16 Discharge

J.H. Verkerke

1 Introduction

Discharge occurs when a firm initiates an involuntary termination of the employment relationship. This chapter focuses on economic analysis of the legal rules that govern these terminations.

We begin with the legal framework. While US law principally restricts individual discharges, in other parts of the world redundancy statutes also impose significant constraints on large-scale reductions in force. Scholars often use the term ‘employment protection law’ (EPL) or, less accurately, ‘employment protection legislation’ (also EPL) to refer collectively to any limitation on termination. It will be helpful, however, to define more carefully the various types of ‘employment protection’ with which we will be concerned.

A series of dichotomous distinctions clarifies both the source of this protection and the circumstances under which it will bind. First, we should distinguish between legally enforceable constraints and those that arise instead from industry or workplace norms. For many workers, the self-enforcing norms that govern internal labor markets provide more practically significant job protection than any of the myriad legal grounds for challenging a discharge (Rock and Wachter 1996; Wachter 2004). Second, we need to distinguish individual from collective employment rights. Countries vary widely in the relative importance of unions and collective bargaining, but even in the US, a country notoriously inhospitable to collective representation, significant segments of the workforce look principally to collective agreements for employment protection. Third, we must consider the distinction between statute and common law. Legislation provides the bulk of employment protection in Europe, for example, but US state courts have developed important common law protections that supplement the many federal, state and local statutory constraints on discharge.

We also should attend carefully to the conditions that trigger specific legal protections. The broadest possible severance pay obligation requires firms to pay workers for both voluntary and involuntary terminations (Lazear 1990). More commonly, as unemployment insurance schemes typically provide, severance is payable only to workers who suffer an involuntary termination. Many forms of employment protection also exclude terminations for which the employer has a sufficient business justification.
The standard of business justification may be quite strict, as in US unemployment insurance systems which demand proof of ‘gross misconduct’ by the worker. Or it may allow the employer to justify a discharge with evidence of mere negligence, incompetence or poor job performance. However, an even more permissive version of the business justification standard is quite common in the US and allows termination for any legitimate business reason, including lack of work or reorganization of the employer’s operation.

Yet another distinction exists between objective and subjective standards of justification. The more restrictive objective approach considers whether a reasonable person would have acted as the employer did. In contrast, a purely subjective standard inquires only whether the employer’s agent genuinely believed that he or she had a good reason for termination. Finally, many US employment laws only protect workers against termination for specific prohibited reasons such as racial or gender discrimination.

Although the legal literature on employment regulations attends carefully to these distinctions, the economic literature on the same subject often neglects legal differences or blurs important distinctions between fundamentally different types of regulation. As we will see below, this imprecision undermines the policy relevance and probity of econometric work that purports to identify how employment protection affects the labor market.

These two branches of the literature also address somewhat different questions. US legal commentators have focused on debating whether courts or legislatures should create new legal protections against unjustified discharge from employment. This ‘just cause debate’ has raged for several decades and shows no sign of abating any time soon. Accordingly, the legal discussion that follows describes the traditional doctrine of employment-at-will and explains the myriad statutory and common law exceptions that have developed in recent decades.

The economic literature relevant to our topic addresses a somewhat broader range of issues. First, many scholars have theorized about how labor regulations will affect employment, labor force participation, unemployment, productivity and other economic variables of interest. Although everyone agrees that raising firing costs must necessarily deter both discharges and new hiring, predictions for all other variables depend heavily on the structure of the model and assumptions about crucial parameters (Bentolila and Bertola 1990; Bertola 1992; Addison and Teixeira 2003). Other theorists have sought to explain why unregulated markets may produce a suboptimal level of employment protection (Levine 1991; Schwab 1993; Kim 1997; Sunstein 2001). Although each of these market failure stories is plausible, they too generate conflicting predictions about observable behavior.
In order to resolve this pervasive theoretical ambiguity, empirical economists have sought to measure levels of employment protection and investigate the correlation between that legal protection and important economic outcomes. These studies fall into two broad groups. In the first, we find cross-country comparison studies, and in the second, studies exploiting the variation in US state laws concerning unjust discharge. Both of these empirical strands of the literature have increased in technical sophistication and developed more robust approaches to estimation. As we will see, however, an inability to measure accurately the stringency of legal rules severely limits the value of these empirical results as a guide to public policy.

2 The legal framework
This section offers a quick sketch of the principal features of US law concerning discharge and some of the important ways that labor regulations differ in other countries. It also highlights significant recent changes and trends in employment protection.

A US employment and labor laws
It is conventional, and at least partly justified, to treat the US as a country in which firms have broad scope to terminate employees at any time, without notice, for any reason, or for no reason at all. This so-called rule of ‘employment-at-will’ ostensibly maximizes managerial discretion and minimizes workers’ grounds for challenging a discharge. The at-will rule prevails as the default termination provision for indefinite term employment contracts in every state but Montana. Courts there had developed an unusually intrusive unjust discharge doctrine and business groups lobbied for legislation to repeal this judicial innovation. State legislators responded with an employer-friendly termination statute that imposes a permissive good cause standard and sharply limits worker remedies (Krueger 1991; Ewing et al. 2005). With this solitary exception, US law appears to protect firms’ right to terminate and to preclude protection for workers who lose their jobs.

This picture of the US employment scene, however, is incomplete and profoundly misleading. Contrary to the conventional understanding, significant segments of the workforce receive legal protection against unjust termination. Although only 7.5 per cent of private-sector workers were union members in 2007, nearly all union members worked under a collective bargaining agreement including a provision that requires just cause for termination, and those contracts also covered an additional 0.7 per cent of workers who were not union members (US Department of Labor, Bureau of Labor Statistics 2008; Dau-Schmidt and Haley 2007). Although the usual standard of justification for terminating union workers permits
both discharge for employee misconduct and layoffs in response to slack demand, labor arbitrators typically require an objectively reasonable basis for discharge and have the authority to order reinstatement with back pay. An even larger fraction of US workers are federal, state and local government employees subject to the stringent protections of civil service systems. In 2007, about 16 per cent of the workforce had civil service protection closely resembling the union just cause standard. There is also evidence suggesting that as many as 15 per cent of nonunion private sector employers in the US have agreed to contractual provisions limiting the grounds for termination (Verkerke 1995). In addition, a much smaller fraction of employees work under definite term contracts that provide robust protection against discharge during the term, although these workers remain exposed to the risk of non-renewal. Summing these pockets of employment protection implies that approximately 34 per cent of all US workers have some form of generally applicable legal right against unjust termination.

The conventional story about US employment law also ignores a staggering array of federal and state statutory rules barring discharge for various enumerated reasons. Under federal law alone, prohibited grounds for termination include race, color, sex, national origin, religion, age, disability, union membership, collective activity in the workplace, military service and the vesting of pension benefits. Federal law also protects against termination in retaliation for filing a claim or assisting in an investigation under any of more than a dozen employment statutes and it extends limited protection to several categories of whistleblowers. Some state laws, and even municipal ordinances, expand the list of prohibited reasons for termination to include categories such as sexual orientation, marital status, political affiliation, participation in lawful off-duty activities, and weight. Judicial decisions have also applied these legislative prohibitions to employer harassment on the basis of an otherwise protected characteristic.

Finally, US employment law now incorporates state common law doctrines that further limit firms’ discretion to terminate workers. A considerable body of empirical work relies on a tripartite classification of the relevant cases into (1) public policy claims, (2) implied contract claims, and (3) good faith claims. The first category encompasses tort cases alleging wrongful discharge in violation of public policy. The paradigmatic public policy claim arises when a firm fires a worker for refusing to commit an unlawful act such as perjury. One unifying feature of these cases is that courts are most likely to recognize a cause of action when the employer’s conduct has significant third party effects. The discharge must cause harm to the general public rather than merely affecting the worker and the firm involved in the dispute (Schwab 1996). In contrast, implied contract claims arise whenever an employer’s words or actions would cause a reasonable
worker to believe that he or she has protection against unjust discharge. The majority of these claims are based on employee handbooks or policy manuals that discuss progressive discipline and enumerate permissible grounds for termination. In the overwhelming majority of jurisdictions, however, employers can avoid forming an implied contract by using a clear and prominent disclaimer (Verkerke 1995, 1998). Finally, a decided minority of US jurisdictions apply the traditional contract doctrine of good faith and fair dealing to the employment relationship. The most widely accepted version of this claim prohibits an employer from using discharge to avoid paying already-earned compensation.

We will examine below several problems with the conventional approach to classifying these doctrines, but our initial introduction to US employment law rules also requires a brief discussion of remedies. Claimants alleging the tort of wrongful discharge in violation of public policy have the largest potential recovery. They may seek the full panoply of tort remedies including compensatory damages – for economic losses and emotional distress – as well as punitive damages designed to deter reckless or willful misconduct. The contractual remedies available for implied contract and good faith claims are far more limited. In these cases, damage rules allow recovery for foreseeable economic losses such as lost wages, but exclude both emotional distress and punitive awards. Federal and state statutory provisions also vary widely in the generosity of their remedies. Some permit plaintiffs to recover full tort damages comparable to the public policy claim discussed above. Others, however, impose limitations or caps on the amount or types of allowable damages. A few of the most limited statutes severely restrict the duration of back pay awards and thus offer an even less generous recovery than plaintiffs receive for contractual claims.

Many statutes allow courts to order employers to reinstate the claimant to his or her original job. Such an order, however, leaves the reinstated employee in a potentially hostile work environment and gives the employer another chance to document a record of poor performance that may be sufficient to justify a subsequent discharge. Parties thus typically settle wrongful termination claims on terms that omit reinstatement. Instead, employers make monetary payments to claimants and often negotiate the terms of the letter of reference that they will provide to prospective employers. In return, the discharged worker agrees to keep the terms of the settlement confidential, to refrain from disparaging the former employer, and to release all legal claims against the firm. Settlements of this sort occur in the overwhelming majority of discharge cases that are not dismissed at an early stage of the proceedings. As a result, it is entirely accurate to think of US employment laws as imposing a tax on discharge equal to the expected value of the payment from the firm to the worker.
B International comparisons

A comprehensive survey of discharge protection around the world is far beyond the scope of this chapter, but a more impressionistic account will highlight important differences between the US and other legal regimes.

In broad outline, US law superimposes specific legal restrictions on a comparatively permissive legal background. In contrast, most other countries begin with general statutory protection against unjust termination. These legal systems then qualify this more restrictive background rule with exceptions for small employers, recently hired workers and fixed-term contracts, all of which increase employment flexibility and mitigate potential labor market rigidities. Large numbers of casual, temporary or guest workers thus bear the brunt of employment adjustment without legal protection from discharge. Although the conventional impression is surely true that legal constraints on discharge are more severe outside the US, firing is far more costly in the US and cheaper elsewhere than contemporary commentary would lead a casual reader to believe.

The literature includes many excellent discussions of employment protection in individual countries. An even larger body of work categorizes and classifies legal provisions in an effort to develop cross-country comparisons (World Bank 2007; Heckman and Pagés-Serra 2000). Typical restrictions on termination include general statutory provisions that require just cause for discharge and mandate severance pay proportional to a worker’s length of service. Severance statutes sometimes differentiate between discharge with and without cause, imposing higher payments for the latter. Redundancy statutes commonly require payments to workers terminated for economic reasons and those payments rise proportionally with the worker’s job tenure. Procedural rules differ widely, but most nations have a specialized labor tribunal that hears cases challenging employment termination. In many countries, rates of union membership far exceed those in the US and thus unions play a more important role in representing discharged workers.

Despite the existence of these muscular discharge regulations, many workers fall outside of the protected group. Exceptions differ widely among countries, but the most important sources of uncovered employment are fixed term contracts, temporary and casual workers, probationary periods, guest worker programs and employees of small firms. Although most countries impose legal restrictions on using fixed term or temporary contracts to evade discharge protections, a significant fraction of the workforce invariably finds employment in the uncovered sector of the economy. It seems likely that this flexibility mitigates somewhat the economic impact of strict statutory employment protection.

In addition, remedies for unjust discharge are ordinarily far less
generous in other countries than in the US. Specialized labor tribunals tend to award much smaller amounts of back pay and compensatory damages. As Estreicher (1985) concludes, ‘[o]utside of Canada and possibly Italy, the monetary awards, while not insignificant, are at modest, predictable levels – certainly in comparison to American jury recoveries’. The same author also observes that in the overwhelming majority of jurisdictions he surveyed, the remedy of reinstatement has little practical importance. As we have already seen, it is difficult to ensure that a reinstated worker will not suffer harassment or be subjected to close supervisory attention designed to discover grounds for justifying another discharge.

Our brief comparison of employment protection in the US and other countries thus reveals significant institutional and formal differences but also some important practical similarities. Job protection in the US arises from the accretion of many individual and comparatively specific employment rights that lack a single organizing principle or unified jurisdictional authority. However, employers who violate these prohibitions must pay relatively large damage awards and high litigation costs. In contrast, most other countries grant workers a general right against termination, but this ostensibly universal protection provides less generous remedies and leaves a significant segment of the workforce outside of its coverage. Thus, the aggregate burden of employment protection law in the US and elsewhere is more similar than the conventional wisdom imagines. The US imposes larger costs less frequently while most other nations make comparatively more modest costs an expected consequence of most employment terminations.

C Trends and legal change
Proposals for legal reform are common. In the US, a significant body of commentary criticizes the at-will rule and seeks greater employment protection. Although these advocates for wholesale change have been thus far unsuccessful, courts and legislatures have created many new employment rights during recent decades. In contrast, most reform proposals outside of the US have involved efforts to reduce firing costs by removing legal constraints on discharge. Legal change in these countries has taken several distinct forms. In most cases, however, the goal has been to encourage hiring and diminish the high levels of structural unemployment that economists often associate with labor market rigidity.

Reforms in Spain, for example, exempted smaller employers from statutory requirements and made it easier for employers to create contracts that are less costly to terminate. During the early 1990s, a wave of reforms swept through many countries in Latin America. Heckman and Pagés-Serra (2000) provide an excellent summary of these developments
and investigate their employment effects. In 2006, the Australian federal government enacted a far-reaching reform called Work Choices (Freyens and Oslington 2007). This legislation provided for at-will dismissal of workers employed at firms with fewer than 100 employees, permitted termination for ‘operational reasons’ at all firms, and made dismissal terms a ‘prohibited topic’ of bargaining for unions at the smaller firms. Although the recently elected Labour government has vowed to restore many of the protections that Work Choices eliminated, as enacted, the statute arguably created an at-will regime more extreme than US law provides.

Naturally, not all reform movements are successful and some have even tightened employment protection rules. In 2006, the French national government tried unsuccessfully to relax the rules governing termination for young people’s first jobs. Although the goal of these reforms was to improve employment prospects for a group that suffers chronically high unemployment, the proposal caused strikes and civil unrest while attracting considerable international attention. Marinescu (2007) studies an amendment to UK law that shortened probationary periods from two years to one year, thus expanding the coverage of that country’s discharge protections.

In contrast, politicians have discussed liberalizing probationary periods in Germany and elsewhere in Europe. Moreover, the World Bank has become a consistent advocate for liberalized termination rules. The group’s Doing Business project measures the stringency of various regulations and argues that greater flexibility will improve economic outcomes (World Bank 2007). Despite some contrary currents, recent history suggests that the most likely direction for reform outside the US involves piecemeal liberalization of existing employment protections.

Countless US legal scholars have advocated for legislation that would replace the at-will rule with a more general requirement of just cause for employment termination. Despite vigorous calls for wholesale reform, US employment regulation has instead evolved far more slowly towards greater protection against discharge. Congress enacted major antidiscrimination statutes in 1964, 1967, 1990 and 1991. Although some state employment discrimination statutes predate these federal laws, many have been amended or expanded in recent decades. State common law employment protections developed through a similarly gradual process of accretion. Beginning roughly in the mid-1970s, increasing numbers of state courts began to entertain the public policy, implied contract and good faith claims that we examined above. As these exceptions to the traditional at-will doctrine spread across the country, many employers discovered new constraints on their discretion to terminate employees. Dau-Schmidt and Haley (2007) provide an excellent summary of these legal developments.
and Morriss (1994, 1995) studies how to date the adoption of each legal innovation.

3 Theoretical ambiguity
Orthodox economic theory warns that legally imposed employment protection, like minimum wages or payroll taxes, will distort an efficient labor market equilibrium and thus reduce social welfare. Thus, only by identifying persistent market failures can we justify government intervention of this sort. Law and economics scholars have risen to the challenge and proposed a panoply of reasons that labor markets may not provide the socially optimal degree of protection against discharge. This section presents a concise overview of this theoretical literature. Unsurprisingly, we will discover that economic theory generates profoundly ambiguous predictions about how employment protection will affect observable labor market outcomes and social welfare. For an analysis of this discussion in the context of minimum wages, see the chapter by Simon Deakin and Frank Wilkinson in this volume (Chapter 5).

A Predicted effects
In what is by far the most frequently cited paper on the economics of employment protection, Lazear (1990) offers a simple static model of wage determination. He investigates what happens when the government requires employers to pay an amount Q to each worker hired in an initial period and then terminated in any subsequent period. In order to maintain an efficient level of employment, Lazear shows that firms must increase the wage from the unregulated equilibrium value $W^*$ by exactly the amount of the severance payment Q. This adjustment ensures ex post efficiency. In the new equilibrium, workers receive Q with certainty – either as part of their wage or as a severance payment. Both workers and firms thus face a marginal payment of $W^*$ when deciding whether to work or employ labor. As Lazear observes, the same workers will find that $W^*$ exceeds their reservation wage and the same firms will find that the marginal revenue product of labor exceeds $W^*$ as in the original equilibrium.

A more difficult problem for this new regulated equilibrium is how to ensure that firms will be willing to hire workers in the initial period. Firms operating under the severance mandate confront total wage payments that exceed the original equilibrium wage by Q in each period of employment. Lazear cleverly solves this problem by suggesting that workers can pay an upfront fee for the privilege of signing an employment contract. This fee equals a weighted sum of Q times the probability of remaining employed in each subsequent period. Since the fee offsets their expected wage bill in subsequent periods, firms are once again willing to
enter employment contracts. Lazear summarizes the argument to this point by observing that ‘[a]ny severance pay arrangement can be offset by an optimal contract that should evolve in a competitive labor market’ (Lazear 1990, p. 702).

It is this conclusion for which Lazear’s paper is most often cited, even more than for the empirical results that we will examine later. However, as Lazear himself acknowledges, this Coasean invariance result is far from robust. If workers are unable to borrow against their future wage income, they will be unable to make the necessary upfront payment for a job. In addition, workers must receive and firms must pay exactly the same amount of severance without any deduction by third parties or cross-subsidy. Imperfectly experience-rated unemployment compensation, for example, breaks this necessary correspondence and produces an inefficient equilibrium. Lazear also suggests that workers may worry about the possibility that the firm will abscond with the upfront payment ‘by declaring bankruptcy’. And, finally, he alludes somewhat vaguely to ‘[o]ther strategic considerations’ that ‘may apply’ (Lazear 1990, p. 704).

To these acknowledged problems, we must add a more fundamental failure to specify with sufficient precision the terms of the employment contract to which the parties agree in the initial period. It is clear from Lazear’s discussion that signing a contract entitles workers to a severance payment $Q$ in the event that they are laid off. Beyond that fact, however, the parties’ obligations are unknown. Does the employer have absolute discretion to lay off workers at will? Or does the contract require employers to justify each discharge with evidence that a worker’s marginal productivity has fallen below the equilibrium wage? These are questions that any useful analysis of employment protection must confront.

If termination at will is permitted, then the Coasean bargain for which the paper is so often cited becomes unsustainable. Firms maximize their returns by receiving the upfront payment from workers and then immediately terminating the contract at a cost of the severance payment $Q$. Anticipating that they will never have a chance to recoup the initial fee by receiving wages in subsequent periods, workers refuse to pay, and the supposed equilibrium collapses. But consider also the fact that Lazear calculates the workers’ upfront fee using a sequence of fixed probabilities for the continuation of employment in the next period. We should perhaps infer from this feature of the model that the separation probability is exogenously determined and thus employers have no discretion about termination. However, such a rule bears no meaningful relationship to real world employment contracts. A regime of fixed continuation probabilities captures neither the general protections of European discharge laws nor the more specific exceptions to the US at-will rule. Both of these legal
systems would permit employers to terminate workers at some cost \( Q \), thus eliminating any equilibrium in which workers pay an upfront fee.

It is doubly ironic that Lazear’s paper is so often cited for its Coasean invariance result. As we have seen, the peculiar contract that ‘offsets’ a severance pay mandate confronts both practical and legal obstacles that far exceed the transaction cost barriers to conventional Coasean bargaining. Indeed, Lazear himself acknowledges at least some of these problems. He treats the invariance point as a theoretical curiosity. Like Coase in his justly famous article on social costs (Coase 1960), Lazear focuses instead on analyzing the ‘somewhat imperfect’ world in which severance payments have real economic effects.

More realistic theoretical analyses of employment protection have also yielded fundamentally ambiguous predictions. To begin, we must expand our analysis beyond severance pay obligations to incorporate other forms of labor regulation. As we have seen, both legislation and court decisions significantly constrain an employer’s discharge decision. Some of these legal rules impose liability that depends on the absence of a good cause for termination and others condition liability on proof of a specific bad motivation for discharge. Still other statutes restrict the use of temporary and fixed-term contracts in order to protect the rights granted to permanent employees. Despite these differences, it is convenient to model all regulatory constraints as an increase in expected firing costs. We will discover below that legal subtleties dramatically affect empirical efforts to measure legal variables, but firing costs work reasonably well for theoretical analysis.

Legally mandated employment protection creates two competing influences on the level of employment. Most obviously, the mandate acts as a tax on discharge and thus reduces the rate of employment separation. Anticipating this additional firing cost, however, employers also reduce hiring. Thus, the net effect on employment is uncertain. If these constraints on firing and hiring increase firms’ total labor costs, then we should expect reduced employment levels (Hamermesh 1993). However, even this seemingly obvious conclusion depends on assumptions about the shape of labor demand (Bertola 1992; Bentolila and Bertola 1990). For example, if the marginal revenue product of labor curve is relatively flat when labor demand is high and comparatively steep when labor demand is low, then average employment across the business cycle will fall as firing costs rise. The intuition behind this result is straightforward. During a boom, increased firing costs will deter a given fraction of the many new hires required to equalize wages and marginal productivity. During a recession, however, the steep marginal revenue product curve implies that firing costs will affect only the comparatively small number of layoffs needed to restore
the labor market equilibrium. Increased firing costs, under these assumptions, deter new hiring more than they prevent layoffs, and thus average employment falls. Conversely, average employment rises if we make the opposite assumptions about the shape of the marginal revenue product of labor.

Similar reasoning suggests that the likely impact of imposing a new legal restriction on termination varies according to the phase of the business cycle (Hamermesh 1993). In an economic expansion, firing costs affect hiring more than they limit terminations and thus we should observe lesser increases in employment levels than would occur in an unregulated environment. In an economic downturn, the constraint on discharge binds more tightly and there is comparatively little hiring to be deterred. Thus employment levels will rise relative to the unregulated state.

The analysis of wages and efficiency implicates similar complications and uncertainty. If workers value discharge protection at more than it costs firms to provide that job security, then wage rates should fall, social welfare will increase, and workers and firms will share the resulting surplus. However, if workers remain unaware of their enhanced job security or if they assign it no value, then labor supply will be unchanged, employers’ costs will rise, and the legal mandate will reduce labor market efficiency. In short, our only unambiguous theoretical prediction is that effective employment protection reduces labor turnover.

B Market failures

It may be difficult to predict how employment protection will influence employment levels and wages, but most economists also seek convincing evidence of a market failure to justify regulatory intervention. What then prevents workers from demanding and firms from offering the optimal contract terms governing discharge? On this question, scholars have been enormously creative in suggesting reasons that an unregulated labor market might produce suboptimal levels of employment protection.

Among legal commentators and would-be reformers, advocates of just cause protection often invoke inequality of bargaining power between employers and workers. This theory contends that employers are particularly prone to abuse their greater bargaining power. They will refuse to grant workers protection against discharge even though these employees would gladly pay the cost of that protection in reduced wages. From an economic perspective, however, this argument, standing alone, verges on incoherence. It offers no explanation for why employers would choose to exploit their bargaining power by offering a suboptimal contract rather than by combining the optimal contract with a lower wage offer. Thus, some other obstacle to efficient contracting must account for employers’
unwillingness to provide the job security terms that maximize workers’ welfare.

Suppose, for example, that most workers know very little about background legal rules concerning termination or about the specific terms of their employment contracts. Their legal ignorance might well prevent them from negotiating for an optimal level of employment protection. Indeed, Kim (1997) finds that many workers erroneously believe the law protects them from discharge for any reason short of serious misconduct or a plant closing. Such a mistaken belief could deter workers from seeking job security terms or lead them to undervalue or ignore protective terms offered by employers (Sunstein 2001). At the very least, this problem argues for an information-forcing default rule of just cause. If workers’ legal ignorance persists, however, then lawmakers must decide whether mandatory employment protection terms will enhance efficiency (Verkerke 1995, 1998, 2003, 2007).

Another informational argument focuses instead on the problems that firms face in screening prospective employees. Applicants ordinarily have more information about their productive characteristics than prospective employers can gather before making an employment offer. This informational asymmetry creates a danger of adverse selection. Contractual job security distinguishes only imperfectly between just and unjust discharges. It thus provides some insurance against job loss for low productivity workers. Fearing that they will attract a disproportionate number of shirkers, firms may be unwilling to offer just cause protection at all. Government-mandated employment protection thus could be justified whenever adverse selection prevents firms from offering the socially optimal job security terms (Levine 1991).

Related problems of moral hazard and adverse selection impede the provision of insurance against the risk of job loss. Bertola (1992) models employment protection as a form of second-best insurance for risk-averse workers. He assumes quite plausibly that firms are risk neutral and workers are risk averse. It is also apparent that the consequences of employment termination for workers can be quite severe. However, market insurance to protect against job loss is largely unavailable. A legal obligation for firms to make a payment to workers contingent on termination thus provides partial insurance and could improve labor market efficiency.

Employment protection also could be underprovided because it exhibits the characteristics of a local public good. Perhaps workers will disclaim any interest in job security and hope to free-ride on the willingness of other employees to accept lower wages in return for stronger protection against discharge. However, this public good argument turns on several assumptions of varying degrees of plausibility. First, it must be impractical for
employers to offer individualized contract terms. Although individually negotiated contracts would be undeniably costly, employers might easily offer a standard menu of job security terms with appropriate compensating wage differentials. Second, employers must have no incentive or opportunity to discover workers’ preferences concerning job security. But firms compete in a labor market for prospective employees and surely have ample opportunities to experiment with terms and imitate successful recruitment strategies. And finally, job security must be nonrivalrous in consumption. Contrary to this assumption, however, legal protection against termination imposes additional expected costs for each employee hired.

Finally, Schwab (1993) offers an entirely different public good argument. He suggests that legal enforcement increases the value of an implicit agreement between workers and firms for career employment. Building on both efficiency wage and human capital models, Schwab argues that employers are especially tempted to renege on their implicit commitment to workers during the late career period when workers receive a wage in excess of their marginal productivity. Judicial enforcement of a ‘life-cycle just cause default’ thus could protect late-career employees from opportunistic termination. Schwab also proposes protection for newly hired workers who may have incurred substantial moving expenses or given up other opportunities to accept a firm’s job offer. Just cause protection for recent hires and late-career workers might enhance employment relationships, but such a rule is comparatively complex and difficult to specify. Thus, Schwab argues that courts should supply this doctrine as a form of public good. However, Verkerke (1995) questions whether life-cycle just cause is a workable standard and challenges Schwab’s assertion that leading cases follow the prescribed pattern of protection.

4 Empirical investigation

We have seen that theoretical arguments produce uncertain predictions about how employment protection will affect the labor market. Likewise, many potential market failures could conceivably justify regulatory intervention. It is difficult to know, however, whether legally mandated employment protection will improve or harm efficiency. Our best hope is therefore to investigate these effects empirically and to determine whether patterns of observable economic behavior shed any light on the plausibility of various market failures.

A International comparative perspective

The earliest efforts to understand how variations in employment protection correlate with important economic outcomes relied on cross-country comparisons.
(i) Early work Lazear (1990) inaugurated this field of inquiry with an early and influential effort to use international comparisons to reveal how employment protection affects the labor market. The paper’s empirical section investigates firing costs with variables that measure legal requirements for severance pay and for notice before termination. Although nominally a panel dataset, most countries altered their notice and severance regulations at most once during the sample period. Thus, as Lazear acknowledges, his results arise principally from cross-sectional rather than time-series variation in the legal variables. He reports that more generous severance pay requirements increase unemployment rates and reduce employment levels, labor force participation and hours worked. In an exercise designed to correlate changes in unemployment with changes in severance pay, Lazear also argues that increased firing costs explain a substantial portion of the increased unemployment in several European countries.

In the years since its original publication, Lazear’s paper has spawned a burgeoning literature of cross-country studies that explore diverse theoretical and empirical speculations. The results of these analyses have been decidedly mixed. Readers interested in a detailed technical exploration of this work will find a superb review in Addison and Teixeira (2003). A few high points from this early literature follow.

The empirical analysis in Lazear’s seminal paper has fared rather poorly under subsequent scrutiny. Addison and Grosso (1996) identify and correct many errors in both the dependent variables and covariates – such as using daily rather than weekly figures to code the hours-worked variable for Italy and misstating the level of severance pay obligations in France, Italy, Norway and Spain. Their revised estimates confirm Lazear’s results for employment, labor force participation and unemployment rates, but contradict his finding that severance pay reduced hours worked. Using corrected data, they also find that longer notice requirements are associated with lower unemployment rates, longer hours, and higher employment and labor force participation – exactly the opposite of Lazear’s findings for notice. More significantly, Addison and Grosso’s corrected calculations reveal that changes in severance pay explain, at best, a far smaller portion of changes in unemployment rates than Lazear reported. In a more comprehensive challenge, Addison et al. (2000) replicate and then critique the econometric methods employed in Lazear’s paper. They explore alternate specifications that include fixed and random effects. They detect and use several different techniques to correct for the presence of serially correlated residuals. They investigate the role of influential observations and conduct other sensitivity analyses. Finally, they analyze a dynamic version of the Lazear model using a generalized method of moments estimator.
These more refined estimation procedures consistently fail to confirm Lazear’s findings that increased severance pay has adverse labor market consequences. After correction for auto-correlation, the severance variable almost never approaches statistical significance and this finding of no effect persists across many alternative specifications.

No one should conclude from these essentially negative results, however, that legally mandated employment protection promotes efficiency. All of these studies use narrow measures of employment protection that ignore other important constraints on discharge. For example, some countries with no severance pay requirements effectively barred individual terminations during part of the sample period. A more complete picture of discharge policies surely would account for procedural obstacles, unfair dismissal standards, limitations on temporary contracts and the overall flexibility of working time (Grubb and Wells 1993). Moreover, conflicting results for severance and notice requirements suggest that these and other legal obligations may influence behavior through subtle and indirect paths (Addison and Grosso 1996).

Many later studies using better econometric methods and diverse sources of data find statistically significant negative effects. For example, Heckman and Pagés-Serra (2000) use a sample of 20 Latin American and Caribbean nations along with 16 largely European countries. They derive a slightly broader index of employment protection that incorporates legal requirements for notice and severance pay measured across the entire job tenure distribution. Despite reforms in the 1990s that substantially reduced regulatory burdens for five nations, the developing countries of Latin America continued to have expected dismissal costs roughly double those of the industrialized countries in the sample. Heckman and Pagés-Serra find that job security has a large negative effect on employment rates. This effect is statistically significant for pooled ordinary least squares and random effects estimates but not significant in fixed effects specifications. The impact on young workers is especially large. The estimated elasticities suggest that job security mandates reduce youth employment rates in Latin America by nearly 10 percentage points. Studies of different groups of countries find similar, though often less dramatic, effects (Grubb and Wells 1993; Scarpetta 1996; Nickell 2007; Di Tella and MacCulloch 2005; Nicoletti and Scarpetta 2003).

Despite the broadly consistent finding that more stringent job security regulations depress employment, critics worry that existing indicators of employment protection are inadequate. Errors interpreting legal requirements are quite common. The effectiveness of enforcement varies widely among industrialized countries and even more so in developing nations. Although they are more easily observable, legislative severance
and notice requirements interact in potentially complex ways with other employment regulations. Efforts to quantify these other legal constraints invariably require researchers to use rather arbitrary criteria to produce a tractable index. The World Bank’s remarkably ambitious effort to quantify hundreds of legal provisions in more than 60 countries, entitled Doing Business, aptly illustrates this concern (Davis and Kruse 2007). The project’s quantitative measures necessarily abstract from many of the complexities and uncertainties in real world legal practice (World Bank 2007). Like Doing Business, most international comparative projects focus exclusively on employment legislation and thus understate regulation in common law countries.

An alternative approach relies on surveys of employer attitudes to measure regulatory flexibility (Pierre and Scarpetta 2004, 2006), but these values appear to be inherently unstable (Addison and Teixeira 2003). Most business people rely on a local perspective to form expectations about legal constraints. These expectations and perceptions thus are likely to vary systematically across nations and phases of the business cycle.

The empirical literature has at least produced a consensus that employment protection significantly reduces flows into and out of unemployment (Addison and Teixeira 2003). How this flow effect influences short-run and long-run employment levels remains a subject open to debate. In a particularly creative effort to move beyond the narrow causal model of most other work, Blanchard and Wolfers (2000) contend that high levels of unemployment in Europe arise from the interaction of employment protections and adverse economic shocks. They observe that shocks to productivity, interest rates and labor demand are too uniform to explain the variation in unemployment levels among countries. Moreover, restrictive labor regulations existed in many countries before their unemployment levels rose in recent decades. Blanchard and Wolfers propose that restrictive labor market policies impede dynamic adjustment to economic shocks. Unemployment thus increases most in those countries whose employment institutions prevent firms from reallocating workers to more productive positions. Despite severe limitations in their data on employment institutions, they find consistent support for the importance of interaction effects. Although Blanchard and Wolfers characterize their results as preliminary, the paper is a welcome movement towards a more nuanced approach to modeling how employment protection may affect unemployment rates.

(ii) More recent investigations In recent years, the most distinctive developments in the empirical literature have been the rise of studies focused on regulatory changes in a single country and a nascent body of work using firm-level data.
Pierre and Scarpetta (2004, 2006), for example, present conventional cross-country comparisons but exploit micro-data on firms rather than national aggregate figures. They find that the stringency of employment protection correlates well with employer reports of regulatory inflexibility. In response to these legal restrictions, firms appear to train workers more intensively and to rely more frequently on temporary employment contracts. Gomez-Salvador et al. (2004) similarly use firm-level data across a sample of 13 countries and find that employment protection reduces job flows.

Despite reliance on firm-level data, concerns remain about such cross-country studies. Omitted variables may confound any effort to correlate national legal systems with economic outcomes or behavior. The possible endogeneity of employment regulations also could interfere with our ability to untangle the true relationships among these variables. Although the problem of endogeneity remains, a number of studies use panel data to investigate the impact of specific legal reforms in a single country. These quasi-natural experiments narrow the range of possible confounding variables and often permit the use of a difference-in-differences approach that requires comparatively weak identifying assumptions.

Kugler and Pica (2008) use an unusually detailed employer-employee panel dataset to study how firing costs affect worker and job flows. In 1990, a legislative amendment expanded the coverage of Italy’s strict employment protections to businesses with fewer than 15 employees. The authors find that this increase in dismissal costs reduced flows into and out of employment by 13 to 15 per cent for small establishments. They also estimate a significant decline in the entry rate for small firms, but no effect on exit rates. Related work by Leonardi and Pica (2007) finds that the same reform reduced entry wages by about 10 per cent and increased the steepness of the tenure-wage profile.

Marinescu (2007) examines a 1999 reform of UK job security law that reduced from two years to just one year the length of service required before workers may sue their employer for unfair dismissal. Marinescu finds a peak in the firing hazard rate just before the end of the probationary period and a trough immediately after it ends. She also observes that, after the reform, newly covered workers and even those with lower job tenure experienced a reduced firing hazard. Firms appear to have increased both recruitment efforts and training in response to the reform. In this case, stricter employment protection had no negative effect on employment or unemployment duration.

Earlier work by Hunt (2000) uses industry-level data to examine how firms responded to reforms in Germany that reduced firing costs and increased the flexibility of working hours. She finds no evidence that these
reforms, which were intended to lower firing costs, speeded employment adjustment. The best explanation for this finding of no effect may be that the amendments were too limited to make a practical difference for employers. Hunt also finds, however, that provisions allowing more flexible hours slightly slowed adjustment in the number of workers, lending some support to models in which those margins of adjustment are substitutes.

Dolado et al. (2007) focus on a 1997 Spanish labor law reform that created a new type of permanent contract with reduced firing costs for a targeted group of youthful workers, long-term and older unemployed, disabled individuals and workers moving from a temporary to a permanent contract. They develop a search and matching model to capture the effects of employment protection on heterogeneous groups of workers and then calibrate their model to the characteristics of the Spanish labor market. Simulation exercises conducted on the calibrated model suggest that targeting reforms on low productivity workers and on jobs subject to frequent productivity shocks may be the most effective strategy for reducing aggregate unemployment.

Behaghel et al. (2008) study a French law that imposes a special tax for laying off workers older than 50. When a 1992 reform created a new exemption from this tax, the transition rate from unemployment to employment for workers over 50 as compared to workers under 50 rose dramatically. The authors conclude that the targeted employment protection for workers over 50 caused significant substitution towards younger workers and thus harmed the ostensibly protected group. Reducing the tax helped these older workers.

Finally, there has also been interesting work on reform measures outside of Europe. Freyens and Oslington (2007) use an extensive survey of small- and medium-size Australian firms to calculate monetary values for the firing costs that these firms incur. They also compute an estimate of the number of jobs that the provision of the Work Choices reform that exempts firms with fewer than 100 employees will create. Their estimate of approximately 6000 jobs gained falls considerably short of the 77 000 figure advanced by government proponents of the reform.

B Identifying the effects of US state laws
In addition to the growing body of work focused on legislative changes in individual countries around the world, an increasingly important strand of the empirical literature uses legal variations among US states to identify the effects of employment protection law. We saw earlier that US labor regulations impose significant and nationally uniform constraints on discharge. However, the public policy, implied contract and good faith exceptions to the at-will doctrine are all creatures of state law. Courts in each
state decide whether to recognize these claims and how liberally to interpret them. Many scholars have exploited this rich source of cross-sectional and time-series variation to estimate how these judicial innovations affect the labor market.

Early work by Dertouzos and Karoly (1992, 1993) finds notably large employment effects. Using a two-stage quasi-instrumental variables procedure, they estimate that states adopting tort exceptions suffer a 2.9 per cent decline in aggregate employment. Recognizing contract exceptions as well imposes an additional employment loss of 1.8 per cent. These estimates suggest that an activist judiciary may cost a state nearly 5 per cent of its jobs solely by eroding an employer’s right to terminate at-will. Although these findings received considerable attention and publicity, subsequent studies have identified smaller effects or none at all. Miles (2000), for example, uses a difference-in-differences estimator and finds that adopting the at-will exceptions has no statistically significant impact on either employment or unemployment. Firms, however, increase their use of temporary help services by 15 per cent when a state has adopted the implied contract doctrine. Autor (2003) confirms this relationship and finds that adoption of the implied contract exception can explain as much as 20 per cent of the dramatic increase in employment outsourcing during the sample period.

Two recent papers by Autor et al. (2004, 2006) break new ground in understanding the impact of at-will exceptions on aggregate employment. One of these papers (Autor et al. 2006) presents estimates that contrast with Miles’s finding of no effect, but nevertheless fall far short of the values that Dertouzos and Karoly reported. Autor et al., like Miles, use a difference-in-differences estimation framework. Their central finding is that the employment-to-population ratio falls by 0.8 to 1.6 per cent in states that recognize the implied contract doctrine. This negative effect is robust to alternate specifications, including state fixed effects, time and region dummies, and a rich set of interactions among the covariates. Their results also contrast starkly with the finding in Dertouzos and Karoly that tort claims affect employment most severely. Neither the public policy nor the good faith doctrine exerts a consistent and statistically significant influence on employment or job flows.

The earlier of the two papers (Autor et al. 2004) offers an insightful discussion of the marked contrast between their results and those reported in the prior literature. Most of the difference arises from a flawed instrumental variables estimation technique that Dertouzos and Karoly used to overcome the possible endogeneity of legal changes. Unfortunately, among their chosen instruments are variables that likely have an independent employment effect that is unrelated to the adoption of at-will exceptions. This direct correlation with the dependent variable implies that the
estimated coefficients for the instrumented legal variables will be biased upward. In addition, their use of many continuously varying instruments creates a danger that the instrumented variables are merely capturing a correlation with other smooth trends in the dependent variable. Adding state time trends to their model, for example, dramatically reduces the estimated impact of state laws. Carefully deployed, the difference-in-differences estimator eliminates these biases.

Explaining the discrepancy between Miles’s finding of no effect and the modest but robust negative effect that Autor et al. observe requires a more subtle inquiry into the methods researchers use to code legal policy variables. As a matter of theory, we are interested in measuring what happens when a jurisdiction ‘recognizes’ each new exception to the traditional at-will rule. How should we define that event? The conventional approach requires the researcher to identify judicial decisions that accept each of the three categories of unjust dismissal doctrines. But even this task remains ambiguous. US state court systems include trial courts, intermediate appellate courts, and a state supreme court. Federal courts also decide employment cases under their diversity jurisdiction and there are notable cases in which federal district courts have either anticipated or impeded legal developments in a state.

Additional complexity arises from the episodic and evolutionary quality of the common law process. We could mark ‘adoption’ with the first time that a state court indicates receptivity to a particular theory of liability. Alternatively, we might require more. Perhaps we should wait until a court has expressly endorsed a new doctrine. Or we could mark adoption with a decision that unequivocally applies the doctrine to the parties before the court and thoroughly explains the rationale for the new rule.

In an exhaustive effort to replicate prior results, Autor et al. (2004) re-estimate their model using three different classification schemes—Dertouzos and Karoly (1992, 1993), Morriss (1995) and Walsh and Schwarz (1996). Although Dertouzos and Karoly, and Morriss, rely on legal classifications that differ in some particulars, using their legal variables confirms Autor et al.’s principal finding that only the implied contract exception has a significant negative effect on employment. However, Miles used the Walsh and Schwarz classification, and re-estimating the Autor et al. model with this scheme replicates Miles’s finding of no statistically significant effects. As the authors explain, Walsh and Schwarz select cases at least in part because they best articulate a court’s rationale for adopting a new doctrine. In contrast, Autor et al. choose to focus instead on the first sign of receptivity to a new theory. Miles’s reliance on Walsh and Schwarz implies that he sometimes dates legal changes too late after firms have already incorporated the expected costs of the new doctrine into their hiring decisions. For
this reason, the Autor et al. approach more accurately models the process of legal change.

While we should admire the efforts of Autor et al. to attend carefully to the differences among legal classifications and to construct legal variables that measure more accurately the moment at which a new doctrine is likely to affect employers’ behavior, even their diligent implementation of the conventional approach to measuring US law does a poor job of capturing the main thrust of state common law surrounding discharge. Walsh and Schwarz’s classification of doctrines distinguishes ‘narrower’ and ‘broader’ versions of each of the three exceptions. Although this approach is more nuanced, it still fails to reflect the essential differences among jurisdictions. Dau-Schmidt and Haley (2007) provide an excellent discussion of the nuances of various at-will exceptions. We can most meaningfully classify state courts’ receptiveness to employment contract claims into liberal, centrist and conservative jurisdictions (Verkerke 1995). Dramatic differences exist among jurisdictions that all ostensibly recognize a particular theory of employment protection. Accordingly, the conventional doctrinal classification measures the practical significance of these legal rules with substantial error.

Other work looks beyond aggregate employment effects to consider how the adoption of at-will exceptions might influence labor market dynamics. Kugler and Saint-Paul (2004) investigate whether employment protection might have differential effects on employed and unemployed job seekers. They find that the adoption of at-will exceptions reduced re-employment probabilities for unemployed but not for employed workers. Schanzenbach (2003) focuses on how the exceptions affect job tenure and the tenure-wage profile. His results are broadly consistent with prior work in finding that neither public policy nor good faith doctrine had a robust statistically significant effect. Although the implied contract exception may have increased job tenure, it did not affect returns to tenure. Finally, Autor et al. (2007) use establishment-level data to study how variations in employment protection affect productivity. They find that the good faith exception alone is associated with a significant decline in job flows and that firms responded to these constraints by increasing capital investment and employment of non-production workers. More tentatively, they suggest that although labor productivity rose, total factor productivity fell, indicating a possibly adverse effect on efficiency. Potentially confounding investment shocks, however, may also explain these effects.

Taking a somewhat different tack, MacLeod and Nakavachara (2007) theorize that legally enforceable just cause protection may solve a problem of inefficient contracting. In their model, firms undervalue workers’ relationship-specific investments principally because of an informational
asymmetry. Courts collect additional information about an employee’s performance and unjust dismissal litigation thus reduces turnover by improving employers’ incentives to screen more carefully. Employment protection also encourages workers to make more efficient relationship-specific investments. To implement their model empirically, MacLeod and Nakavachara distinguish occupation groups according to the level of training workers receive in their jobs. They find statistically significant effects for the good faith exception and, less consistently, for the implied contract exception. Estimates for the high-investment group show positive and significant employment and wage effects, while the low-investment sector experiences negative and significant effects. From the positive employment and wage effects that they observe for high-skill workers, the authors conclude that the good faith doctrine in particular may have economically beneficial effects.

Finally, Malani (2008) offers a fundamentally different strategy for assessing the consequences of employment protection regulations. He estimates hedonic regressions measuring how at-will exceptions – as well as no-fault automobile insurance laws, health insurance mandates and tort reforms – affect home values, rents and wages. Using this novel technique, Malani finds highly significant negative welfare effects of $128 per month for the good faith exception and $57 per month for the implied contract exception. Although he entertains the possibility that these per capita effects are ‘implausibly large’, Malani defends his method and concludes that perhaps these at-will exceptions ‘simply have larger than expected effects on welfare’. It would be premature to endorse or reject the hedonic approach to valuing legal rules, but it is indeed difficult to believe that one extremely narrow and rarely invoked employment doctrine reduces each citizen’s welfare by more than $1500 per year. Interestingly, the size of the estimated welfare effects is inversely correlated with the number of states that have recognized each exception. Perhaps some other important characteristic of these states correlates closely with the pattern of at-will exceptions. This question, along with the issue of measuring legal doctrines, discussed below, should receive more attention in future work.

C Political economy of employment protection

Another strand of the literature seeks to explain the origins and political dynamics surrounding the adoption of employment protection. In a seminal contribution, Krueger (1991) proposed that unjust dismissal legislation in the US was proposed (and in Montana adopted) in response to judicial erosions of an employer’s right to terminate at-will. His case study of Montana aptly illustrates the hypothesized mechanism. As we have seen, decisions of the Montana Supreme Court created an unusually intrusive cause of action
for wrongful discharge. Violations of a rather vague concept of good faith entitled successful plaintiffs to receive extensive tort damages. Business groups vigorously assailed these decisions, objecting to both large damage awards and legal uncertainty about when termination was impermissible. In response, the Montana legislature passed the Wrongful Discharge from Employment Act, which codified a more permissive standard for termination, narrowly limited remedies and created an incentive to use less costly arbitration procedures.

Krueger supports his causal conjecture with an analysis of panel data on proposed unjust dismissal legislation and corresponding court decisions recognizing exceptions to the at-will doctrine. He finds a statistically significant association between legislative proposals and both the good faith and public policy exceptions as well as the total number of exceptions recognized in a state. Serial correlation in the data on at-will exceptions severely constrains Krueger’s ability to investigate causation empirically. However, his exploration of alternate lag and lead structures, as well as a limited Granger-Sims causality test, lend some additional support to the hypothesis that judicial recognition of at-will exceptions spurs proposals for unjust dismissal legislation.

A second line of research seeks to explain the origin of differences in employment protection between nations. Saint-Paul (1997, 2000) advances a theory based on complementary rigidities in other legal and social institutions. Thus, powerful unions and highly regulated markets tend to support an equilibrium with significant legal constraints on discharge. Employment protection also tends to reinforce the power of insiders and creates a constituency of workers whose current position is largely dependent on those regulations. Saint-Paul observes that greater employment protection in the past thus generates stronger political support for maintaining (or perhaps even increasing) protection today. He concludes that once a country has enacted rigorous employment protection, it is difficult to displace.

Other scholars have offered theories that supplement Saint-Paul’s account of protection legislation. Some work has considered national differences in civic attitudes, religious belief and credit markets (Algan and Cahuc 2004, 2006). A recent paper by Belot (2007) proposes that employment protection is more (less) valuable for workers in countries with higher (lower) migration costs. In the US, greater labor mobility causes the median employed voter to prefer a legal regime with low protection and comparatively high turnover, because greater turnover improves a mobile worker’s chances of finding a good match. Conversely, barriers to labor mobility in Europe make workers value the more stringent employment protection common in those nations.

Brügemann (2007) takes yet another approach to explaining why
differences persist between countries’ labor regulations. He considers the multi-stage political process through which new employment protections are proposed, debated, enacted and implemented. In his model, firms have an opportunity to adjust employment after enactment but before implementation of a new level of employment protection. Voters, who are also workers, anticipate that employers will try to shed labor if the legislature adopts a more stringent law. In nations that already have strict protections against discharge, workers have little to fear and their support for new legislation will be undiminished. In countries with comparatively weak employment protections, however, workers are more vulnerable during the period before implementation. They will be less likely to support new protective legislation. Brügemann thus concludes that employment protection is both tough to scrap and tough to get.

Finally, a growing body of work has extended the analysis of legal origins to the task of explaining differences in labor regulation. Botero et al. (2004) investigate employment, collective bargaining and social security laws in 85 countries. They gather extensive legal data on statutory provisions and use them to construct summary indices for each area of regulation. Their empirical work examines three competing explanatory hypotheses for these labor market interventions – efficiency, political power and legal origins. For employment laws particularly, Botero et al. (2004) find support for the theory that common law countries regulate less intrusively than nations with a civil code of French legal origin. Unsurprisingly, leftist governments are also more protective, but the size of the legal origin effect is larger. The results for labor regulation broadly follow those obtained in other legal domains. This correlation lends some additional support to the notion that countries adopt a regulatory style determined in large part by its legal origin. Individual laws tend to be adapted to that style of regulation.

5 Challenges
An impressive and sophisticated body of research has investigated how the legal regulation of discharge affects the labor market. Despite this literature’s many admirable qualities, several persistent problems beyond those discussed above undermine its relevance to policy debates. Many scholars have noted, for example, that proving adverse effects on employment and wages leaves open the possibility that workers value the remaining jobs more highly than they would a larger number of less secure positions. Even ignoring this somewhat intractable problem of calculating social welfare, however, we should be cautious about relying too heavily on these undoubtedly creative efforts to understand employment protection.
A Some inconvenient facts

Many of the theoretical arguments frequently invoked to support increased legal protection against termination have rather obvious empirical implications. Surprisingly little effort has been directed at determining whether the available data support those implications. There are a few exceptions, as the earlier discussion of Schanzenbach (2003) and the following discussion suggest.

It has become commonplace to argue that workers may not negotiate for job security because they mistakenly believe that they already have legal protection against unjust discharge (Kim 1997). One implication of this argument is that better informed workers should be more likely to seek and obtain more protective contract terms governing discharge. A survey of employment contract terms among nonunion firms provides some contrary evidence (Verkerke 1995). One of the most striking features of the data is that contract terms were uniform for all employees at a given employer. Thus, middle managers and janitors, chemical engineers and manufacturing operatives all worked under identical terms governing discharge. Moreover, none of the law firms in the sample offered a just cause contract suggesting that even law firm associates, who we must hope are reasonably well informed about the law, failed to contract for just cause protection.

Yet another argument for just cause protection depends on asymmetric information about worker productivity (Levine 1991). A worker might well fear that asking for just cause protection will signal that he or she is a comparatively unproductive, ‘bad’ worker. Similarly, employers may be reluctant to offer a just cause contract for fear of attracting a disproportionate share of ‘bad’ applicants. The bad worker signaling theory implies that individually negotiated job security terms should be exceedingly rare. The adverse selection argument implies that firms should offer just cause only when the availability of easy screening devices allows them to prevent adverse selection.

Perhaps surprisingly, case law is filled with accounts of individual negotiations. Many leading cases involve precisely the sort of discussion about job security that the bad worker signaling theory precludes. Moreover, my own survey of employment contract practices reveals that 15 per cent of employers offer contractual job security (Verkerke 1995). According to the adverse selection theory, bad workers should flock to these employers, but they reported no plans to abandon their wholly voluntary commitment to employment protection. It is a similarly inconvenient fact that union collective bargaining agreements uniformly require just cause for termination. The available evidence suggests that, rather than being inundated with unproductive workers, the union wage premium tends to attract higher productivity workers.
Similar problems afflict theories based on bargaining power and economies of scale in providing just cause protection (Verkerke 1995). The point here is not to suggest that this limited empirical evidence conclusively refutes all of these arguments. Instead, the exercise suggests the potential value of looking more closely at existing contract practices. Patterns in these data can help us to assess the plausibility of common theoretical explanations for the prevalence of at-will contracts.

B Spillover effects

Existing work has also paid insufficient attention to the problem of spillover effects. Perhaps most obviously, national firms tend to standardize their human resource policies and employment documents. Thus, legal innovations in a few leading jurisdictions have a direct effect on how national employers behave in other locations around the country. Additional spillovers arise from the fact that an innovative decision in one jurisdiction is likely to cause firms to anticipate similar developments in other jurisdictions. Recent work by Bird and Smythe (2006) suggests that these lines of influence follow the jurisdiction of federal circuits rather than strict geographic proximity. But whatever the precise locus of these spillover effects, an estimation strategy that treats each new decision as arriving on a blank slate will underestimate the employment effects flowing from earlier legal changes.

The focus in Autor et al. (2004, 2006) on developments likely to generate a client alert letter is helpful. But an employment law firm’s advice about legal risks ordinarily will draw on a much broader range of influences than just those local decisions that are the subject of a client letter. Moreover, experiences at industry conferences, information from professional education seminars and general scuttlebutt about the developing legal climate undoubtedly affect the decisions of human resources managers and executives about hiring and firing. As with the problem of classifying legal doctrines correctly, these spillover effects unequivocally bias downward the coefficients on the employment protection variables.

The theoretically optimal approach would attempt to measure the probability that courts in a particular jurisdiction would, if presented with the opportunity, embrace a given doctrine. A decision in an influential jurisdiction is likely to be a strong predictor of what will happen in similarly inclined jurisdictions. Leading decisions thus cause firms to update their probability estimate, to adjust their hiring and retention of employees, and to take appropriate legal precautions. Rather than measuring the date of adoption in a jurisdiction, we should instead be measuring lines of influence among jurisdictions and information that would cause employers to update probabilities.
Finally, the conventional classification of jurisdictions into those that have adopted and those that have not adopted a doctrine exacerbates the mismeasurement of legal spillover effects. This approach treats as identical jurisdictions that have rejected a doctrine and those that have not yet spoken on the question. However, there is an enormous difference between a jurisdiction that has unequivocally rejected a theory of recovery and one that has simply not ruled on the question of whether or not to adopt it. A comparatively simple amendment to the system of doctrinal classification thus might distinguish three possible conditions for each doctrine in each jurisdiction – adopted, no ruling and rejected.

Bibliography


Hamermesh, Daniel S. (1988), ‘The Demand for Workers and Hours and the Effects of Job


