing over time, where Z=0 represents the pre-treatment and Z=1 the post-treatment period. In this case, we are simply observing changes in outcomes before and after implementation of a universal employment mandate.

We rarely have pure public policy experiments and must assess as best we can the mostly non-experimental evidence at hand. Typically, we use regression analysis to calculate something of the form

\[
(2) \quad E[Y|X,Z=1] - E[Y|X,Z=0],
\]
where $X$ represents a set of measurable control variables intended to account for (at least some of) the
differences in the treatment and non-treatment group that are correlated with outcome $Y$. Or in the case
where we are differencing over time, $X$ accounts for changes in $Y$ over time not related to $Z$.

What are the concerns regarding the single differencing approach described above, in which we
either compare outcomes between observations with and without a policy or the change in outcomes among
a group before and after a policy? Meyer (1994) identifies factors that threaten the internal and external
validity of simple experimental approaches such as (1) or (2). Internal validity refers to whether or not one
can draw the inference that measured differences in the outcome variables are in fact *caused* by the
treatment, within the context of the particular study. External validity refers to whether or not the results in
a specific study can be generalized to different settings, times periods, groups of individuals, and the like.

An important threat to internal validity or, stated alternatively, a major cause of false inference, is
*omitted variables*. The failure to account for factors that affect $Y$ and are correlated with $Z$ (or $Z$ and $X$) will
lead to biased estimates of the treatment effect (omitted variables uncorrelated with $Z$ and $X$ will not bias
estimates). In the context of parental leave, this would include unmeasured factors that affect employment
or wages and are correlated with the presence of a family leave policy across, for example, time, firms,
states, or countries. *Treatment endogeneity*, whereby adoption of a treatment or passage of a law requiring
treatment is in response to past, present, or expected future outcomes, also may bias estimated treatment
effects. For example, the adoption of parental leave policies by companies, states, or countries may be
determined in part by the level of (or changes in) wages and employment. Emphasized in the experimental
literature are issues of *selectivity* or non-randomness with respect to assignment to a treatment group. In the
context of non-experimental data, selectivity can be thought of as a form of omitted variable bias -- those
receiving treatment (e.g., those in companies or locations with family leave) differ in a systematic way from
measurably identical individuals not receiving treatment. Bias also can result from a failure to account for
underlying trends in outcomes, in effect, a form of omitted variable bias. As discussed subsequently,
experimental designs that include (appropriate) non-treatment comparison groups are intended to account
for underlying trends.

*Data quality* problems arise from imperfect measurement of variables. Estimates of the effects of
family leave may be biased if policies treated as identical in fact differ across time, firms, states, or countries. For example, a single dummy variable denoting presence or absence of maternity leave would not account for whether leave is paid, whether leave is voluntary or mandatory, the length of leave, job retention rights, or a host of issues regarding policy interpretation and implementation. Another potential source of faulty inference is an overly restrictive model specification that does not properly measure the causal effect of explanatory variables. For example, the effects of family leave may differ with respect to other factors (e.g., gender, marital status, or number and age of children), necessitating a specification that interacts the treatment variable with other explanatory variables.

A lack of external validity can arise where the population from which a researcher's data are drawn differs substantially from the larger population to which one seeks to generalize results. Interactions of the treatment with the setting, time period, or the outcome variables often make it inappropriate to generalize results from a specific study or data set to other settings. In this respect, the selectivity and endogeneity concerns discussed above threaten not only internal validity (i.e., the inference of causality) but also external validity or the ability to generalize outside the particular setting. For example, the effects of voluntary family leave measured among companies adopting such policies is not likely to provide an accurate measure of outcomes that would result from those same policies if adopted or mandated among companies not previously adopting them. Likewise, effects observed in the past may not be identical to effects in the future.

In practice, researchers use a number of alternative approaches for measuring the treatment effect associated with a particular policy. As emphasized by Meyer (1994), a goal of the researcher ought to be either to have exogenous variation in the treatment variable or, if not exogenous, to understand the source of its variation. And as emphasized below, it is important that there be appropriate comparison groups with which outcomes from the treatment group can be compared. The key feature of the alternative approaches is that they employ multiple comparison groups and differencing techniques to measure treatment effects.

A common research design is what we refer to as the single difference approach, as seen previously in equations (1) and (2). Yet the difference approach is not likely to provide reliable estimates of the treatment effect, unless one is confident that one has controlled for all important factors influencing outcome
Y and that treatment Z is exogenous. In order to improve the accuracy of estimates, researchers often employ one or more comparison groups and calculate a difference-in-differences (DD) estimator of the treatment effect. For example, if a state (or several states) mandated family leave policies during some period, one can measure the treatment effect by taking the difference between the change in outcomes among companies in the state(s) that adopted family leave policies and the change in outcomes among companies in the comparison state(s) not adopting family leave. One can also include additional control variables reflecting factors that influence employment and wage outcomes but whose influence over the period differ between states adopting and not adopting the policy. Here it is particularly important that the comparison group(s) be similar to the treatment group and, if not, that adequate control variables be included that account for differences. The use of multiple comparison groups provides a check on the robustness of estimates.

Multiple comparison groups also make possible alternative ways of measuring treatment effects if the groups differ in ways likely to influence their response to a policy. For example, suppose we expect women but not men to be affected by a mandated family leave policy. This enables researchers to provide a difference-in-differences-in-differences (DDD) estimate of the treatment effect. This estimator might measure the difference between states that did and did not enact family leave policies in the changes over time (the pre- and post-enactment periods) in female-to-male employment (or wages). Whereas the DD estimator provides for time period controls by introducing comparable non-treated states during the same years, the DDD estimator additionally controls for state-specific effects by introducing a within-state comparison group (i.e., males) assumed not to be affected by the policy. Of course, the quality of an estimator need not be a function of the extent of differencing. In this example, the DDD approach may well provide a less accurate measure of the treatment effect than the DD approach if men are significantly affected by family leave and if unmeasured state-specific effects are unimportant. The important point is that policy evaluation may be enhanced by the use of multiple comparison groups. It is important the researcher exercise good judgment as to what questions should be addressed and the appropriateness of
alternative research designs.\(^5\)

Additional points regarding research design warrant mention. An advantage of examining changes in rather than levels of outcomes is that one may thereby control for what would otherwise be endogenous policy change. If the adoption of a policy is related to the (past, present, or future) outcome level, treatment effects estimated in levels are likely to be biased. While adoption of a policy is often affected by the outcome level that it in turn is likely to affect, it need not follow that adoption is related to changes in outcomes. Also, while the method of analysis is important, it is critical that there be a high degree of signal-to-noise in the data itself. In the example above, reliable estimates are likely only if a number of countries make substantial changes in family leave policy during one or more periods, and if changes in outcomes occur shortly following the policy change.

The methods described in this section can be used to characterize much of the empirical literature measuring the effects of employment regulations. Three themes warrant emphasis. First, empirical studies cannot be conducted in boilerplate fashion; rather, reliable empirical work requires good judgment, knowledge of the data, and an understanding of the processes that generate both the employment policy and the labor market outcomes under study. Second, it is rare that one has reliable empirical estimates on which to base policy decisions. And third, even were reliable estimates of policy effects available (say, the employment and wage effects from family leave), this would not constitute a full-fledged benefit-cost or welfare analysis of alternative policy proposals. Ideally it is the latter that might best inform (if not influence) policy decisions.

**Workplace Mandates, Regulations, and Selected Public Policies: An Analysis**

*Workers’ Compensation*

State workers’ compensation laws provide for employer-mandated no-fault insurance covering workplace injuries, coupled with limits on liability from lawsuits. The passage of workers' compensation laws in several states during the 1910s constituted one of the earliest and more important government interventions into the workplace. Currently, workers’ compensation is compulsory in all but three states

---

\(^5\) Studies using the *DDD* approach in the context of family leave are Gruber (1994) and Ruhm (1996), described subsequently.
(New Jersey, South Carolina, and Texas); even in those states most employers voluntarily choose coverage in order to limit their liability. Total payments from workers' compensation are sizable, amounting to $44.1 billion in 1992. Of this total, 41 percent was for hospital and medical payments. Payments for workers' compensation exceed those for state and federal unemployment insurance, food stamps, supplemental security income (SSI), veteran programs, or housing programs (U.S. Bureau of the Census 1995: Table 585).

Workers are eligible for medical and indemnity (lost wage) benefits when disabled by job-related injury or illness. Employers are liable regardless of fault, but may dispute the severity of an injury or illness and challenge whether it is work related. Workers' compensation costs are nominally paid for by employer payroll taxes. A few states require that employers insure through a state operated insurance system. Many states operate a state system but permit insurance through private insurance companies or self-insurance. In most states, the typical small employer purchases private insurance and large employers self-insure. Insurance companies set rates based on a combination of manual rates, which vary on the basis of rather detailed industry/occupation breakdowns, and experience rating (larger established firms have full or close to full experience rating).

Is there a strong economic rationale for workers' compensation? We believe there is. Absent some form of no-fault insurance for workplace injuries, a large number of accidents would be handled by the courts using a negligence standard. The joint costs of determining liability under these circumstances would be large, substantially larger than the indemnity and medical costs of most accidents. Only a minority of workers would be compensated for injuries if it were necessary to prove company (or co-worker) negligence, and payment would be received long after most medical expenses occurred and wages were foregone. In principle, workers and firms could enter into employment contracts that include forms of no-fault injury insurance not unlike workers' compensation. Were it not for a mandatory system, it is likely we would see such contracts in some workplaces, assuming adverse selection were not too serious. But contracts of such detail involve considerable transaction costs and are rarely the norm. Workers' compensation systems, in effect, provide a standard contract, albeit one in which the parties cannot bargain away. Indeed, the fact that workers' compensation arose prior to the depression and New Deal social legislation suggests that there was not unduly strong business opposition to mandatory workers'
compensation. Workers' compensation might be justified by forms of market failure discussed in the previous section. Workers (at the relevant margin) may not have good knowledge about workplace dangers and employers have little incentive to truthfully reveal such information. This implies that compensating risk differentials would be inefficiently small and workplace safety too low. Mandatory compensation for injuries forces the employer to take account of the cost of workplace injuries (even if the costs are fully shifted to workers in the form of lower wages). Externality arguments might also be made. Absent workers' compensation, much of the medical costs and some of the indemnity costs from workplace injuries will be shifted to others. Indeed, an advantage of having workers' compensation rather than other forms of health insurance pay for the medical cost of workplace injuries is that it shifts costs to parties whose behavior can affect safety.

Although a strong case can be made for some system of mandatory workers' compensation, such a system creates inherent inefficiencies, primarily by reason of moral hazard. Moral hazard here refers to a situation where insurance coverage affects the actions of insured parties; specifically, the probability and extent of injury and illness claims. Moral hazard is inevitable absent symmetric information between insurers and those insured and complete experience rating of premiums. Workers receiving generous compensation for workplace injuries are more likely to make claims for benefits than they would do were compensation lower, and time away from work is likely to be longer for any given health limitation. If health care providers and injured workers receive compensation for medical treatment, more treatment is likely to be provided than would otherwise occur.

If the level of safety were suboptimal in an unregulated market, the introduction of workers' compensation should be associated with a safer workplace. Likewise, higher benefit (cost) levels should result in fewer workplace injuries. Yet the empirical evidence unambiguously points to a positive relationship between both injury claims and duration and the level of benefits (for reviews, see Ehrenberg 1988; Krueger 1990; Butler 1994). In contrast, there is a negative relationship between workplace fatalities

---

6 Fishback and Kantor (1994b) argue that state workers' compensation was a precursor to the development of the welfare state.
and benefit levels. Taken together, this evidence suggests that employers respond to workers' compensation
by making the workplace safer, but that moral hazard on the part of workers increases claims from non-fatal
accidents. The moral hazard to some extent will take the form of reduced risk avoidance by workers on the
job. The more important behavioral effect is that in the event of an injury, absence from work and
indemnity claims are more likely the more generous are benefits, the more informed are workers, and the
greater protection offered workers to management discouragement of claims (Hirsch, Macpherson, and
DuMond 1997; Weil in this volume). There is also evidence that medical costs are higher for workers'
compensation patients than for patients with similar medical conditions not resulting from work-related
injuries. Although some of the difference may result from higher prices, most of the difference appears to
be a greater use of medical services for workers' compensation patients (Durbin, Corro, and Helvacian 1996;
Johnson, Baldwin, and Burton 1996).

Although employers nominally pay for workers' compensation, theory suggests that most of the
costs should be shifted back to workers in the form of lower wages, given aggregate labor supply curves of
close to zero elasticity. Fishback and Kantor (1994a) provide evidence on relative wage changes among
states enacting and not enacting workers' compensation early in the century, and find full or almost full
wage offset. Gruber and Krueger (1991), using more recent data, report that wage growth is inversely
related to changes in benefit levels.

It is the variation in indemnity benefits across workers that provides the basis for much of the
empirical work on workers' compensation. Because provisions vary by state and benefit replacement rates
vary with worker earnings within states (owing to benefit ceilings), the standard empirical strategy has been
to examine differences in claim rates with respect to benefit levels using industry, state-by-industry, or (less
frequently) individual data. An example of a recent study explicitly using a quasi-experimental approach is
Meyer, Viscusi, and Durbin (1995), which examines evidence in Michigan and Kentucky following
increases in benefits for higher but not lower wage workers. They find evidence that in both states duration

---

7 Although benefit elasticities differ considerably across studies, depending on research design, the outcome measure,
and data, a ballpark estimate of the average benefit elasticity is .40, implying that, say, a 10 percent increase in benefits
is associated with a 4 percent increase in claims. Our evaluation of the evidence places the benefit elasticity at the
lower end of the range of estimates in the literature, closer to .20 to .30.
of absence from work increased among higher wage workers following the increase in benefit levels, but that there was no change in behavior among lower wage workers.

Workers' compensation provides an important source of insurance to workers for medical costs and indemnity losses arising out of workplace injuries. A system mandating coverage may well be superior to what would exist absent government mandates and regulation, although we cannot describe with confidence what the unregulated counterfactual would be (this would depend in part on the nature of a country's health and unemployment insurance systems). But our current system of state workers' compensation insurance is costly and suffers from no small degree of moral hazard. As evident with other government programs, inefficiencies increase as program generosity expands. It is difficult to provide adequate compensation for workplace injuries in a way that does not entail substantial inefficiency costs.

The Fair Labor Standards Act: Minimum Wages and the Overtime Premium

The Fair Labor Standards Act of 1938 (FLSA) introduced, *inter alia*, two important forms of federal wage regulation -- minimum wages and an overtime premium. Because the FLSA is examined elsewhere in this volume, our treatment is brief. The overtime provision requires that overtime wages of at least one and one-half times the straight-time hourly wage be paid for hours worked per week in excess of 40. Although interpretation is not always clear, the provision applies to workers whose pay varies directly with hours worked. Employees are exempt if paid a bona fide salary and if their duties are performed independently of a supervisor or detailed company procedures. Roughly two-thirds of all workers are subject to the overtime pay provision (Ehrenberg and Smith 1994: 138; Zachary 1996), with the major categories of excluded employees being executive, administrative, and professional workers, outside salespersons, and agricultural workers, as well as some groups of workers covered by other labor legislation (e.g., truck drivers, airline personnel, and railroad workers).  

A principal argument used to support the overtime premium is that it will increase employment. To discourage routine use of overtime and increase employment, proposals were made in Congress during 1979

---

8 Compliance with the overtime provision is not universal. Estimates by Ehrenberg and Schumann (1982) and Trejo (1991) suggest that about 10 percent of covered employees working in excess of 40 hours do not receive overtime payments. Noncompliance results because of financial incentives to firms to hold down labor costs, legal ambiguities regarding coverage and a poor understanding of the law among many workers and employers, and relatively weak enforcement and low penalties by the Department of Labor (Zachary 1996).
and 1985 to increase the overtime premium to double time (Ehrenberg and Smith 1994: 138n). An overtime premium mandated economy-wide cannot increase employment significantly, however, unless there exists excess unemployment (we ignore the unlikely possibility that an overtime premium attracts individuals into the labor force). If there exists high unemployment, then the argument that an overtime premium increases employment is logically correct if the straight-time wage (W) remains fixed. Firms determine the optimal mix between employment and hours per worker (see Ehrenberg and Smith 1994: 136-43); an increase in marginal wage costs from W to 1.5W may shift firms' mix toward employment and away from overtime hours.

This argument for increased employment need not hold if the overtime premium causes the straight-time wage to decrease so that the combined wage-hours combination is of equivalent value to workers (Trejo 1991). That is, the availability of jobs offering overtime hours may result in an equilibrium straight-time wage that is slightly lower than it would be in the absence of the premium. Trejo refers to this possibility as the "fixed-job" model, as compared to the "fixed-wage" model assumed previously. Trejo attempts to compare these models empirically, testing for wage and employment effects associated with overtime pay. Because the overtime provision is federal, has remained at 1.5W over time, and applies broadly across the labor market, estimating its effects reliably is difficult. Based on occupation and industry of employment, Trejo compares wages and employment of male hourly workers who are likely to be covered by the overtime provision to those for workers not likely to be covered. His evidence is not fully consistent with either the fixed-wage or fixed-job model, leading him to conclude that the overtime provision may cause both small increases in employment and small decreases in straight-time wages.

We know even less about the costs of the FLSA overtime provision resulting from reduced scheduling flexibility. Absent overtime pay regulation, firms requiring workers to regularly or occasionally work long or variable hours would be required to pay a compensating differential only if workers regarded long hours with pay as a disamenity. Labor market sorting would result in what is likely to be a rather modest wage premium for long hours.9 With the overtime provision in place, firms will often choose to employ existing workers at 1.5W rather than hire additional workers at W, owing to variable product

---

9 The widespread presence of moonlighting (or dual jobs) is consistent with there being for many workers a constraint on hours and the desire to work longer hours on the primary job (Paxson and Sicherman 1996).
demand and fixed employment and training costs. But the overtime premium does raise the cost to firms of using variable work hours, and is likely to increase reliance on temporary workers in positions where firm-specific skills are minimal (for evidence on the use of such workers, see Polivka 1996). Flexibility costs appear to have renewed interest in modifying the FLSA overtime provision, with business groups and the Clinton administration engaged in discussion over legislation providing provisions for compensatory hours or an accounting period longer than a week (Zachary 1996; Stout 1996; Siwolop 1996).  

Although empirical evidence on the effects of the overtime mandate is highly limited, theory and available evidence suggest that a) the effects on aggregate employment are small, b) utility from the compensation package may change little for workers, relative to what would exist absent the mandate, given that compensating premiums might otherwise exist and the straight-time wage would be higher, and c) there exist efficiency costs from reduced scheduling flexibility for workers and firms, but the magnitude of these costs is unknown. As with other employment mandates, we suspect that both the benefits and costs associated with the FLSA overtime provision are lower than those asserted by proponents and opponents. But unlike the case with other mandates, an argument for market failure in the absence of an overtime provision seems difficult to sustain.

The FLSA also established a minimum wage (MW) for workers, with subsequent legislation periodically raising the (nominal) wage and expanding coverage. As is the case for the overtime standard, the rationale for a minimum wage cannot readily be rooted in market failure arguments. Textbooks routinely show that a minimum wage imposed on a monopsonist can increase wages and employment. Yet employers of low-wage workers rarely fit the standard monopsony model. In an effort to account for what is a weak empirical relationship between the minimum wage and teen employment (see below), theorists

---

10 A business-backed bill (the Working Families Flexibility Act) that would allow workers, with employer consent, to choose to take compensatory hours rather than pay at time and one-half, passed the House on July 30, 1996, with no Democratic votes (St. Louis Post-Dispatch 1996: 5A).

11 One could make an externality argument. By discouraging long hours of work, one provides parents the opportunity to spend more time with their children, thus enhancing future human and social capital. Absent the policy, parents would underinvest in their children either because they are not fully cognizant of the impact of time with their children, or because they do not take into account the non-private benefits. Even if one were to accept the argument of parental underinvestment, it does not follow that the overtime premium is an effective way to increase investment in children. Many families will circumvent the constraint on work hours through moonlighting or increased hours of work by another family member.

12 Recent legislation raised the federal minimum wage from $4.25 to $4.75 beginning in October 1996, and to $5.15 in September 1997. Some states require wages in excess of the federal minimum.
recently have proposed models in which small employers in competitive markets face upward-sloping supply curves or behave as if they were monopsonists (see Card and Krueger, 1995). It is far too early, however, to evaluate the generality or empirical importance of such models. Despite the absence of a clear efficiency-based rationale for minimum wage laws, public support is widespread, albeit on grounds other than economic efficiency. The most typically stated rationale for the minimum wage is to improve the well-being of workers least well off. Public support may also spring from a belief that fairness dictates some minimum level of compensation for work, or some minimum spread in relative wages within the workplace. The principal critique of the minimum wage by employers and economists is that higher labor costs for low-skill workers will decrease employment among those least well off, and that the minimum wage does little to reduce poverty or family income inequality.

We will not attempt a thorough analysis of the burgeoning economic literature, much of it focusing on teenage employment effects (see surveys by Brown, Gilroy, and Kohen 1982; Card and Krueger 1995). Rather, we will briefly summarize what we believe can be concluded from the research, focus on the methodological approach used in recent studies, and relate our conclusions regarding minimum wage laws to the larger issue of workplace mandates and regulations.

Most variants of a standard neoclassical model predict that a binding minimum wage will reduce employment (e.g., Card and Krueger 1995: Ch. 11). Until recently, much of the literature estimating the employment effects of MW were based on quarterly time-series data relating changes in the teen employment to population ratio to a prevailing minimum wage measure (typically a coverage-adjusted relative wage), holding constant other determinants of teen employment. An often stated range for the teen elasticity of employment with respect to MW is -.1 to -.3 (Brown, Gilroy, and Kohen 1982), although studies using more recent data produce estimates closer to -.1 (Wellington 1991; Card and Krueger 1995: Ch. 6).13

Recent analyses, most notably those authored (or co-authored) by Card or Krueger, have used new

---

13 The finding that teen MW elasticities are far smaller than standard labor demand elasticity estimates need not be inconsistent, because a relatively small fraction of teens (particularly older teens) are affected by the MW. For example, if one-quarter of teens are affected by the MW, a teen elasticity with respect to the MW of -.2 would correspond roughly to a labor demand elasticity of -.8.
methods and touched off a number of new studies. The Card and Krueger studies correspond closely to the quasi-experimental approach described in Section III. For example, Card (1992b) examines the effects of the April 1989 MW increase from $3.35 to $3.80 on teen employment. At the time the new MW went into effect, several high wage states (including those with state minima exceeding the federal minimum) had as few as 10 percent of teens that should have been affected by the federal minimum (i.e., with pre-April 1989 wages ranging from $3.35 to $3.80), whereas low wage states (typically those in the south) had as many as half of their teenagers affected. The expectation from standard theory is that one should observe lower teen employment growth (or a larger reduction) in low wage states with a high fraction of teens affected than in high wage states with a low fraction affected. Card finds no significant differences, however, in employment growth based on differences in the fraction of teens affected. One can interpret this as a difference-in-differences (DD) approach, comparing the difference in employment changes between markets most affected and those least affected by the treatment (MW), controlling for other factors affecting teen employment such as demand conditions.

Other studies from Card and Krueger use a similar methodology. Card (1992a) examines the employment effects of an increase in California's MW, but finds differences in employment growth to be largely unaffected by the proportion of workers who should be impacted, or relative to comparison states. A study by Katz and Krueger (1992) using self-collected data on Texas fast-food restaurants found differences in employment growth following an increase in the MW were largely unrelated to the previous wage level in the restaurants. And in one of the more publicized (and disputed) studies, Card and Krueger (1994) collected data on fast-food restaurants in New Jersey and Pennsylvania. Using a DD methodology similar to the Texas study, they found that employment growth among the New Jersey restaurants was unrelated to the proportion of workers affected by the New Jersey state minimum wage law. In addition, they used the Pennsylvania restaurants as an alternative comparison group, comparing changes in employment growth in New Jersey restaurants (the treatment group) to growth in Pennsylvania restaurants, the latter serving as a comparison group not subject to a minimum wage increase. In neither case did they find that MW increases

---

14 For a description, as well as an analysis of prior literature, see Card and Krueger (1995). For a critique of the "revisionist" literature, see the July 1995 Industrial and Labor Relations Review symposium.
had a negative and significant impact on employment.\textsuperscript{15}

The new research on the minimum wage reinforces and strengthens rather considerably prior evidence concluding that employment effects associated with changes in the minimum wage, at least at levels historically adopted in the U.S., are rather small. Our understanding of why MW employment effects are so small is on less firm grounds. But the absence of large effects of MW on employment clearly undercuts the case for opposing moderate increases in the minimum.\textsuperscript{16} Less encouraging is the fact that the anti-poverty effectiveness of minimum wages is rather limited (see Gramlich 1976; Addison and Blackburn 1996a; for a more positive view, see Card and Krueger 1995: Ch. 9). This is not surprising, given the weak linkage between low wage workers and family income. MW laws disproportionately affect teenagers, yet teen workers are distributed evenly throughout the family income distribution. Despite Card and Krueger’s more upbeat evaluation, we are not convinced that MW increases have a substantial anti-poverty effect, even if they do have an equalizing effect on the distribution of earnings. The MW appears to be a far less effective anti-poverty tool than direct income maintenance programs or alternative earnings-related policies such as the earned income tax credit (see Burkhauser, Couch, and Glenn, forthcoming).

Ultimately, one is reduced to making second-best arguments for the minimum wage. Not unlike the conclusion that will be made with respect to other worker mandates or regulations, the often vitriolic debate over minimum wage laws exaggerates its importance. Neither the benefits nor costs match those asserted by its more vocal proponents and opponents.

\textit{Employment-at-Will/Wrongful Discharge}

Outside of the union sector, the employment relation in the U.S. is largely governed by the common law employment-at-will doctrine, the legal basis of which rests on notions of freedom of contract and

\textsuperscript{15} In fact, Card and Krueger report that employment fell by more in Pennsylvania than in New Jersey. This study has received considerable criticism based on measurement error in data collected by their phone survey. In particular, Neumark and Wascher (1995) reexamine the evidence, substituting establishment records provided by restaurants. They find a small but negative MW employment effect based on the New Jersey/Pennsylvania comparison, but fail to find a significant MW effect based on the within-state New Jersey data.

\textsuperscript{16} It should be noted that weak employment effects from the MW are not a result of low compliance. Evidence from Card and Krueger and others indicates clearly that MW laws do increase wages. Nor does the evidence support the proposition that small employment effects can be explained by a changed mix in the compensation package, with higher wages offset by lower training and fringes. Fringes and training costs are low on most MW jobs. A higher MW does, in fact, increase costs to businesses, some of which are passed forward to consumers. Card and Krueger (1995, Ch. 10) provide evidence from an events study surrounding the 1989 legislation showing that expectations of a higher MW are associated with lower market values among companies that are low-skill labor intensive.
mutuality of obligation or consideration. The at-will principle is often referred to as "fire-at-will," because the employer can legally terminate an individual open-ended contract (i.e., one of indefinite duration) without cause (we ignore for now implicit contracts binding workers and firms). The at-will principle, however, has been attenuated over time as a result of incursions of state legislatures and the courts.

Courts in almost all states have recognized some exceptions to the traditional common law. The most commonly recognized exception has been a so-called "public policy exception," designed to protect whistle blowers, those exercising statutory rights, and those refusing to act unlawfully. Secondly, courts in a majority of states have held that company personnel handbooks and oral statements to the employee constitute implied-in-fact contracts that preclude dismissal without proper cause. Thirdly, a minority of states have proceeded well beyond these two exceptions to argue that there is an implied covenant of good faith and fair dealing that governs the employment relation, effectively requiring just cause. Despite this judicial activism, only Montana has thus far enacted just cause legislation per se. Other states have introduced such legislation (for details, see Krueger 1991: 650-52).

The case for unjust dismissals legislation may be made on either first- or second-best grounds. The former would typically have a basis in freeing up valuable information possessed by workers and in encouraging consummate as opposed to perfunctory cooperation. Absent a mandate, adverse selection would result in "problem workers" (i.e., those who might lose jobs with employment-at-will but could only be dismissed with difficulty under just cause) being attracted to those firms who voluntarily adopt an unjust dismissal policy. Firms, thus, are reluctant to adopt such policies and too little job security is produced absent a mandated policy. The second-best case might be predicated on the argument that such a mandate might offset imperfections in the UI system that encourage too much turnover. Since unjust dismissals legislation raise employment adjustment costs, however, the second-best argument is not altogether a comfortable one.

But as we have indicated there is no national unjust dismissal legislation in the U.S. Empirical analyses have thus been devoted to an examination of the attenuation of hire-at-will by the courts. One irony of this is the suggestion that a mandate might secure a low-cost and predictable alternative solution to
the casuistic rulings of the judiciary.\textsuperscript{17} Surveys of average awards granted in unjust-dismissal cases point to substantial (and positively skewed) settlements. In successful cases, the median initial award approximates $180,000 (this figure does not account for post-trial actions that may reduce awards) (Dertouzos, Holland, and Ebener, 1988; Shepard, Heylman, and Duston, 1989). That being said, total costs (including legal fees) to employers would appear to be less than .1 percent of the total wage bill. But note that one cannot directly infer the effects of the legal system based on total award and legal costs, since high awards and an unfavorable judicial climate will deter businesses from dismissing workers without clear-cut cause, lowering the number of total cases and awards.

Turning to the empirical evidence, Dertouzos and Karoly (1992, 1993) have provided a direct test of the effects of legal incursions, using state level data for 1980-87. States are distinguished according to which of three (hybrid) wrongful dismissal doctrines their courts have embraced, and whether or not the remedies provided for are contractual or tort-based. They examine a fixed-effects employment model in which regressors include gross state product, the growth in gross state product, year dummies, and dummies for legal doctrine and type of remedy. The latter are instrumented to account for their nonrandom distribution across states (i.e., adoption of a particular doctrine or remedy may be influenced by unmeasured factors affecting employment).

Dertouzos and Karoly report that aggregate employment is on average 2.9 percent (1.8 percent) lower in the years following a state's recognition of tort (contractual) damages for wrongful termination. Regressions run for other combinations of doctrine and remedy confirm that it is the availability of tort remedies rather than type of exception that drives the disemployment result. Recalling that legal costs represent a tiny proportion of the wage bill, the authors conclude that in their employment decisions employers seem to be reacting as if the costs were much greater. Despite this evidence, Dertouzos and Karoly argue that the benefits of unjust dismissal legislation may be worth the employment sacrifice. Among other things, they speculate that a clear legal doctrine limiting unjust dismissal might reduce uncertainty about the enforceability of implicit labor market contracts and allow the parties to more fully

\textsuperscript{17} An analogy can be drawn with workers' compensation laws, passed with some support from labor and management.
reap the benefits of long-term employment relationships.¹⁸

Other analysts have interpreted court rulings in a rather different sense. For example, Krueger (1991) takes the position that the courts and current legal system are costly, if not yet so costly as to lead to the adoption of legislation. His maintained hypothesis is that the prospects for actual legislation are improved by draft legislation, which in turn is directly linked to the erosion of employment-at-will by the courts. He models the determinants of the likelihood that a state will introduce unjust dismissal draft legislation, using state data for 1981-88. His logit estimates suggest that the good faith and public policy exceptions by the courts (he does not examine remedies) are significant determinants of states drafting legislation -- the probability that a state legislature will introduce legislation is increased by 6.7 and 8.5 percent if its court system has recognized good faith and public policy exceptions, respectively. Causality appears to run from legal incursion by the courts to legislation, rather than the converse. Features of the draft legislation he examines include limits on employer liability. Krueger links the success of unjust dismissal legislation in Montana to a substantial prior reduction in employer liability. To repeat, for states that have introduced but not passed legislation, it is suggested that the threat raised by court attenuation is not yet great enough.

The evidence marshalled by Krueger offers some support for the notion that unjust dismissal legislation may come to represent for employers "an acceptable political compromise" (Krueger, 1992).¹⁹

This interpretation is consistent with the idea expressed elsewhere in this paper that enabling legislation for mandates may ultimately represent a less radical departure from existing practice than is often alleged.

There has been much debate about employment protection more generally, and invidious

¹⁸ An indirect test is Hamermesh (1993), who examines the speed of adjustment of employment to output in nine two-digit industries for the period 1973:4 to 1988:3. He finds a reduction in the adjustment speed in retail trade and finance, and an increase in responsiveness in construction, mining, durable manufacturing, nondurable manufacturing, wholesale trade, and transportation, public utilities and communications. Since the former group of industries have levels of union density that are low and most of the latter have relatively high union densities, Hamermesh speculates that the observed reduction in flexibility in the former group could well reflect an erosion of hire-at-will, namely, a reduced willingness on the part of nonunion employers to both lay off and hire workers.

¹⁹ But see also Stieber and Block (1992: 795), who argue that Montana is simply an outlier. They claim that employers will oppose and continue to oppose such legislation because it "would extend protection against unjust dismissal to a vastly larger number of discharged employees than are now generally involved in court suits." The latter are typically middle- and upper-level management groups rather than hourly workers. Krueger (1992) responds that there are degrees of opposition, and that business resistance to legislation will be less fierce in states where the status quo is least attractive.
comparisons drawn between Europe and the U.S. in this regard. Yet the evidence is both sketchy and contentious. Perhaps the best-known study is that of Lazear (1990), who examines the effect of severance pay on employment and unemployment in 20 nations over the interval 1954-86. His evidence suggests that more generous statutory severance pay (and longer advance notice) is associated with lower employment and elevated joblessness. A replication of this study by Addison and Grosso (1996) confirms the directional influence of severance on the two outcome measures but their country fixed-effects estimates show the contribution of severance pay to unemployment is not material. That still leaves intact the adverse impact of severance on employment, although it remains to be seen whether this result survives reestimation with a richer mix of country-level controls. Interestingly, Addison and Grosso also report that advance notice seems to be associated with generally favorable labor market outcomes. One conclusion is, then, that different job protection mandates may have different effects on the outcome indicators. But the broader and more important conclusion from this research is that more substantive progress awaits formal parameterization of individual regulations, recognition of the broader institutional structure in which they are adopted and administered, and, importantly, more widespread (but not universal) adoption so that outcomes can be measured using quasi-experimental methods.

Unemployment Insurance

The principal rationale for UI is probably less one of economic efficiency, and more one of stabilization and income maintenance. UI acts to stabilize consumer spending and provides incentive to firms to smooth employment and production (Moss, 1996). Danziger and Gottschalk (1990) report that UI payments reduce the poverty rate of unemployed individuals by about 20 percent, although, as with minimum wages, UI benefits are not received disproportionately by the poor (Hutchens 1981). Efficiency arguments for UI can also be made. While some workers desire insurance against losses from unemployment, adverse selection coupled with asymmetric information makes voluntary private insurance markets incomplete at best. Hence a government mandate that employers provide insurance (as with workers’ compensation) can be welfare improving. Public provision rather than mandates might be justified on the grounds that unemployment risks are not easily diversifiable for private insurance companies (Anderson and Meyer 1993) or, for that matter, states. Indeed, UI is structured so that more federal funds
flow to states with high unemployment.

Much attention has focused on the inefficiencies associated with UI and, in particular, the impact of imperfect experience rating on layoffs and the effects of UI benefits on unemployment duration. On the former question, it is clear that imperfect or incomplete experience rating -- firms do not bear the full cost of worker layoffs through higher taxes -- exacerbates unemployment by causing excessive use of temporary layoffs (e.g., Topel 1983, 1984). Although there has been no major reform of the system to make UI taxes more closely accord with the layoff risk, we should note that among Western countries the U.S. is an anomaly in having even partial experience rating.

Analysis of the effects of UI on the behavior of the insured unemployed has received close attention. From a theoretical perspective it is clear that subsidizing unemployment should lead to longer joblessness. Both the static labor-leisure model (Moffitt and Nicholson 1982) and the alternative job search model (Mortensen 1970) produce this result. The difference between the models resides in the prediction of the latter that UI may improve the quality of job matches. Empirical research has confirmed the longer unemployment duration of those covered by UI. Although there are acknowledged statistical problems associated with left censoring and accounting for unobserved individual heterogeneity, a typical result from U.S. research is that a 10 percent increase in the UI replacement rate increases unemployment by between 0.5 and 1 week, with some studies reporting higher estimates (Meyer 1990). Similarly, longer entitlement periods are associated with elevated unemployment (Katz and Meyer 1990). At issue is whether this longer unemployment is productive of income. Unfortunately, the evidence on the effects of UI on subsequent earnings is both sketchy and sufficiently varied in approach and results to lead experienced observers to reach opposing interpretations (cf. Burtless 1990; Cox and Oaxaca 1990). Recent research by Addison and Blackburn (1996b) evaluates this issue and concludes that a modest favorable impact of UI on earnings is evident in the data, but only when a comparison is effected between UI recipients and nonrecipients (as opposed to samples consisting of claimants only). These favorable effects, where observed, nevertheless fall well below those reported in earlier studies (e.g. Burgess and Kingston 1976). Furthermore, Addison and Blackburn find no evidence of lesser subsequent job changing among UI recipients than among nonrecipients.
Possible efficiency effects of UI remain empirically elusive and are likely to remain so in the absence of experimental studies. Some such studies have been conducted on the use of reemployment bonuses awarded to those who quickly find jobs. In a survey of evidence from the bonus experiments, Meyer (1995) concludes that bonuses speed up job finding on the part of the insured unemployed, that this does not detract from subsequent earnings growth, and that the benefits and costs of the programs may be roughly equal. Meyer, however, notes the difficulty of proceeding from these experiments to permanent policies. The problem in a nutshell is that a reemployment bonus makes filing for a benefit more valuable, leading those who are eligible for benefits but who currently do not claim them to do so. The empirical suggestion is that this effect could be substantial. The bonus program also acts as an indirect subsidy toward the growth of relatively unstable jobs and industries with short unemployment spells. This pessimistic evaluation does not, however, carry over to experiments involving job search assistance, also reviewed by Meyer. Here the evidence from five programs points to reduced UI benefit receipt, increased earnings, and savings to the UI system and government as a whole. Note that these job search experiments strengthened not only the employment service, but also the UI work test. (Johnson and Klepinger 1994)

An overall evaluation of UI cannot ignore the inefficiencies associated with the system as presently constituted. Imperfections of the UI system are often deployed to present a second-best case for a number of other mandates. Improvements made to UI, while welcome, appear marginal from this perspective since they do not tackle incomplete experience rating. And from the perspective of income maintenance, the safety net remains incomplete for those with long durations of joblessness. The potential efficiency improvements to UI from job search assistance and reemployment bonuses only partially address the problems of disadvantaged workers for whom more fundamental changes are likely to be required. Such assistance is not likely to come from changes in the UI system.

Pensions

Pensions provide an important mechanism by which employers influence worker selection, turnover, effort, and retirement behavior (for reviews, see Ippolito 1987; Gustman, Mitchell, and Steinmeier 1994). Pensions facilitate a more optimal matching of workers with job slots in the face of asymmetric information. For example, pension contributions by a firm, a high degree of wage tilt (i.e., low initial wages
but rapid wage growth), and a lengthy vesting period discourage applications from workers with high
discount rates and workers most likely to leave the firm. Such a compensation structure is particularly
attractive for companies where turnover is costly. Traditional defined benefit (DB) pension plans are
structured in a way that penalizes both early exit or delayed retirement from a firm, with pension value
maximized at an optimal retirement age. And as argued by Lazear (1995: Ch. 4) and others, backend
loading of compensation through pensions (or wage tilt) may increase effort (i.e., reduce shirking) among
workers to avoid dismissal from the firm.

The Employment Retirement Income Security Act of 1974 (ERISA), the Tax Reform Act of 1986,
and additional federal regulations have restricted firm discretion with respect to pensions (for a summary of
regulations, see Hoopes and Maroney 1992). We will focus on two forms of pension regulation -- vesting
provisions and pension insurance. First, beginning in 1986 most private sector pensions were required to be
fully vested within five years of employment. Employees leaving after five years must receive either a
lump-sum payout or be eligible for future pension benefits, based on employer contributions or promises
(workers receive their own contributions even if they leave prior to vesting). Second, ERISA created the
Pension Benefit Guarantee Corporation (PBGC), which acted to insure at least partial pension payments to
workers whose employer terminated a pension plan, while at the same time placing restrictions on firms’
ability to terminate plans. Companies with DB plans were required to maintain a minimum level of funding,
disclose information to the PBGC, and pay insurance premiums (not risk rated) to the PBGC to cover the
cost of failed plans.

What are the possible economic justifications for mandated vesting requirements, information

20 Defined benefit plans promise workers specific benefits, based on a formula that is typically the product of an
earnings base (e.g., earnings averaged over the last several years), years of service, and a “generosity factor” (typically
about 1 percent in the private sector and more in the public sector). DB plans discourage early and late exit from the
firm. Because the earnings base is in nominal terms, workers leaving a firm mid-career suffer a substantial loss in
pension wealth since the earnings base will have a low real value at the time of retirement. Pension wealth also
declines if an employee works late into life since payment is received for fewer years. In contrast to DB plans, defined
contribution (DC) plans (including 401(k) plans) provide a payment by the company into individual worker pension
accounts, where funds are then managed largely by workers. DC plans are by definition fully funded, portable
(following vesting), and are not structured to penalize early or late retirement. In recent years, there has been a large
shift away from DB and toward DC plans.

21 These provisions do not directly affect DC plans, since they are fully funded. While ERISA provides for minimum
funding requirements, Internal Revenue Service (IRS) provisions place maximum limits on funding, since company
earnings placed in pension funds are not taxed until distributed to workers. Many firms have terminated DB plans and
switched to DC plans because they believe IRS limits produce less than optimal funding of future pension liabilities.
disclosure, pension insurance, and other governmental regulations of pensions? As stated forcefully by Ippolito (1987, pp. 459-60):

"On its face, the pension contract is tenuous. In exchange for lower cash wages, the firm promises workers pension payments many years in the future. Yet it can either terminate the plan any time or fire workers prior to retirement; in either case, the firm can impose large capital losses on workers. Because of the complexity of the contract and its long-term nature, informational problems would appear to abound. And, the contract is largely implicit, making it unenforceable in the courts. Moreover, there is a potential for a lemons market; if some firms perpetrate fraud, the expected "quality" of all pensions is reduced. Pensions appear to offer a classic example of a product that could not survive in an unfettered competitive market, one that would require at least some government regulation to ensure its survival."

In short, a case can be made for government regulation on grounds of asymmetric information, adverse selection, and externalities. To this list, we would add the goal of increasing savings. Insuring a minimum amount of pension savings can be justified on externality grounds, because without such savings older persons will require greater public support. It might further be argued that individuals might prefer a form of forced savings as a disciplining device, since individuals' short-run behavior is often relatively myopic and inconsistent with long-run discount rates (Thaler 1992, Ch. 8).

How do such arguments apply specifically to vesting requirements and pension insurance? With respect to vesting, it can be argued that workers have inadequate knowledge about pension provisions and firms may default on implicit contract promises. Such arguments appear more convincing for regulating financial practices and mandating information disclosure than for vesting regulations, however, since vesting provisions are relatively easy to understand and reputational effects limit firms' opportunistic short-run behavior. We believe a stronger argument for mandatory vesting is to encourage greater pension savings than would otherwise occur among workers who regularly switch employers. One would expect a

---

22 Despite the strong a priori case for market failure, Ippolito and others do not believe that there is evidence for widespread market failure and violation of implicit contracts, owing primarily to the disciplinary role reputational effects have on employers.
five-year cliff vesting provision to be associated with low quit rates just prior to five years of tenure, followed by a spike in quits. We are not aware of strong evidence for a vesting spike in quits, suggesting that young workers are not well informed about pension rules or that they highly discount the accrued pension benefits. Either explanation would lend support to a mandated vesting rule, although not necessarily at five years.

Why mandatory federal pension insurance? Prior to the mandate, private markets did not develop because of adverse selection (firms in poor financial shape are most likely to insure), asymmetric information (companies have better knowledge than insurers), and moral hazard (insured companies may underfund absent minimum financing requirements). Federal insurance pooling across risk classes overcomes some of these difficulties, and might be warranted on the grounds of imperfect worker knowledge and a reduction in externalities resulting from terminated pension plans. Mandated federal insurance makes the most sense where firm failure is a possibility and where there are few private market alternatives.

That being said, federal insurance is not without its own set of problems, in particular adverse selection. Many firms have dropped out of the PBGC insurance pool by phasing out their DB plans, instead making DC plans available to workers. Plans remaining with the PBGC then become more risky and require higher premiums, in turn accelerating the movement out of DB plans. Terminated plans, whose pension liabilities are shifted to the PBGC, primarily have been underfunded plans of union firms. Despite the ability for a union, as workers' agent, to monitor a firm's pension funding, this has been outweighed by the incentive for union firms to rationally underfund plans (and use debt financing) as a means to moderate future wage demands (Ippolito 1985; Bronars and Deere 1991).

Although there has been much recent theoretical and empirical research on pensions (see Ippolito 1987; Gustman, Mitchell, and Steinmeier 1994), the literature provides insufficient information to assess the benefits and costs associated with federal regulations. We believe that government policies intended to mandate or encourage private savings are appropriate, the principal lever here being the subsidy given pensions through their tax deferral status. Likewise, disclosure and financial regulation of pensions seems justified on information and externality grounds. A case can be made for federal insurance owing to
asymmetric information and adverse selection, although the PBGC has not been immune to these same forces. But regulation comes with a cost, in particular distortions in the value-enhancing contractual arrangements between workers and firms. The magnitude of benefits resulting from government pension policies is largely unknown, while at the same time costs and unintended consequences associated with such policies have been readily apparent.

Advance Notice

After many false starts the U.S. now has an advance notice mandate.23 The Worker Adjustment and Retraining Notification Act (WARN) (Public Law 100-379) was enacted on August 4, 1988, and became effective February 4, 1989. The Act requires employers with 100 or more full-time employees to provide 60-days' written notice of a plant closing or mass layoff to workers or their representatives. A plant closing is defined as the shutdown of a single site of employment, or part thereof, involving 50 or more employees. A mass layoff is a layoff with duration of more than six months that affects at least one-third of the workforce (but not less than 50 employees) at a single site of employment. If 500 workers or more are involved, the one-third rule does not apply and notification is automatic (for further details, see Addison and Portugal 1991).

Considerable effort has been devoted to analyzing the effects of voluntary notice on unemployment and earnings. Less attention has been accorded the effects of WARN, but there have been interesting theoretical developments that might usefully be addressed. In what follows, we briefly review the theoretical work, and then note the difficulty of making inferences about the effects of a mandate based on evidence from voluntary notice. We conclude with observations on the practical impact of notice legislation.

The general case for an advance notice mandate is typically predicated on the existence of externalities or preexisting distortions in UI. Recent theoretical work, however, has focused on other considerations. As was noted in our introduction, models based on asymmetric information and prohibitive transactions costs have produced the result that notice, although valued by the parties, will not be provided

in equilibrium under freedom of contract (Kuhn, 1992). The alternative mechanism of using the wage to
signal the temporary or permanent status of the firm provides a solution – with permanent firms offering
higher wages that serve to increase their future retention rates and which temporary firms find unprofitable
to mimic – but only if firms can reset wages in response to information about the need to lay off workers.
Where firms have to fix the wage before they know their type (i.e., viability), there will result a pooling
equilibrium characterized by a uniform, noncontingent wage. In these latter circumstances, a mandate can
benefit both sides because of the better informed separation decisions of workers.

It is possible to erect alternative models in which a mix of notice and no notice contracts typify
equilibrium under freedom of contract, even in the presence of market failure, once the restrictive notion of
prohibitive contracting costs is relaxed. Addison and Chilton (1997) argue that the firm's option of not
giving notice substitutes for the inability of the worker to commit by virtue of the prohibition on involuntary
servitude. In these circumstances, a no-notice contract (at an adequate wage) is a simple arrangement for
retaining the worker and avoiding inefficient quits. Nevertheless, a mix of notice and no-notice contracts
will emerge in the equilibrium, their relative frequency being determined by the distribution of firm-specific
parameters of the model. It is shown that there is no under-provision of notice under unrestricted
contracting, with the result that a notice mandate cannot increase the joint surplus of the employment
relation. Indeed, in some cases the mandate will actually reduce the joint surplus by causing excessive quits.
Also, although the mandate may be simply redistributive, there are also instances in which workers, and not
simply firms, are adversely affected.

Here as elsewhere our discussion points to the sensitivity of expected outcomes to assumptions.

The plot thickens somewhat when we come to examine the empirical evidence, most of which is based on
the CPS Displaced Worker Surveys (DWS). It can be concluded from evidence on voluntary notice that
prenotification "works" in the sense that it is associated with reductions in joblessness and perhaps with
improved earnings development as well. But it is a large leap from these results to the conclusion that
notice should be mandated. After all, opponents of a notice mandate would argue that beneficial effects of

24 The evidence is reviewed in Addison and Portugal (1991). More recent treatments are Addison and Portugal (1992a)
and Ruhm (1992, 1994).
voluntary notice are to be expected; if notice is valuable, it will be contracted (and paid) for by the parties. Empirically, the endogeneity of notice is addressed by instrumenting notice or employing Heckman-type selection procedures. In the latter case, we may in principle then use selection coefficients from notified and non-notified worker equations to predict the effect on, say, unemployment of mandating notice (Addison and Portugal, 1992b). Unfortunately, it has proved difficult to model the notice endogeneity using available data. Even if it were concluded that notice were exogenous, so that the beneficial effects of notice on, say, unemployment duration obtained from single-equation specifications could be generalized to a regime of mandatory notice, debate would not end. All that is being measured in these studies are benefits from notice. The costs of notice have still to be reckoned with and we have even less information on these.

What are the costs associated with advance notice? One potential cost for employers is premature quits by notified workers. Indirect evidence in Fallick (1994) suggests that these costs could be substantial. Fallick, who examines the determinants of notice, finds that concerns about early exit reduce significantly the probability that an employer will provide notice. A larger cost may be a plant having to continue operations longer than is privately (and perhaps socially) optimal. Deere and Wiggins (1996) suggest that workers often receive little notice in those firms where the need for closing is known only a short time in advance. The longer the time horizon (proxied by facility obsolescence, for example) the longer the amount of notice given. The considerable variation in the amount of notice across firms may be driven by heterogeneity in firm circumstances. In the Deere and Wiggins sample, actual notice was frequently longer than that contractually required, but was also far shorter than the 60 days set under WARN. The legislation does allow for a reduction in the notice period in cases of unforeseeable business circumstances, but Deere and Wiggins caution that this provision is not satisfactory given the potential for legal wrangling, and requirements that employers bear the burden of proof.

Enough has been said to indicate that our knowledge of the effects of WARN is rudimentary, particularly on the costs side. Our understanding of the potential benefits is better, but even here, unemployment may not be reduced if the observed gains of notified workers come at the expense of non-

25 Although we note in passing that Ehrenberg and Jakubson (1988) fail to detect any evidence of compensating differentials.
notified workers -- although there are few concrete signs of this in the (cross-national) data (Addison and Grosso 1996). The evidence is still such that an opponent of legislation might readily accept the evidence that those who have been notified in the past have benefitted, yet nonetheless believe that the costs of making such notice mandatory exceed the benefits.

Might not the effects of WARN be considerably more muted in practice than opponents and antagonists alike have speculated? To examine this possibility we turn in conclusion to a descriptive statement of the impact of the Act on the incidence of notice. In the six-year period prior to WARN, roughly one half of all (displaced) workers received some form of notice. But this was predominantly informal in nature -- three-quarters of those with notice either 'expected' to be laid off or received verbal notice. Just 8.6 percent of workers received written notice of greater than one month (3.9 percent having between one and two months' written notice and 4.4 percent having more than two months' written notice).26 In the three-year period following the Act, a rather smaller proportion of workers (8.2 percent) received written notice of greater than one month.27 Once we control for observable factors that might be related to the provision of notice (e.g. nature of displacement, expected eligibility for UI, the state unemployment rate, firm size, union density, etc.), the conclusion that WARN failed to affect receipt of written notice still stands (Addison and Blackburn 1994). One interesting result, however, is that short written notice of less than one month became more prevalent after WARN. This hints at an increased formalization of layoff procedures that may be linked to the Act, although such notice is not of the type actually mandated under WARN.

Why did the legislation not lead to greater notice being given workers? The most likely explanation is that usual employer layoff behavior leaves most displacements uncovered by the Act. Thus, for example, a recent General Accounting Office (1993) study which investigated 650 layoffs that affected 50 or more workers in facilities employing at least 100 workers found that 49 percent of these layoffs would anyway have been exempted because they fell below the one-third rule. Another 15 percent of the cases were also identified as being exempted on other grounds. And it will be recalled that WARN only covers firms with at

26 The principal data set available to researchers, the DWS, identifies three lengths of written notice: less than one month, between one and two months, and more than two months. The ambiguity is that WARN grants 60 days' notice, an interval that might be reported by the DWS respondent as either of the two longest categories.
27 Of the post-WARN sample, 4.4 and 3.8 percent received between one and two months' notice, and more than two months' notice, respectively.
least 100 workers, thereby excluding at a stroke 35 percent of the workforce from potential entitlement to notice. Given that the Act has had no discernible impact on notice -- at least for its first three years of operation -- it is tempting to conclude that this may have been the intention of its designers in Congress. Politicians simultaneously could claim support from one constituency (workers) in voting for the legislation, while those same (or other) politicians could appeal to another constituency (employers) by voting to attenuate the mandate's impact through the adoption of escape routes in those instances where economic and political costs may be high.

Maternity Benefits/Family Leave

Maternity and family leave arrangements provide an interesting case because a variety of theoretical arguments can be used to make the case for market failure. The externality argument is that imperfect information on the part of a parent as to the importance of child care, or undervaluation of the welfare of children in the parental utility function, raises costs to society via the subsequent well-being or behavior of their children. More compelling is the argument that asymmetric information, in particular the limited information available to the employer as to high-risk workers (i.e., those likely to become pregnant or access leave) would lead to adverse selection among firms voluntarily adopting maternity benefits or family leave provisions. A mandate may provide a (potential) Pareto improvement in which high-risk workers' gains exceed the low-risk workers' losses (Addison, Barrett, and Siebert 1995). Complications are introduced by firm heterogeneity. Where high-risk workers cause greater difficulties for some firms than others, and if separation has achieved the appropriate allocation of labor to begin with, then it automatically follows that a mandate, which amounts to enforced pooling, must adversely affect the allocation of labor. The benefits provided are group specific, applying primarily to younger women. Mandating higher benefits for this group means that their wages must fall. Where this is constrained by antidiscrimination rules, fair wage laws, or relative-pay norms there will be disemployment among the target group.

U.S. research on maternity benefits per se has focused on whether the cost of maternity insurance has in fact been shifted. The principal study is that of Gruber (1994) who, as previously noted, employs a DDD methodology. The basis of his inquiry is legislation in selected states during 1975-79 that outlawed treating pregnancy differently from other health insurance benefits. Prior to legislation, there typically was
no coverage or pregnancy benefits were treated differently. He seeks to ascertain the effect on earnings from the state laws (inter alia) among a "treatment group" of married women aged 20 to 40. In order to identify this effect, his estimating equation controls for year effects to capture common national trends in earnings, state effects to control for state-specific differences across areas fixed over time, and finally state-by-year effects to control for state-specific shocks correlated with the passage of the laws. The earnings of treatment individuals (married women aged 20-40) in the states passing laws are compared with a set of control individuals in those same states (individuals aged over 40 and single males aged 20-40), and the differences before and after the change in state laws are expressed in relative terms; that is, relative to earnings changes among the same groups in states that did not make changes in the law.

Gruber reports a 5.4 percent fall in the relative wages of prime-aged married women in states that passed laws compared to the change in relative wages in the nonexperimental states, ignoring other observables that affect earnings, and 4.3 percent following control for observables. Since the cost of expanding maternity benefits amounted to between 1 and 5 percent of wages, these estimates (as well as others provided by Gruber) suggest that the full cost of the mandates were shifted back to prime-aged women in the form of lower wages. Although women's hours worked increased, numbers employed fell by almost 2 percent. The suggestion is that although the mandate is valued by the treatment group, some disadvantage is experienced by part-time workers -- not surprising given that pregnancy insurance is a fixed cost with respect to hours. Nevertheless, the main conclusion by Gruber is that efficiency may not have been adversely impacted.

Finally, we turn to a study with a bearing, albeit indirect, on the Family and Medical Leave Act of 1993 (FMLA). Ruhm (1996) provides a cross-national study of 16 European countries during 1969-88, using a DDD methodology much like Gruber's analysis for states, to measure the effects of alternative mandated parental leave provisions on employment, working hours, and wages. He reports that total leave and its paid leave component are on net positively related to employment, although the effects ultimately

---

28 Ruhm (1996; forthcoming) provides a useful, succinct critique of the parental leave literature. He notes that the exemptions and restrictions of the FMLA limit its reach to around 31 percent of working women. The Clinton administration is expected to propose expansion of family leave to require most employers to grant up to 24 unpaid annual hours of leave for family obligations (e.g., parent-teacher conferences, transporting a parent to a doctor's appointment) (Stout 1996; Siwolop 1996).
become negative for extended leave provisions (total leave beyond 52 weeks and paid leave after 26 weeks). Ruhm finds rather material reductions in wages at long durations of (paid) time off work, namely, a 3 percent wage loss associated with paid leave beyond 26 weeks. Provisions with respect to total (as opposed to paid) leave suggest modest wage gains at 27 weeks and minor losses at 52 weeks. Ruhm interprets his evidence as supportive of parental leave of up to three months (the FMLA standard) and as suggesting that employers are able to shift to workers some of the costs, particularly at lengthier durations. We are wary of drawing strong inferences for efficiency from this single study, but nonetheless find it interesting that the research again suggests that the level of a mandate appears to affect outcomes. This result underscores the need for caution in designing a mandate, even where the case for market failure is a strong one. The relatively weak standards associated with what are politically feasible mandates may have no small degree of economic justification.

The National Labor Relations Act: Worker Voice and Economic Consequences

Discussion to this point has centered on what can be considered "direct" mandates, whereby the state requires that covered firms provide specific personnel policies (e.g., family leave, advance notice), employment terms (e.g., minimum wages, an overtime premium), or insurance coverage (e.g., workers' compensation, UI, and ERISA). In this section, we consider what can be considered an "indirect" form of regulation -- U.S. labor law. The National Labor Relations Act (NLRA) provides the legal structure for private sector union organizing and collective bargaining. The regulation is "indirect" in the sense that government does not dictate terms of employment but, rather, establishes a framework in which workplace outcomes may be determined locally through collective bargaining.

The principal economic justification offered for the NLRA or Wagner Act at the time of its passage was the need to redress what was widely regarded as a fundamental imbalance of power in the labor market. Important non-economic arguments for the legislation included the desirability of supporting industrial democracy and due process. The NLRA thus sought to encourage collective bargaining and to prohibit specified unfair labor practices on the part of management. Recent discussion of the NLRA has focused more on its role in maintaining a labor relations system often dominated by confrontation rather than cooperation, and the Act's alleged chilling effect on worker participation within nonunion companies.
The sources of unequal bargaining power have been addressed by Kaufman (1989, 1991), who identifies generalized unemployment, monopsony power, and discrimination/labor market segmentation as the chief culprits. The diagnosis of a need for countervailing power during the 1930s remains controversial (see, for example, Reynolds, 1991). Kaufman (1996) has recently added an important clarification of the macroeconomic context of his earlier analysis. Drawing on Congressional testimony, the language of the act, and Senator Wagner's legislative record and speeches, Kaufman concludes that the ultimate goal of the Act was "greater economic stability through better economic balance," and as a component part of a coordinated economic program designed to protect against the downward spiral of wages and labor standards or "destructive competition." Regardless of the validity of these economic arguments, we find Kaufman's historical description compelling. We would also note that the notion of destructive competition is alive and well in Europe, where the justification of mandates establishing a floor of worker rights has often rested on precisely this rationale. Here it is usually also argued that unfettered competition threatens consensus and may call into question the achievement of greater economic integration (Addison and Siebert 1991, 1994, 1997). We would also note that modern-day concerns with widening income inequality, sharpened by an improving environment for profitability, offer political appeal for mandates and other forms of government intervention. In short, many of the same arguments that motivated the Wagner Act still lurk in the contemporary wings and once again signify that distributional concerns may transcend those of efficiency.

With these preliminaries behind us, we now turn toward an examination of the economic effects of the Wagner Act. The principal consideration is of course the effect of labor cartelization. If the monopoly face of unionism were all that occupied us, our narrative would be comparatively straightforward, but as is well known unions have been endowed in modern research with potentially important offsetting collective voice attributes that dilute the monopoly effects. Moreover, the consequences of a precipitous decline in private sector unionism have to be addressed, issues which much exercised the Dunlop Commission on the Future of Worker-Management Relations (Commission 1994).29

---

29 Union density among private sector workers declined from 24.2 percent in 1973 to 10.3 percent in 1995 (Hirsch and Macpherson 1996: 10, Table 1)
Absent widespread monopsony or other distortions, the encouragement given to unions by the Wagner Act can initially be evaluated using the standard on-the-demand curve union monopoly model. The result is allocative inefficiency stemming from the union wage premium. Too few workers will be employed in the union sector and too many in the nonunion sector. Any wage rigidities in the nonunion sector (by reason of minimum wages, the floor of welfare payments, and queuing for union jobs) serve to exacerbate the problem since output losses from unemployment now have to be factored in. Yet conventional estimates of the efficiency loss are tiny. The crude estimates offered by Rees (1963) amount to just 0.14 percent of GNP. More sophisticated estimates by DeFina (1983) that allow for corporate income tax distortions -- which artificially reduce capital intensity -- are even smaller. To be sure, such calculations assume that union gains are effected by means of a costless transfer from the rest of the community but they do not give tremendous traction to the cartelization argument. (A more serious criticism that centers on the static nature of such empirical applications will be dealt with below.)

Moreover, conventional estimates of the costs of the union rule book and featherbedding do not wildly inflate the output loss. Rees (1963) guesstimates that manning standards, workplace restrictions, and other working practices probably exceed 0.3 percent of GNP, while Allen's (1986) careful analysis of the construction industry concludes that removal of work rules would reduce staffing levels by 3 percent and total costs by 2 percent. These are nontrivial but still modest estimates, and may potentially be offset by other union job regulatory practices. Finally, nearly all studies indicate that the prototypical union bugbear, strikes, even if laid at the door of unions -- an assumption that is vastly naive (Bertram, Siebert, and Addison 1985) -- are unlikely to affect output materially by reason of intertemporal substitution of production and the emergence of bargaining protocols designed to take strikes out of competition (Reder and Neumann 1980; Neumann and Reder 1984).

All of this provides thin gruel for opponents of unionism in general and of the enabling legislation of the NLRA in particular. Developments associated with the notion of collective voice (Freeman 1976) further seem to attenuate the economic case against unions as combinations in restraint of trade. The collective voice argument is, of course, that unions lower turnover and establish more effective governance structures in workplaces characterized by public goods (i.e. shared working conditions), complementarities
in production, and long-term contractual relationships. Following Brown and Medoff’s (1978) important paper, numerous empirical studies sought to obtain estimates of the effects of unions on productivity using a unions-in-the-production-function test. We have summarized the unions and productivity literature elsewhere (Addison and Hirsch, 1989). Our conclusion at that time was that no compelling case existed to support the presence of a statistically or quantitatively significant positive union productivity effect on average, and that such effects were anyway inconsistent with other pieces of evidence concerning profitability and employment. This conclusion seems to have stood the test of time and the publication of more refined estimates based on firm data (e.g., Clark 1984; Hirsch 1991a). Although we would not wish to overstate the precision with which union productivity effects can be estimated, or deny the existence of substantial positive and negative effects in particular settings, productivity effects appear to be small on average.30

This evidence is of course not exactly favorable to the collective voice model. And work on profit effects seemed meantime to further qualify the reach of that model. That is to say, the evidence points rather clearly to negative union effects on firm profitability, irrespective of the particular profit indicator used (rate of return on capital, Tobin’s q, price-cost margin, etc.) and level of aggregation. Addison and Hirsch (1989) survey the earlier evidence; more recent studies include Hirsch (1991a, 1991b), Becker and Olson (1992), and Bronars, Deere, and Tracy (1994). One response to the adverse impact of unionism on this aspect of firm performance has been to argue that unions simply tax away monopoly rents (Freeman and Medoff 1984). But there are few signs to indicate that union profits or wage gains stem from concentration-related profits (Hirsch and Connolly 1987). This does not deny that unions capture rents, although these are more likely to accrue because of regulatory barriers (e.g., airlines, trucking, railroads, and telecommunications prior to deregulation) or quasi-rents associated with long-lived tangible and intangible capital (Hirsch 1991a).

Yet it is entirely possible that these facts, namely, lower profitability in union regimes coupled with little or no effect on productivity, are nonetheless consistent with (static) efficiency. This theoretical point is 30 For a virtually equivalent conclusion, see the recent survey by Booth (1995). An evaluation more favorable to the voice interpretation is offered by Belman (1992).
derived from the so-called "efficient contracts" model (McDonald and Solow 1981), which demonstrates that wage-employment outcomes on the demand curve -- the context of the monopoly union model -- are inferior from the point of view of the bargaining parties to some alternative combination off that curve with lower wages and higher employment, affording the union higher utility and the firm higher profit. Although efficiency from the perspective of the parties is not in general Pareto optimal from a societal perspective, in the strong efficiency case (i.e., a vertical contract curve) at least the outcome is neutral; that is, output, prices, investment and employment are identical to the competitive (i.e. union-free case). The parties can thus be envisaged as maximizing the size of the pie and then bargaining over the division of the surplus. The negative effect of unions on firm profitability, noted earlier, may constitute no more than a lump sum tax.

But the story does not end here because in all of this it is assumed that capital is fixed. Not surprisingly, therefore, recent theoretical and applied research has sought to endogenize capital (on the former, see Hirsch and Prasad 1995; Addison and Chilton 1996). One of the more thorough applied studies is that of Hirsch (1991a) because of its attempt to link the issues of profit and investment in physical and intangible capital. He is concerned to measure both direct and indirect effects of unions on investment. The direct effect stems from the union tax on the quasi-rents to long-lived and relation-specific capital, leading firms to cut back on investment so as to equate the marginal post-tax rate of return with the marginal financing cost. The indirect effect of unions on investment arises from the higher financing costs in the wake of reduced profitability.

Using data for 1968-80 on approximately 500 publicly-traded U.S. manufacturing firms, Hirsch first estimates the profit effect for a typical unionized firm relative to a nonunion firm, reporting a reduction in market value of roughly 20 percent (for a somewhat lower estimate, see Hirsch 1991b). He next estimates a physical capital investment equation in which the independent variables are union coverage, profits, and detailed firm and industry controls. Other things equal, it is found that the typical unionized firm has 6 percent lower capital investment than its observationally equivalent nonunion counterpart. Allowing for the profit effect increases the estimate to about 13 percent; that is, about half of the overall impact of unions is an indirect effect. Hirsch repeats the exercise for intangible capital (annual investments in R&D). His findings imply that the average unionized firm has 15 percent lower R&D, holding constant profitability and
the other determinants. Allowing for the indirect effects induced by lower profitability only modestly raises
the estimate. These deleterious union effects on capital investment have been confirmed in subsequent
studies (e.g., Hirsch 1992; Becker and Olson 1992; Bronars and Deere 1993; Bronars, Deere, and Tracy
1994). Another interesting finding is that debt-equity ratios are higher in unionized firms (Bronars and
Deere 1991, 1993). The basic idea is that firms increase debt to reduce opportunistic future union wage
demands, especially in the presence of firm-specific capital.

Overall, then, the evidence indicates that union effects are real and distortionary. Whatever the
positive benefits of collective voice on firm performance, these seem to have been overshadowed by rent-
seeking behavior as reflected in reduced profitability and lower investment. It is not surprising, therefore,
that we observe substantial management resistance to unions and slower growth in union than in nonunion
employment. Indeed, what is commonly characterized as "deindustrialization" is in no small part
"deunionization" (Linneman, Wachter, and Carter 1990). The literature on union organizational strength has
of course stressed a number of alternative explanations for private sector union decline. One such theme is
the contribution of the growth in unfair labor practices by management (Freeman 1988), which is not
unrelated to union effects on wages and profits. Another closely related theme is the apparent shift in the
National Labor Relations Board's interpretation and enforcement of labor law (Sockell and Delaney 1987;
Allen 1994). Both views suggest that the Wagner Act has been diluted and that labor law now needs to be
strengthened (Weiler 1990), whereas the rent-seeking approach, while not condoning flouting of the law by
management, would view the erosion of union bargaining power as both natural and not altogether
unwelcome, other things being equal.

Other things may not be equal, however. One interesting issue here is raised by the apparent decline
in the demand for unionism on the part of workers (Farber and Krueger 1992). A change in worker
sentiment could in part reflect changes in the industrial relations environment noted above. It may also be
expected to reflect the expanding role of the government (and the courts) in extending worker rights.
Causation likely runs in both directions, however, raising the question of whether the decline in union
density may have gone too far. This necessarily imprecise argument rests on the problems with mandates
identified in this paper as well as the litigation stemming from laws that rely wholly or in part on individual
lawsuits for their enforcement.

Stated bluntly, the decline in private sector unionism (i.e., "indirect" regulation) and absence of a formal mechanism for worker voice in most private sector workplaces may be leading to an undue reliance on costly "direct" mandates and a litigious labor relations environment. The Dunlop Commission was of course much exercised by the impact of law and administrative regulation on the workplace. Implicitly its "solutions" look to unionism. For example, in recommending experimentation with in-house dispute resolution the Commission (1994: 29) seems to favor unions in expressing its concern "about the potential for abuse of ADR [alternative dispute resolution] created by the imbalance of power between employer and employee."

A final issue raised by the decline in unionization relates even more directly to the Wagner Act. It is widely felt that worker participation is pro-productive but is likely to be underprovided by the market, especially if the law places constraints on participative arrangements within nonunion regimes. There may thus be first- and second-best arguments favoring a participation mandate. It has also been argued that participation is less effective in nonunion regimes or where employee groups are not largely independent of management control, although employee autonomy also can foster the same type of rent-seeking behavior associated with unions.

As is well known, section 8(a)(2) of the Wagner Act prohibits employer-sponsored employee involvement schemes from engaging in functions similar to those of independent unions. Although recent legal cases are consistent with a strict interpretation of the law,31 it has not been established that section 8(a)(2) has in fact had a chilling effect on the introduction of participation schemes in nonunion regimes (Rundle 1994), despite common assertions to the contrary. There remains the possibility of repealing or modifying it, the latter approach having been adopted in the TEAM (Teamwork for Employment and Management) Act.32

32 The TEAM Act limits the scope of 8(a)(2) and allows employer organized and funded worker participation groups in nonunion plants and offices. The measure was passed by the House of Representatives in September 1995 and by the Senate in July 1996. The President vetoed the bill on July 30, 1996, on the grounds that the legislation "would undermine the system of collective bargaining that has served this country so well for many decades" (St. Louis Post-Dispatch 1996: 5A).
Potentially more important, then, is the issue of the underprovision of participation in unregulated markets. The general case for a participation mandate is supplied by Levine and Tyson (1990), who offer a prisoners' dilemma argument. They contend that were all firms to adopt participative machinery each would benefit. But participative firms require among other things compressed wage structures to encourage group cohesiveness and dismissals protection to lengthen the time horizon of workers. Unlike the participatory firm, "traditional" firms are said to motivate their workforces through the threat of unemployment and also via sharply differentiated wage structures. The scene is thus set for the emergence of a nonparticipatory equilibrium, since the viability of the single participative firm will be prejudiced by adverse selection such that it will attract the "work-shy" while losing highly-productive workers to traditional firms with a less compressed wage structure. In this way, so the argument runs, the market will be systematically biased against participatory workplaces and the economy will be locked in a suboptimal equilibrium. It is also argued by Levine and Tyson that participation works better in unionized regimes because union workers have greater job security (among other reasons), although they downplay rent-seeking insider behavior and provide little more than a caricature of "traditional" firm behavior.

Freeman and Lazear (1995), on the other hand, are alert to the rent-seeking problem. In their model, works councils both increase the joint (shareholder plus worker) surplus, at least over some range of council power, and engage in rent-seeking that reduces shareholder profitability and eventually dissipates the surplus.\footnote{In fact, the authors accept that workers will always demand more than the socially optimal level of power.} Because works councils reduce profitability, they are either not established or are given insufficient authority by management; hence an inefficient underprovision of participation absent a mandate. The sources of improved joint surplus identified by Freeman and Lazear are very much those emphasized by the collective voice model, this time underwritten by high quality information exchange and the enhanced job security made possible by a codetermination process which also inculcates in workers a longer-run view of the prospects of the firm. Ultimately, through recommending that participation be mandated, Freeman and Lazear seek to decouple pay from the factors that determine the size of the pie. This explains why they alight on the German institution as an exemplar or template. There is of course a certain irony in all of this - if the goal is to decouple issues of production and distribution, the grounds for managerial opposition to
works councils are opaque.

But at least we see here an attempt to define the content of a participation mandate. There remains the issue of whether efficient participation can be mandated, as well as the lingering ambiguity over the contribution of independent worker representation to outcomes. Advocates of participation mandates in general and German-style mandates in particular have rested their arguments on rather partial evidence concerning the efficacy of participation (on the German evidence, often ignored by works council proponents, see Addison, Kraft, and Wagner 1993; Addison, Schnabel, and Wagner 1996). As for the specific contribution of unionism to participative outcomes, the most favorable interpretation of that evidence is that participative programs "work" in both union and nonunion regimes vis-à-vis the nonparticipative nonunion firm (Cooke 1992, 1994). Any such interpretation has to be considered alongside the unfavorable dynamic effects of unions on firm performance noted earlier.

Evidence to the effect that workers would generally welcome greater involvement in their companies and that participation appears to be lower in nonunion regimes (Delaney 1996) might at first blush appear to complement recent theoretical arguments pointing to under-provision in regular markets. Even were it established that there exists a systematic market bias against participation, there is scant knowledge of the type of public policies that might encourage effective worker participation in what is a largely nonunion private sector. More troublesome is the difficulty in disentangling policies that might enhance worker participation from the rather contentious debate over the appropriate role for unions and labor law. The NLRA has undoubtedly strengthened the bargaining power of organized labor in the private sector, with net effects that may well have hastened union decline. This conclusion is of course quite consistent with the argument that certain aspects of the union decline raise legitimate grounds for concern.

Even a crude benefit-cost evaluation of the NLRA is difficult, absent a clear counterfactual. Two possible comparisons are a) the present NLRA relative to a labor market largely free of direct or indirect regulation and mandates or b) a strengthened NLRA and union sector as compared to a largely nonunion but heavily regulated labor market (i.e., one with an expanded role for direct mandates). By either standard of comparison, the NLRA (in weak or strengthened form) is arguably superior. Perhaps the more interesting comparison is of the current mix of a relatively weak but rigid NLRA, a diminishing private sector union
presence, and a growing albeit limited role for direct mandates, versus alternative regulatory and legal structures that might better facilitate worker voice and participation in union and nonunion companies. Of course, by this comparison our current system does not fare so well. What is far less certain is that achievable political pathways to a superior system of labor law and regulation can be found.

The outline of an ideal system of labor law and regulation lies well beyond the scope of this paper, and our expertise. Such a system, however, must be one that simultaneously offers workers many of the types of organizing rights and legal protections offered by current labor law, while at the same time allowing considerably greater flexibility and enhancing worker participation and cooperation at both union and nonunion workplaces.34

Conclusions

Economic theory can support the case for a number of employment mandates, at least at a general level, based on a variety of market failure arguments. Concerns about equity and worker rights often strengthen the case for state intervention in the workplace. At a more precise level, the predicted effects of mandates depend crucially on assumptions being made and the setting in which regulations are implemented. Policy debate over mandates often accords all too little attention to the milieu of considerable firm and worker heterogeneity in which policies must be implemented, as well as to the market adjustments that evolve in response to mandates. Even where a strong case can be made for a specific workplace mandate, it does not follow that the actual policies adopted and implemented via the political process will be Pareto-improving.

The role of economic analysis is potentially important. Theory often provides a good understanding of the qualitative effects of mandates and a reasonable framework for identifying benefits and costs. Empirical evidence, although suggestive, rarely provides information sufficiently specialized for policy makers. Well-designed studies providing internally compelling results need not be generalizable to the

---

34 For examples of labor law reforms that satisfy these criteria and promote "value-added" unionism, see Estreicher (1994, 1996). Levine (this volume), among others, proposes a system that would lessen direct regulation while maintaining a minimum set of labor standards for firms that voluntarily adopt alternative regulatory systems with employee oversight and approval. He would maintain the current system of standards for firms not adopting alternative systems (see, relatedly, Kochan and Osterman 1994). Levine argues that movement in this direction, while weakening workers’ de jure rights, would strengthen their rights de facto and produce net welfare gains.
policy issue at hand. Much of the evidence reviewed here suggests rather muted benefits and costs resulting from workplace mandates. The effects of mandates are mitigated in part through market escape routes, the shifting of costs, and the mobility of resources, and in part via a political process that shows some sensitivity to both benefits and costs. Politicians in recent years have been unable to devise policies that provide large benefits to their constituencies yet entail low economic costs. This of course comes as little surprise to economists, whose role in all of this has been to cast a rather skeptical eye over governmental intervention. The analytical stance of the economist confronts head on not only the enthusiasm of those who see mandates as a deliverance in a world of pervasive market failure, but also those who stubbornly defend the status quo based on equally doctrinaire grounds.

Our analysis has emphasized that workplace mandates face substantial economic limits. Older and larger programs such as workers’ compensation and UI, whose existence can be readily justified on economic grounds, in practice entail no small amount of inefficiency, arising in particular from moral hazard. There is far less evidence of substantial costs or benefits associated with recent programs such as WARN and the FMLA. We are even less sanguine of the role of the NLRA, which on the one hand serves as a less than ideal framework for what is a shrinking and rigid union labor relations system, while on the other hand either restricting or doing little to facilitate worker voice in the mostly nonunion private sector. It seems appropriate that governmental labor law and regulations should better facilitate the development of worker participation and voice. At the same time, it is important that the NLRA not be replaced with a plethora of federal mandates dictating specific terms of employment -- outcomes which might better be determined by market forces and decentralized communications and bargaining in union and nonunion workplaces.
REFERENCES


St. Louis Post-Dispatch. 1996. "Clinton Vetoes Republican Labor Measure." July 31, p. 5A.


