

Great Expectations: Law, Employment Contracts, and Labor Market Performance

W. Bentley MacLeod¹

Columbia University, Department of Economics, 420 West 118th, MC 3308, New York, NY 10027-7296, USA

Contents

1. Introduction	1592
2. The Law	1596
2.1. What is law?	1599
2.2. Employment law	1605
2.3. Exceptions to employment at will	1609
2.3.1. <i>Public policy exception</i>	1610
2.3.2. <i>Implied contract exception</i>	1611
2.3.3. <i>Good faith exception</i>	1612
2.4. Discussion	1614
3. The Economics of the Employment Relationship	1616
3.1. Why do we need models?	1616
3.2. Economics of the employment contract	1621
3.2.1. <i>Insurance</i>	1622
3.2.2. <i>Asymmetric information and the employment relationship</i>	1626
3.2.3. <i>The reliance interest</i>	1629
3.3. Implementing the efficient employment contract in a market	1634
3.4. Summary	1642
4. The Evidence	1643
4.1. Employment law	1643
4.2. The effect of unions	1662
5. Discussion	1685
References	1688

Abstract

This chapter reviews the literature on employment and labor law. The goal of the review is to understand why every jurisdiction in the world has extensive employment law, particularly

¹ I am grateful to Janet Currie, Harold Demsetz, Victor Goldberg, Lance Liebman, and seminar participants at the Columbia Law School Faculty seminar and the American Law and Economics Association meetings for helpful discussions and comments. I am also grateful to Elliott Ash, Wil Lim, and Uliana Logina for invaluable research assistance. E-mail address: bentley.macleod@columbia.edu.

employment protection law, while most economic analysis of the law suggests that less employment protection would enhance welfare. The review has three parts. The first part discusses the structure of the common law and the evolution of employment protection law. The second part discusses the economic theory of contract. Finally, the empirical literature on employment and labor law is reviewed. I conclude that many aspects of employment law are consistent with the economic theory of contract—namely, that contracts are written and enforced to enhance *ex ante* match efficiency in the presence of asymmetric information and relationship specific investments. In contrast, empirical labor market research focuses upon *ex post* match efficiency in the face of an exogenous productivity shock. Hence, in order to understand the form and structure of existing employment law we need better empirical tools to assess the *ex ante* benefits of employment contracts.

JEL classification: J08; J33; J41; J5; K31

Keywords: Employment law; Labor law; Employment contract; Employment contract Law and economics

"Now, I return to this young fellow. And the communication I have got to make is, that he has great expectations."

Charles Dickens, *Great Expectations*

"Take nothing on its looks; take everything on evidence. There's no better rule."

Charles Dickens, *Great Expectations*

1. INTRODUCTION

New jobs and relationships are often founded with great expectations. Yet, despite one's best efforts, jobs and relationships may end prematurely. These transitions might be the result of an involved search for better opportunities elsewhere, or in the less happy cases they may stem from problems in the existing relationship. These endings can be difficult, especially when parties have made significant relationship-specific investments. The purpose of this chapter is to review the role that employment and labor law play in regulating such transitions. This body of law seeks a balance between the need to enforce promises made under great expectations and the need to modify those promises in the face of changed circumstances.

The chapter's scope complements the earlier chapter on labor-market institutions in Volume 3 of this handbook by Blau and Kahn (1999). That chapter focused on policies affecting wage-setting institutions. Like much of modern empirical labor economics, Blau and Kahn (1999) use the competitive model of wage determination as the central organizing framework. Economists begin with the competitive model because it provides an excellent first-order model of wage and employment determination. The competitive model assumes that wages reflect the abilities of workers as observed by the market; this information, combined with information about a worker's training, provides sufficient information for the efficient allocation of labor.

Even if a labor market achieves production efficiency, it may nevertheless result in an inequitable distribution of income, as well as inadequate insurance for workers against unforeseen labor shocks. A number of institutions—such as a minimum wage, unions, mandated severance pay, unemployment insurance, and centralized bargaining—are viewed as ways to address these inequities and risks. Given that a competitive market achieves allocative efficiency, then these interventions necessarily result in allocative inefficiency. Hence, the appropriate policy entails a trade-off between equity and efficiency. For example, Lazear (1990) views employment law as the imposition of a separation cost upon firms wishing to terminate or replace workers. From this perspective, the policy issue is whether or not the equity gains from employment law are worth the efficiency costs. Many policymakers, such as the OECD and the World Bank, have taken the view that these employment regulations have for the most part gone too far—that they restrict the ability of countries to effectively adjust to economic changes and make workers worse off in the long run.²

Notwithstanding the mainstream skepticism toward efforts to regulate the employment relationship, it remains true that some form of employment law has operated in every complex market society for at least the last 4000 years—for example, the first minimum wage laws on record date back to Hammurabi's code in 2000 BC. This chapter therefore takes a somewhat different perspective, drawing upon the literature in transaction-cost economics pioneered by Coase (1937), Simon (1957) and Williamson (1975), as well as the work on law and institutions by Posner (1974) and Aoki (2001). This research on the economics of institutions, like empirical labor economics, begins with the hypothesis that long-lived institutions are successful precisely because they are solving some potential market failure. Accordingly, this chapter is organized around the following question: How can labor-market institutions be viewed as an *efficient* response to some market failure? Just as the competitive-equilibrium model supposes that wages are the market's best estimate of workers' abilities, the institutional-economics program views successful institutions as solving a resource-allocation problem.

This approach does not assume that these institutions are perfect. On the contrary, just as the competitive model yields predictions regarding how wages and employment respond to shock, the hypothesis that institutions efficiently solve a resource-allocation problem generates predictions regarding the rise and fall of these institutions. Important precursors to this approach are found in labor economics. In a classic paper, Ashenfelter and Johnson (1969) suggest that we should be able to understand union behavior, including the strike decision, as the outcome of the interaction between several interested parties. The work of Card (1986) demonstrates that observed contracts cannot be viewed as achieving the first best, and hence transaction costs are a necessary ingredient for understanding the observed structure of negotiated employment contracts. Despite this

² See the influential Jobs Study by the OECD (1994) and the recent work by the World Bank economists Djankov and Ramalho (2009).

early progress, the literature I review is still undeveloped. We do not have a good understanding of how the law works, nor do we understand the impact of legal rules on economic performance. In an effort to summarize what *is* known, this review is divided into three sections that correspond to coherent bodies of research.

Section 2 briefly reviews the structure of employment law and discusses some exemplary cases. A full review of employment law is not provided—for an excellent review, see Jolls (2007). My more modest goal is to provide some relevant insights into what law is and how it works. This sort of targeted inquiry is desirable because the standard assumption in economics is that the law enforces contracts as written. In practice, private law imposes *no* restrictions on behavior. It is mainly an adjudication system that can, after a careful review of the evidence, exact monetary penalties upon parties who have breached a duty. Hence, private law is a complex system of incentive mechanisms that affects the payoffs of individuals but does not typically constrain their choices. The distinction is important for economics because it is convenient to model legal rules as hard constraints on behavior—that is, as structuring the *available moves* in a game rather than just altering some of the expected payoffs. This approach also implies, wrongly, that rules apply equally to all individuals. Treating private law as an incentive system, instead, implies that the impact of the law is heterogeneous—an individual's response to a legal rule will vary with an individual's characteristics, such as wealth, attitudes toward risk, and the evidence that one can present in court.

Heterogeneity and information also play key roles in the theory of employment contracts reviewed in Section 3. The past forty years have witnessed tremendous progress in the economic theory of contract, especially in terms of teasing out how a particular set of parties should design a contract given the transaction costs characterizing the employment relationship. The influential principal-agent model, for example, was developed in the context of the insurance contract, which specifies state-contingent payments.³ The modern theory of contract, building on the work of Grossman and Hart (1986), recognizes that an important function of economic institutions and contracts is the efficient allocation of authority and decision rights within a relationship.

Most economists agree that unions and employment law affect the relative bargaining power of individuals. These institutions are usually interpreted as mere re-distribution of rents, and so any allocation of bargaining power that results in prices diverging from competitive levels is inherently inefficient. The modern literature on contract views authority as an instrument for mitigating transaction costs due to asymmetric information and holdup. This perspective also naturally admits a role for fairness in decision-making, and hence can provide an economic rationale for why fairness concerns are important in the adjudication of a dispute.⁴

³ See Pauly (1968).

⁴ See Kornhauser and MacLeod (2010) for a fuller discussion of this point.

Section 3 also discusses the empirical content of these models. The modern empirical literature is concerned with identifying a causal link between various labor-market interventions and performance. Unfortunately, much of the economic theory of contract is not amenable to this approach. These models typically describe how matches with certain observable features (X variables) result in an employment-compensation package (Y variables). As Holland (1986) makes clear, these are not causal relationships, but merely associations. For example, the predicted relation between the sex of a worker and the form of his/her employment contract is not causal, since the sex of a worker is not a treatment variable.

This distinction is useful because it helps explain the gulf between much of the theory discussed in Section 3 and the empirical evidence discussed in Section 4. It is also worthwhile to keep in mind that *all economic models are false*. This does not imply that these models are not useful. As Wolfgang Pauli quipped regarding a paper by a young colleague — “it’s not even wrong!”⁵ Rather, economic models are decision aids that guide further data collection and help in selecting between different policy interventions. The upshot for empirical researchers is that one typically tests the associations that the theory predicts, rather than the theory itself. Empirical determinations of the validity of these associations can help us decide whether and to what extent we can rely on the model as a decision aid.

The theory section discusses how contract theory can be used to understand employment law, and the conditions under which it may be desirable. We begin with a discussion of how contract design is affected by the interplay between risk, asymmetric information and the holdup that arises from the need for parties to make relationship specific investments. These models can be used to explain the role of the courts in enforcing the employment contract. The recent property rights theory of the firm developed by Grossman and Hart (1986) illustrates the importance of governance and the associated allocation of decision rights in order to achieve an efficient allocation.⁶ A contract is an instrument that explicitly allocates certain decision rights between the contract parties. This can also be achieved with unions. This section discusses how the appropriate allocation of power and decision rights can enhance productive efficiency. Thus, these theories provide conditions under which union power may enhance productive efficiency, as suggested by Freeman and Medoff (1984).

Section 4 reviews the empirical evidence on employment and labor law. Here we are concerned with explicitly causal statements such as the question of whether or not a reduction in dismissal barriers will reduce unemployment. This is a causal inquiry because it compares the outcomes from two different choices: having more or less employment protection law. Section 4.2 discusses the literature on unions that addresses

⁵ See Peierls (1960). It is also worthwhile to point out that even though Newtonian physics is false, it is all that one needs for most practical applications.

⁶ See Hart (1995) for a full discussion of this approach.

two questions. First, does unionization of a workforce increase productive efficiency? Second, does unionization increase or decrease firm profits? Even if a union increases productive efficiency, if the increase in rents extracted by the unions is greater than the increase in productive efficiency, then profits would fall as a consequence, which in turn may lead to a decrease in unionization. The chapter concludes with an assessment of the evidence and a discussion of future directions for research.

2. THE LAW

"The law embodies the story of a nation's development through many centuries, and it cannot be dealt with as if it contained only the axioms and corollaries of a book of mathematics. In order to know what it is, we must know what it has been, and what it tends to become."

Oliver Wendell Holmes, *The Common Law*, 1881

The purpose of this section is not to provide a comprehensive review of employment law. Rather, the goal is to provide a sense of how employment law has developed so that one might better understand its impact on the employment relationship.⁷ In the United States, employment law is primarily the domain of the states. Section 4 reviews several empirical studies that have exploited the natural experiments resulting from variations in state laws to measure the impact of various laws on economic performance. Overlaying the state laws are a number of federal statutes affecting the employment relationship, including the National Labor Relations Act of 1935 (allowing workers to organize collective bargaining units), Fair Standards Act of 1938 (establishing minimum employment standards), Employee Retirement Income Security Act of 1973 (ERISA) (ensuring that employee benefits meet national standards), Occupational Safety and Health Act of 1970 (OSHA) (establishing minimum health and safety standards in the workplace), and Family and Medical Leave Act (establishing protections for leave related to personal sickness or family emergencies). These laws are enumerated in Table 1.

While the primary concern of the present chapter is the United States, studies of employment law in other countries are also discussed. Blanpain (2003), for example, provides a comprehensive review of European law. As in the United States, European law is complicated by the fact that both individual countries and the European Parliament create rules that affect the employment relationship. More generally, all countries in the world have some system of employment laws, created and adapted to the circumstances of each jurisdiction.

One chapter cannot do justice to the dizzying complexity of the law across jurisdictions, even if attention were restricted to a narrow area such as employee-dismissal law. As the quote from Holmes (1881) illustrates, legal systems are complex

⁷ See Jolls (2007) for a survey of employment law, and Rothstein and Liebman (2003) for a more comprehensive review of US law. Gould (2004) provides an accessible discussion of American labor law, and how it differs from European labor law.

Table 1 United States employment laws.

Racial discrimination	Civil Rights Acts of 1866, 1964	1866	Bars racial discrimination by employers.
Social security	Social Security Act	1935	Distributes social security benefits to those of retirement age.
Minimum wage	Fair Labor Standards Act	1938	Establishes a minimum hourly wage.
Overtime rights	Fair Labor Standards Act	1938	Requires that employers pay a higher wage for work exceeding 40 hours a week.
Child labor	Fair Labor Standards Act	1938	Places limits on many forms of child labor.
Gender discrimination	Civil Rights Act of 1964	1964	Prohibits gender-based discrimination.
Religious discrimination	Civil Rights Act of 1964	1964	Prohibits discrimination against employees on the basis of religion.
Age discrimination	Age Discrimination in Employment Act	1967	Prohibits discrimination against workers above the age of 40.
Workplace safety	Occupational Safety and Health Act	1970	Establishes minimum standards for workplace safety.
Good faith exception	<i>Fortune v. National Cash Register Co.</i> , 373 Mass. 96	1977	Provides for a wrongful discharge claim against employers violating the common-law contract duty of good faith and fair dealing.
Public policy exception	<i>Tameny v. Atlantic Richfield Co.</i> , 27 Cal.3rd 167	1980	Provides for a wrongful discharge claim against employers when the discharge would be a violation of public policy, for example, when the employee is fired for refusing to commit a crime.
Implied contract exception	<i>Wooley v. Hoffmann-La Roch</i> , 99 NJ 284	1985	Provides for a wrongful discharge claim against employers when an employment contract is implied.

(continued on next page)

Table 1 (continued)

Sexual harassment	<i>Meritor savings bank v. Vinson</i> , 477 US 57	1986	Recognizes sexual harassment as a violation of the Civil Rights Act of 1964.
Layoff notice	Worker Adjustment and Retraining Notification Act	1988	Companies must give 60 days' notice before large-scale layoffs.
Whistleblower protection	Whistleblower Protection Act of 1989	1989	Prohibits retaliation against employees for reporting illegal acts against the federal government.
Disability discrimination	Americans with Disabilities Act	1990	Prohibits discrimination based on disability.
Pregnancy discrimination	Civil Rights Act of 1991	1991	Prohibits discrimination against employees because they are pregnant.
Medical leave protection	Family and Medical Leave Act	1993	Requires employers to allow workers 12 weeks of unpaid medical leave for certain medical conditions of themselves or close relatives.

systems that evolve over time to resolve the variegated disputes faced by parties of commercial transactions. To make sense of this complexity, I follow the lead of the law-and-economics movement as epitomized in the work of Richard Posner, who argues that the law, especially the common law, has evolved over time to address the needs of individuals trading in a market economy.⁸ Posner (2003) explicitly poses the rhetorical question: “How is it possible, the reader may ask, for the common law—an ancient body of legal doctrine, which has changed only incrementally in the last century—to make as much economic sense as it seems to?”⁹

The claim is not that the entire body of rules and norms governing economic activity can be viewed as the solution to the problem of efficiently organizing economic activity. Rather, the claim is that individual rules have evolved to solve particular problems that appeared repeatedly before the courts. From this perspective flow some observations that may help explain the theoretical and rhetorical gaps between legal rule making, economic models of the labor market, and economic policy making. The advantage of the economic approach is that it allows one to explore the empirical implications of a simplified representation of the law. The disadvantage, particularly for purposes of

⁸ See Ehrlich and Posner (1974), and also Posner's classic work, *Economic Analysis of Law*, now in its seventh edition.

⁹ See p. 252.

economic policy analysis, is that the simplifying assumptions may miss key characteristics of the law that are important in practice.

This gap between law and economics is particularly salient in laws protecting employees from discharge, wrongful or otherwise. In economics, employment protections are typically modeled as a form of turnover costs, leading to the view that such laws probably interfere with an economy's efficient response to shocks. As a consequence, organizations such as the OECD (see OECD (1994)) have advocated reducing or abandoning employment protections. Recent work by Blanchard and Tirole (2008) addressing how France and other countries should design unemployment insurance and employment protection concludes that there is no role for the law beyond enforcing employment contracts.

Nonetheless, as a matter of law, there are no jurisdictions where courts enforce all privately agreed-upon contracts. Labor contracts with young children, for example, are almost universally prohibited. Generally speaking, there is a substantial gap between the law in practice and the law as represented in many economic models of employment. One reason for this gap is that legal practitioners rarely have any reason to use an explicitly economic approach to understand the form of a particular contract. Lawyers typically represent clients in cases after the fact; the question of why a legal rule exists is not important to them. The real issue is to predict how a judge will rule and then present the case so that their client will do as well as possible in what is essentially a negative-sum game between the plaintiff and defendant. In this game, the details of the law are crucial, but the reasons why they have a particular form are not usually relevant.¹⁰

In consequence, the concerns of the legal scholar are quite different from those of the economist, which in turn creates a gap between the law as it exists and the law as modeled in economics.¹¹ The goal of the next two sections is to narrow that gap ever so slightly, at least in the context of employment regulation. The next subsection discusses the generic structure of the law. Even though legal rules vary greatly from jurisdiction to jurisdiction, the notion of law and how it works has some universal features. More specifically, Section 2.1 examines three well-known employment-law cases in the United States and the United Kingdom. These cases illustrate the complex problem faced by a judge in an employment dispute. Moreover, as these cases demonstrate, the courts are not passive agents; they play an active role in resource allocation. Consistent with the "Posnerian" perspective, the decisions in some cases can be viewed as enhancing economic performance.

2.1. What is law?

The notion of a legal rule dates back at least as far as 2000 BC, with the Babylonians following Hammurabi's code and the Mesopotamians following the code of Urukagina.

¹⁰ Furthermore, judges are often economically illiterate—and are often suspicious of statistical evidence—so economic arguments might even be counterproductive. See the discussion by Posner (2008) and Breyer (2009).

¹¹ See discussion in MacLeod (2007b).

While some of these ancient edicts imply curious beliefs about causality,¹² it is clear that the purpose of many of these rules is to modify or constrain human behavior. For as long as law has existed, its purpose has been to facilitate the efficient functioning of civil society.

The aforementioned codes even had room for what we would now call labor law. Hammurabi's code includes, for example:

- Minimum wage rules: "If any one hire a day laborer, he shall pay him from the New Year until the fifth month (April to August, when days are long and the work hard) six gerahs in money per day; from the sixth month to the end of the year he shall give him five gerahs per day".
- Liability rule: "If a herdsman, to whom cattle or sheep have been entrusted for watching over, and who has received his wages as agreed upon, and is satisfied, diminish the number of the cattle or sheep, or make the increase by birth less, he shall make good the increase or profit that was lost in the terms of settlement".

These are remarkable examples of how legal rules are created to regulate the employment relationship. That minimum wage rules persist to this day suggests that there are robust reasons for the existence of such rules. Recognizing this possibility, the law-and-economics approach seeks to explain the rules as solutions to well-defined market imperfections. Ostrom (2000), for example, has shown that many societies have developed efficient systems of rules and adjudication for regulating the use of common-pool resources, thereby avoiding the tragedy of the commons (Hardin, 1968). Ostrom observes that all successful commons-governance regimes consist of a set of rules that have the following features:

1. The rules are commonly known;
2. There are penalties for breaking rules that increase in intensity with the severity and frequency of violation;
3. There is an organization or an individual who is responsible for imposing penalties when informed; and
4. There is a process of adjudication when there are disagreements regarding whether an offense has occurred and what penalty should be imposed.

All organizations, including firms and families, have rule systems with these features. The economic analysis of such rule systems typically entails asking what set of rules would achieve an efficient allocation.¹³ In the common-pool-resource problem, for

¹² Hammurabi's code, Paragraph 2, decrees: "If any one bring an accusation against a man, and the accused go to the river and leap into the river, if he sink in the river his accuser shall take possession of his house. But if the river prove that the accused is not guilty, and he escape unhurt, then he who had brought the accusation shall be put to death, while he who leaped into the river shall take possession of the house that had belonged to his accuser".

¹³ Determining the optimal system for selecting a course of action given the preferences of individuals, the technology of the environment, and the information available is the subject of mechanism design theory. See Jackson (2001).

example, one seeks a set of rules ensuring that each person with access does not overuse the resource. Organizational economics tries to determine which systems of rules and compensation in firms or public entities ensure that agents reveal useful information and choose efficient levels of effort.

What distinguishes the law's rule system from a family's or organization's is not the existence of binding rules, but the sources of enforcement and adjudication. When we speak of a legal system, we mean the set of rules and associated penalties that are enforced by the state.¹⁴ As discussed above, however, the jurisdiction associated with a dispute is often poorly defined. Even when the jurisdiction is well-defined, disputes can often be decided by multiple judicial bodies (including private mediation and arbitration). Many countries, such as Italy and the United Kingdom, maintain a separate system of courts for ruling on employment disputes. In the United States, disputes regarding a collective-bargaining agreement may be brought before the National Labor Relations Board, employment disputes before a state court, and Title VII discrimination suits before a federal court.

To illustrate what we mean by a legal system (and employment law in particular), let us consider the proverbial worker-firm relationship. In the standard neoclassical model of employment, the worker agrees to supply L hours of labor for a wage w . This can be viewed as a supply contract where the hours are, say, consulting time. In that case, the relationship would be governed by contract law. The worker must supply L hours, and the employer must pay wL dollars.

Should the worker supply less than L hours, the worker has *breached* the contract. Suppose that the firm has paid P_0 in advance, a common practice if the worker is, for example, a lawyer. Some sort of binding agreement is needed; otherwise, the worker would simply take the P_0 and try to find employment elsewhere. The question, then, is: What incentives does the worker have *not* to breach the agreement?

One possibility is the use of an informal enforcement mechanism. This would include firms' telling each other that the individual has breached, and hence should not be dealt with.¹⁵ Another alternative is to use physical violence against the individual, a common technique in the illegal drug trade and other black markets.¹⁶ While both enforcement systems are still widely used, societies have evolved more legalistic systems of adjudication for the simple reason that parties sometimes fail to perform even if they act in good faith.

¹⁴ It is worth highlighting the fact that a set of written rules is not a requirement, nor even a defining feature of a legal system. In the Middle Ages, the subjects of a feudal lord faced a number of rules regulating their life on the manor, most of which were not written (Bloch, 1961); nonetheless, the set of rules was widely understood. Written rules do have good social consequences, however, by assisting with evidentiary issues and facilitating agreement among parties regarding the applicable rule. Inter-subjective agreement on the law is an important issue in modern labor law, as we shall see in the case of employee handbooks.

¹⁵ Greif (1989) has some nice examples from the Maghrebi traders, who did write letters to each other in this regard.

¹⁶ Naidu and Yuchtman (2009) document that criminal law, with punitive sanctions, was widely used to enforce employment contracts in 19th century England.

Enforcement systems that trigger punishment regardless of the *reason* for breach, such as violence among drug dealers, are simply not always efficient.¹⁷

In contrast, if the contract is viewed as a *legally binding* agreement, then the breach of contract by the worker gives the firm the *right* to seek damages in court. If the firm prevails, the court can order the worker to pay damages to the firm; if the worker refuses to pay, the court can still enforce the decision by ordering the seizure of the worker's assets.

The fact that contract breach leads to the right to file suit—as opposed to an automatic penalty—is a feature that distinguishes legal enforcement from other forms of rule systems, such as rewards within a firm. In the economics literature, it is common to view any agreement between parties that links future rewards to actions as a “contract”. Jensen and Meckling (1976), for example, famously proposed that one should conceptualize a firm as a “nexus of contracts”. If all contracts are enforceable at negligible cost, a reward system that promotes an employee for good performance and an agreement with an outside supplier to pay a bonus for sufficient quality are assumed to be equally enforceable. If employment contracts are enforceable at no cost, subject only to information constraints, then explaining the contract form requires only that one carefully specify the environment and then use principal-agent theory to work out the optimal contract.

The difficulty, as Williamson (1991) observes, is that the law uses forbearance for transactions within a firm. Even if a firm promises a promotion, that does not confer a right upon the worker to sue the firm should the promotion not be offered. All dispute resolutions of this sort occur strictly within the firm, with no appeal to an outside legal authority available. A full discussion of the role of courts in such disputes must wait until the next subsection, but for now the relevant point is that for any contract between two legal persons (in our example, firm and consultant), both parties *always* have the right to seek damages in court should there be a breach of contract.¹⁸

That the decision to sue is discretionary implies that the same rule may have different effects depending on the characteristics of the parties of the contract. The characteristics of the parties might even be the most important factor in whether a lawsuit is filed. For example, suppose that a company has illegal discriminatory hiring practices. If the market is thick, with plenty of employment alternatives, potential employees may not find it worthwhile to bring suit against the company. In a less friendly employment market, if an individual believes he has been the victim of illegal discrimination and cannot find another job, he may bring suit. This point is illustrative of the fact that we will always find differences between a legal rule on paper and the same rule in practice.¹⁹

¹⁷ See MacLeod (2007a) for discussion and proof of this point.

¹⁸ See Kornhauser and MacLeod (2010) for a discussion of the concept of a legal person.

¹⁹ Dunford and Devin (1998) have a nice discussion of how employee perceptions affect the decision whether to file suit. For a review of the literature on the decision to file a suit see Kessler and Rubinfeld (2007).

A second source of uncertainty is how the court will decide a case. In the case of the breaching consultant, the court has to make two decisions, each of which is prone to statistical error: (1) deciding whether a breach has occurred, and (2) if so, the damages to be paid. For damages, the general rule is that courts order expectations damages—namely, the losses associated with the worker's breach. An example of a formula that the court might use is:

$$\begin{aligned}\text{damages} &= P_0 + w_1 L - wL, \\ &= \text{prepayment} + \text{cost of replacement} - \text{promised wages not paid}.\end{aligned}$$

Here w_1 is the wage for the replacement worker. The worker has to pay the costs associated with finding a replacement worker rather than the value of the work done, due to contract law's requirement that injured parties make every effort to mitigate losses arising from breach. The mitigation rule is mandatory, meaning that parties cannot contract around it. Employment law has many other mandatory rules; slavery is prohibited, for example, as is discrimination on the basis of race, age, or sex.

Expectation is not the only way to calculate damages from contract breach. In a famous paper on contract damages, Fuller and Perdue (1936) identified restitution and reliance as other possible measures of damages.²⁰ *Restitution* damages, for one, are intended to put the firm into the same financial position—as if the contract had never been signed. In our consultant example, restitution damages would only require the worker to repay P_0 . Matters are less clear if P_0 is paid to the worker so that he can buy passage to the job site and secure accommodation. Suppose upon arriving, the worker is injured and cannot perform his duties; the firm sues for breach. If the contract does not specify what happens in this contingency, a court may be asked to fill in the gaps in the agreement. Some other issues that might not be described in the contract (and which the court must adjudicate) include whether the injured worker may be fired or whether he must take out a loan to repay P_0 .

Alternatively, the firm, in the expectation of being able to rely upon performance, may have paid for passage and accommodation. In that case, if the worker does not perform, then she may be asked to compensate the firm for these *reliance* expenditures.

Finally, the contract could have provided that a non-performing worker pay a fine P_0 . In this case, non-performance would *not* be a breach of contract, and a breach would occur only if the worker failed to pay back P_0 . Notice that in this case, the damages would simply be P_0 , even under the expectation-damages rule. This example illustrates that for the same exchange relationship there is no unique notion of contract breach. Rather, the

²⁰ This work has been influential in law and economics. Economists have used the three damages measures—expectations, restitution, and reliance—as alternative legal rules that can be analyzed using the standard tools of economics. See, in particular, Shavell (1984) and Rogerson (1984).

contract *defines* what constitutes breach, and then the courts must decide whether or not to enforce the terms set out in the agreement.²¹

Defining the conditions for breach is not a clear-cut exercise. For example, the contract could have specified *liquidated damages* in the amount P_0 . In this case, breach would occur for non-performance, but then the contract would direct the court to set damages at P_0 . This contract seems to be equivalent to the previous one (indeed, most economic theories of contract would see them as equivalent), but there is an important distinction. In the previous case, if the worker pays P_0 , no breach has occurred, and the firm has no right to bring suit against the worker. In the liquidated-damages case, even if the worker offers to pay P_0 , the firm still has a right to sue the worker because technically a breach has still occurred. Admittedly, the firm would face an uphill battle in court if the worker had offered to pay—and normally would have no reason not to accept the offer as part of a settlement agreement—but the right exists nonetheless. The firm might believe, for example, that the worker abused the liquidated-damages clause, accepting the consulting job only as a contingency in case another opportunity did not work out. If the firm's belief is true, the worker has arguably violated the requirement of good faith and fair dealing—another mandatory rule in the common law of contracts. Arguing that the liquidated-damages clause no longer applies, the firm might ask for expectation damages larger than P_0 .

Conversely, assume that the firm sets liquidated damages at three times P_0 . Breach occurs, and the worker declares these damages to be unconscionably high. The worker may have a valid claim under contract law's prohibition against penalty clauses. This doctrine—another mandatory rule—provides that liquidated damages far exceeding the losses to the injured party will not be enforced by courts.

The prohibition on penalty clauses, along with the other examples of mandatory rules discussed above, illustrate that the law allows for a significant degree of judicial intervention into private contracts. Among other things, a court can fill in missing terms and refuse to enforce unreasonable terms. The decision whether to intervene is often at the court's discretion, but courts usually turn to the Uniform Commercial Code and other statutes, as well as previous court decisions, to justify these interventions. Thus the legal system is an adjudication process that modifies contracts in the face of a breach as a function of past experience and practice.

The extent to which courts should intervene into freely entered agreements has proven to be controversial. Early scholars, such as MacNeil (1974), proposed that courts rely on many sources of information, including industry custom and the history of the relationship between the plaintiff and the defendant, when making a decision. If parties have a longstanding relationship, Macneil argued, contract terms should be enforced

²¹ See MacLeod (2007a) for a discussion of whose reputational concerns interact with the breach decision. In particular, the theory predicts that the party whose reputation is the most valuable should be the one who is responsible for initiating breach.

within the context of the relationship. Macneil and many others believed that this sort of context-sensitive adjudication could help repair the parties' relationship and facilitate the continuation of mutually beneficial exchange. Most economic theories of contract, in contrast, work from the assumption that parties have well-defined interests and can draft agreements efficiently, implying that contracts should be enforced as written. This presumption has led to a law-and-economics scholarship that mostly argues for the curtailment of judicial discretion, and for a more systematic dependence upon basic economic reasoning when ruling on a case (see Goetz and Scott (1980) and more recently Schwartz and Scott (2003)).

Regardless of one's theoretical commitments, it remains the case that the law does not simply enforce a set of well-defined rules. The law does include a set of rules, but along with a system of adjudication that results in a context-sensitive application of these rules to individual cases. This context sensitivity includes, among other things, consideration for the idiosyncratic features of the parties. The basic principles of contract law apply to all agreements between two parties, but more specialized bodies of law have evolved to regulate specific classes of contracts. The insurance industry is regulated by a specialized area of contract rules (see for example Baker (2003)) as is employment law. It is to this latter body of law that we now turn.

2.2. Employment law

Employment law evolved from contract law and master-servant law to deal with the unique problems characterizing the modern employment relationship. The first task is to determine the difference between (1) a firm's relationship with an outside contractor selling services, and (2) its relationship with an employee. The difference not only affects the area of law that regulates the relationship, but it also affects the relevant tax law. In the United States, the Internal Revenue Service will find that an employment relationship exists when "the person for whom services are performed has the right to control and direct the individual who performs the services, not only as to the result to be accomplished by the work but also as to the details and means by which that result is accomplished".²²

As this tax regulation exemplifies, the obligation of the employee is to follow his employer's directions, not to produce a specific service with particular characteristics. Simon (1951)'s model nicely captures this distinction between sales and employment. In a *sales contract*, says Simon, the seller agrees to supply a particular good or service x from the set of all possible goods and services X , and in exchange the buyer agrees to pay a sum P . An *employment relationship*, in contrast, is characterized by a subset of all possible goods and services, $A \subset X$, that represents the set of duties that the employer might ask the employee to carry out. A might include the service x defined in the aforementioned

²² Treas. Reg. Sections 31.3121(d)-1(c)(2).

sales contract, but that single task would normally be just one component in a broad complex of obligations defining an employment relationship. In exchange for a promise to carry out these duties, the employer agrees to pay a wage w .

Simon's simple model highlights an essential feature of the employment relationship, namely, the admissible scope of a person's job as represented by the set A . The admissible set of tasks $S \subset X$ —that is, the set of acts that an employer is allowed by law to command—has been subject to a plethora of regulation and litigation. For example, is it conscionable for a firm to require 50 hours a week? Can a manager ask her assistant to commit crimes?

An employment relationship often begins with little formal agreement about the tasks the employee will be asked to carry out. The longer the potential duration of the employment, the more incomplete the initial employment agreement. Given the informal nature of such agreements, when disputes do arise the courts will have little to rely upon when constructing the obligations of each party. With poorly defined obligations, determining whether a breach occurred presents a difficult task, as does choosing an appropriate remedy.

The combination of extreme contract incompleteness and daunting litigation costs have convinced many legal scholars that the appropriate default rule is at-will employment. The courts have converged to this default rule partly because they now view employment law as an extension of contract law. That view diverges significantly from the early case law on employment disputes, which was mostly governed by “master-servant law”.²³ That old body of law consisted of a set of legal default rules developed in England and the United States to deal with cases involving domestic servants. In the master-servant relationship, the customary period of employment lasted one year; courts held that neither party should sever the relationship before then.

In a widely cited work, Wood (1877) argued for replacing this law with the rule of at-will employment, where both parties can sever the relationship whenever they wish and face no liability beyond the requirement that the employer pay her employee the agreed-upon wage for work already completed. Wood's argument was a pragmatic one, based on the bad experiences of many employers and employees with the inflexibility of master-servant law. As detailed in Feinman (1978), the new rule was quickly adopted by the New York courts and remains the default rule today. In California, the legislature adopted what is now Section 2922 of the California Labor code, which provides that “employment, having no specified term, may be terminated at the will of either party on notice to the other”.

The at-will-employment rule figures prominently in most economic models of the labor market. As these models have it, workers and firms enter into relationships that are preferred to the alternatives in the marketplace. Should a firm mistreat a worker, or

²³ See Feinman (1978) for a review of the development of this law.

have high standards for performance or number of hours worked, the firm will have to pay relatively high wages or else the worker will leave. Similarly, if a worker demands a higher wage or better working conditions, the firm is free to search for another worker who will abide by the current arrangements. In equilibrium, all firms and workers are satisfied with their lot relative to the alternatives.

The hypothesis of a perfectly competitive market can explain many broad features of wages and employment over time, but it cannot explain the emergence of the at-will-employment rule. This is an example of the model's inability to explain the emergence of laws that seem to constitute reasonable responses to real economic issues. That failure indicates flaws in popular economic models—but not in economic reasoning generally. Consider the case of child labor. As societies have become more wealthy, they have gradually imposed stricter legal constraints on the minimum age and maximum hours for minors in the workplace. By the perfectly-competitive-market hypothesis, these restrictions would be unjustified because only those families for which child labor is efficient would put their children to work rather than in school. On the other hand, a more realistic economic inquiry recognizes the market imperfection imposed by liquidity constraints: children (and parents) cannot borrow against future income arising from education, so many families send their children to work for a short-term gain in income rather than invest in a long-term gain from education. Investing in education results in superior overall welfare, so the choice to put children to work is inefficient. Laws that regulate child labor, like the Fair Labor Standards Act, are justified because, by increasing the cost of child labor, they motivate families to substitute education for labor. By increasing investments in education, these laws increase social welfare.

The analysis in the previous paragraph indicates the potential insight to be gained from an evolutionary perspective when investigating the law and economics of employment law. Applying this perspective to our law's historical origins, we observe that employment law adapts to the changing macroeconomic environment. One of the earliest labor statutes on record, the Ordinance of Labourers, addressed the problems of unharvested crops, rising wages, and poaching of workers faced by English landowners at the height of the Black Death. Similarly, in 1630, the Massachusetts General Court placed a wage cap of 2 shillings a day on skilled craftsmen, who were at that time taking advantage of limited supply. More recently, the Fair Labor Standards Act became law in the midst of the Great Depression, when workers lacked market power to ensure good working conditions. Mandatory overtime pay, meanwhile, incentivized firms to hire more workers at fewer hours each, thereby serving as an income- and risk-sharing function.

These legal adaptations to changes in the labor-market environment can all be conceived as forms of insurance, whether against the waste of unharvested crops, gouging by craftsmen, or unemployment. In the next subsection, we discuss recent work based on the hypothesis that workers are risk-averse, which might help explain some of the

features of these laws. Certainly, both minimum-wage laws and unemployment insurance can be viewed as forms of imperfect insurance.²⁴

Once the issue of risk is put aside, the law-and-economics movement has tended to take the view that employment at will is the optimal default rule (see, e.g., Epstein (1984)). Within economics, a tradition including Friedman (1962) and Alchian and Demsetz (1972) has viewed labor services from the perspective of the buyer-seller contract—with *no* remedies for contract breach. Specifically, Friedman argued that a competitive market with free entry and exit is the most efficient market form, even when contracts cannot clearly specify quality. In his vision of the world, workers and firms trade freely within the context of the sales contract (quality x in exchange for price p); should performance be inadequate, the worker would gain a poor reputation and thereafter be excluded from the market.

There are two difficulties with this argument. The first, as MacLeod (2007a) discusses, is that the literature on relational contracts shows that reputational concerns are neither necessary nor sufficient to ensure efficient exchange. Second, the common law has developed many doctrines that limit the freedom of contract in the context of the simple buyer-seller model. Posner (2003) suggests that these developments tend to enhance contract performance. Chakravarty and MacLeod (2009) present evidence that this is indeed the case for a large class of contracts that are common in the construction industry.

As the Simon model highlights, the employment relationship is different because a performance obligation is created *ex post*. If the worker accepts a contract with scope $A \subset X$, then the performance obligation is created when the employer asks the employee to carry out $x^* \in A$. If the relationship were governed by standard contract law, then if the employee chooses $x^b \neq x^*$ the employer-cum-buyer could sue for damages $B(x^*) - B(x^b)$, where $B(x)$ is the benefit to the employer of action x . Under employment at will the general rule would be that the employer has no right to sue, but she can freely dismiss the employee, even if performance is satisfactory.

Correspondingly, the employee has the right to leave whenever he wishes. For example, if the employer asks the employee to carry out an action outside the scope of his duties, then under at-will employment the employee has the right to refuse to carry out the task and find another job. In contrast, in the case of a construction contract, if the buyer were to ask for a modification to the building plan, then under US law the contractor would have an *obligation* to carry out the modification, but he would also have the right to sue for the additional cost of the change if the buyer/builder does not adequately compensate him for the changes.

The defining feature of at-will employment is that in each period parties are free to renegotiate the contract, with the outside options defined by each parties'

²⁴ The reader is referred to Blau and Kahn (1999) and Rogerson et al. (2005) for excellent reviews of this literature.

market opportunities. Consistent with the Coase theorem, we should expect at-will employment to give rise to arrangements that are *ex post* efficient. This observation has led some legal scholars (e.g., Epstein (1984)) to suggest that exceptions to employment at will are inefficient. Yet today in the United States, as in most other jurisdictions worldwide, the law of wrongful discharge is alive and well. Indeed, there are clear exceptions to the rule of employment at will. In the next section, we discuss three of the most important exceptions figuring in recent empirical work on employment law.

2.3. Exceptions to employment at will

This section discusses three exceptions to employment at will that have found broad support in US courts. These exceptions are judge-made laws, created in response to difficult cases; hence, they are good examples of how the common law evolves in response to the disputes that arise in practice. The three exceptions we consider are (1) the public policy exception, protecting from employer retaliation those workers that act in a way consistent with accepted state policy, (2) the implied contract exception, protecting workers who can show that the implicit contract with the employer entails just-cause dismissal, and (3) the good-faith exception, requiring employers and employees to behave in ways consistent with fair dealing.

US courts rarely order specific performance—that is, the losing employer typically still has the right to discharge the employee—and hence the issue is usually one of damages: How much should she have to pay for this right? In other jurisdictions, however, reinstatement is sometimes considered an acceptable remedy. One of the few US cases in which specific performance was granted in an employment dispute was *Silva v. University of New Hampshire*.²⁵

The question of damage awards is not straightforward, but economics can assist in organizing our thinking. If markets are perfectly competitive—and a worker's compensation is equal to his best market alternative—dismissal does not entail any harm. However, as Mincer (1962) has shown, the second assumption can break down when the worker's training costs were significant. If the worker paid for some of the training costs, he will be compensated for them through increased future compensation, and therefore his income may be in excess of the best market alternatives. More often, dismissal entails a costly job search and possibly relocation. When an employee has been wrongfully discharged, the court will award damages that reflect these costs.

An additional complication is whether the wrongful-discharge action comes under tort or contract law. The first exception to the at-will employment rule, a claim for wrongful discharge as violation of public policy, is considered a tort claim.²⁶ A tort claim, put briefly, is distinguished from contract disputes by there being no requirement

²⁵ 888 F.Supp. 293 (D.N.H. 1994) (granting a preliminary injunction preventing a tenured professor's suspension for comments that offended some students).

²⁶ See Rothstein and Liebman (2003), Chapter 10.B.

for a prior contractual relationship. Standard examples include traffic accidents and medical malpractice. The practical implication of this distinction is that tort law allows for the recovery of both consequential damages and punitive damages, which may far exceed the direct economic harm suffered by the discharged employee. In contract law, consequential damages and punitive damages are in general not recoverable.

2.3.1. Public policy exception

Under the public-policy exception, an employee may sue for wrongful discharge if he is dismissed for conduct that is protected by law. Miles (2000) summarizes the four types of terminations that fit under this class of exception.²⁷ They are (1) “an employee’s refusal to commit an illegal act, such as perjury or price-fixing”; (2) “an employee’s missing work to perform a legal duty, such as jury duty or military service”; (3) “an employee’s exercise of a legal right, such as filing a workman’s compensation claim”; and (4) “an employee’s ‘blowing the whistle,’ or disclosing wrongdoing by the employer or fellow employees”.

A well-known example of the first type of public-policy exception is the 1980 case *Tameny v. Atlantic Richfield Co.*²⁸ Plaintiff Tameny, the dismissed employee, claimed that his discharge resulted from a refusal to participate in the company’s unlawful price-fixing scheme. Defendant Atlantic Richfield argued that since there was no employment contract, Tameny’s employment was at-will and could be terminated at any time. The California Supreme Court ruled for Tameny, holding that an employer cannot discharge an employee for refusing to perform an illegal act. The court further held that the employee can recover under tort law, thereby allowing for potentially higher damages. On this last point, the moral distinction between tort and contract—specifically, that a breach is blameless, but a tort is wrongful—is relevant. Atlantic Richfield did not just breach a contract, it retaliated against an employee for refusing to do its criminal dirty work. Consistent with our moral intuitions, the court considered the company’s conduct to be morally wrongful—not just business as usual—and therefore established a legal mechanism for increased punishment of such conduct.

Economic models of employment mostly ignore illegal activity on the part of employees, yet *Tameny* and other cases involving the public-policy exception clearly demonstrate that some employers do ask employees to commit crimes. The economic implications of this rule are difficult to tease out. In *Tameny*, at least, the employee was asked to engage in anti-competitive activity, so in this case the public-policy exception probably enhanced economic efficiency. But economic evaluations of public-policy cases are generally more difficult. If the illegal activity entails consumer goods, such as drugs or gambling, then the public-policy exception likely decreases output, albeit in a direction that arguably enhances social welfare. The prohibition against discharge for military

²⁷ See page 78.

²⁸ 27 Cal.3rd 167 (Calif, 1980).

service, meanwhile, reduces economic efficiency because it prevents the employer from finding a more productive replacement. As with the illegal-consumer-goods exception, the military-service exception to at-will employment arguably serves other social-welfare goals.

2.3.2. Implied contract exception

When a worker can verify that a permanent employment relationship is promised by his employer, such employment can no longer be regarded as at-will and can be terminated only under just cause.²⁹ Under reigning court precedent in some states, if a personnel manual given to employees specifies that termination is only with cause, a binding contract exists. *Woolley v. Hoffmann-La Roche* was the first opinion to hold that employee handbooks can be part of a legally binding employment contract.³⁰

The facts of *Woolley* are as follows. Plaintiff Richard Woolley was hired by defendant Hoffmann-La Roche, Inc. in 1969 as section head in one of defendant's engineering departments. The parties did not sign a written employment contract, but the plaintiff received a personnel manual which read, in part, that "[i]t is the policy of Hoffmann-La Roche to retain to the extent consistent with company requirements, the services of all employees who perform their duties efficiently and effectively". In 1978, after Woolley's submission of a report on piping problems at one of defendant's buildings, defendant requested that he resign. Plaintiff refused, and he was fired.

The trial court judge held for the defendant on summary judgment. On Woolley's appeal, the New Jersey Supreme Court reversed and remanded the case for trial, holding that an employee's handbook could be evidence of a binding contract. The court couched its ruling in notions of fairness:

All that this opinion requires of an employer is that it be fair. It would be unfair to allow an employer to distribute a policy manual that makes the workforce believe that certain promises have been made and then to allow the employer to renege on those promises. What is sought here is basic honesty: if the employer, for whatever reason, does not want the manual to be capable of being construed by the court as a binding contract, there are simple ways to attain that goal. All that need be done is the inclusion in a very prominent position of an appropriate statement that there is no promise of any kind by the employer contained in the manual...

In this case, as in many others, one party is not completely truthful with the other party. This possibility is ignored by most economic models of contract. Economists typically assume that both parties do what they say they will do, and if they do not, any malfeasance is anticipated by the other party. The *Woolley* opinion can be seen as requiring employers

²⁹ See, e.g., *Toussaint v. Blue Cross & Blue Shield*, 292 N.W.2d 880 (Mich. 1980) ("When a prospective employee inquires about job security and the employer agrees that the employee shall be employed as long as he does his job, a fair construction is that the employer has agreed to give up his right to discharge at will... and may only discharge for cause").

³⁰ *Woolley v. Hoffmann-La Roche, Inc.*, 499 A.2d 515 (N.J. 1985).

to comply with previous agreements not to engage in malfeasance. The judgment does not prohibit dismissal without cause; it simply requires that employers honor promises not to dismiss without cause.

Employee handbooks are not the only example of an implied contract. For example, *Pugh v. See's Candies* held that a long employment with regular promotion can establish a long-term contract.³¹ In this case, the plaintiff-worker Pugh reported to company higher-ups that his current supervisor was a convicted embezzler, for which the supervisor subsequently fired him. Pugh filed suit, but the trial court dismissed the case at summary judgment. On appeal from the dismissal, the appellate court agreed that Pugh's reporting his supervisor's past conviction was not "whistle-blowing" under the public policy exception, but the long duration of Pugh's good service was sufficient to establish an implied contract. The court therefore reversed and remanded the case for trial.³²

This example illustrates a concrete case in which an employee is dismissed not because of an objective failing (otherwise one could provide cause for dismissal) but because, essentially, he did not get along with his new supervisor. If the contract were at-will, then dismissal would be immediate. This rule prohibits the dismissal of long-term employees who may not fit in, or, if delinquent in their performance, the employers are unable to provide sufficient evidence of this poor performance.

2.3.3. *Good faith exception*

The requirement of good faith and fair dealing is a mandatory rule in contract law, and consequently in employment law. The employment cases involving this exception typically turn on the use of at-will employment by the employer to deprive the employee of compensation. In *Mitford v. Lasala*,³³ the discharged employee, who was a party of a profit-sharing agreement with the defendant, was fired to ensure that he would not share profits. The court held that "good faith and fair dealing... would prohibit firing [an employee] for the purpose of preventing him from sharing in future profits".

Currently, courts typically find a rather narrow application of this rule to the timing of dismissal and payment of compensation, rather than to other forms of bad behavior by employers.³⁴ Typical examples of wrongful terminations that fit under this class are: (1) a salesman being fired right before his commissions should be paid to him, and (2) an employee being dismissed in order to avoid paying retirement benefits.

As we can see from Fig. 1 there are many fewer states adopting this law than in the case of the implied contract rule. Given the more narrow applicability of the rule,

³¹ *Pugh v. See's Candies*, 171 Cal. Rptr. 917 (Cal. Ct. App. 1981).

³² At trial, the jury rendered a verdict in favor of the company notwithstanding the appellate court's holding on implied contracts. *Pugh v. See's Candies*, 250 Cal. Rptr. 195 (Cal. Ct. App. 1988).

³³ 666 P.2d 1000 (Alaska 1963).

³⁴ See Section 10.2 of Rothstein and Liebman (2003).

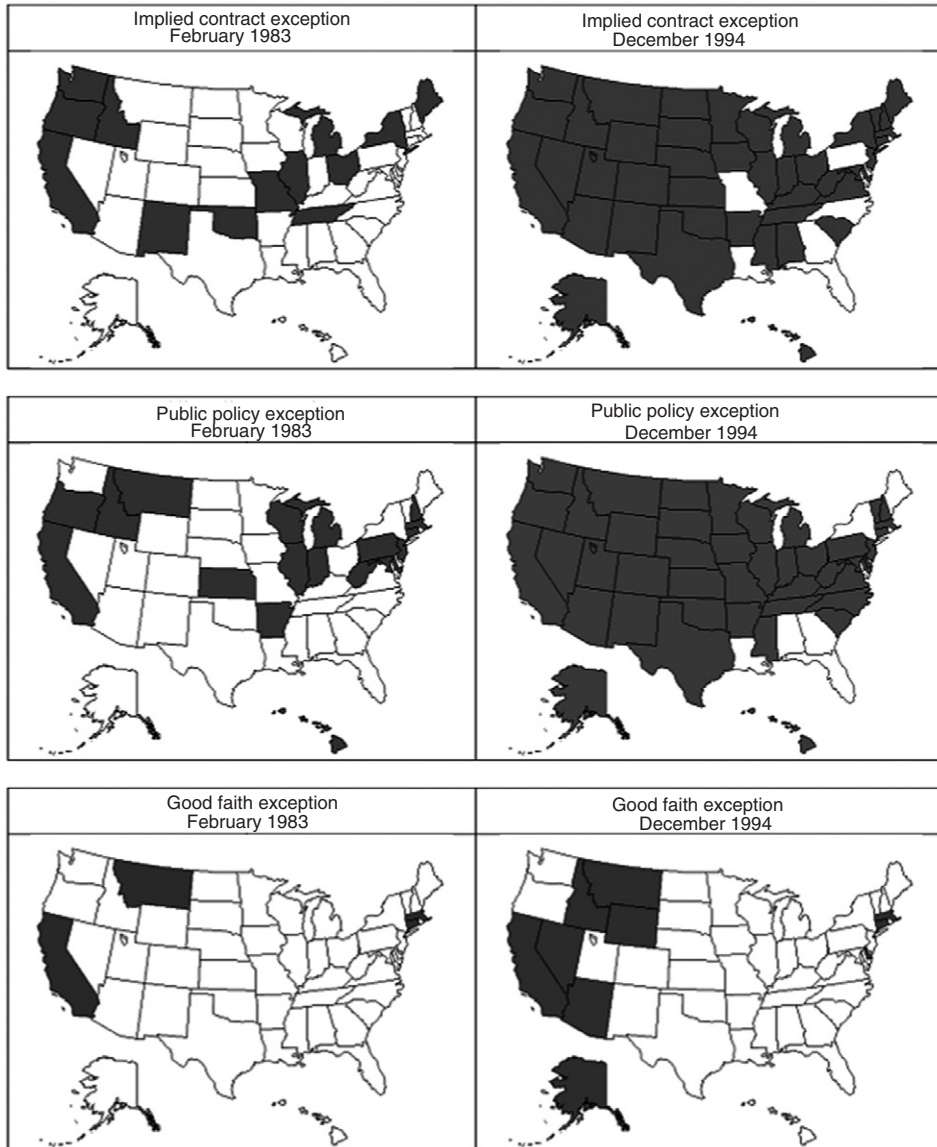


Figure 1 *Pattern of employment law adoption during 1983-1994.*

this may simply reflect the fact that courts in these states have adhered more closely to the common-law principle of at-will employment, and hence there was a need for statutory intervention to deal with cases where employers avoid paying compensation by a preemptive dismissal. If so, then we might expect this rule to have a substantial impact.

This impact is not due to the effect upon firing costs, but rather because it corrects poorly drafted contracts. In the case of *Mitford v. Lasala*, the contract was quite clear, and it implied that the firm had no obligation to pay the bonus. Most employees would expect to be paid in such a case, but at the time of writing the agreement they simply would not expect the deception to occur. In such cases, the courts can enhance productive efficiency by essentially completing an incomplete contract.

Consider now the case of *Fortune v. National Cash Register Co.*³⁵ Plaintiff Orville E. Fortune, a former salesman of National Cash Register Company (NCR), brought a suit to recover certain commissions allegedly due from a sale of cash registers to First National Stores, Inc. Fortune had been employed by NCR under a written contract that provided for at-will mutual terminable with notice. The contract also specified that Fortune would receive an annual bonus computed as a percentage of sales that he performed or supervised. In November 1968, Fortune was involved in a supervisory capacity in a sale of 2008 cash registers to First National, for which the bonus credit was recorded as \$92,079.99. The next month, Fortune was given notice of termination. NCR ended up keeping Fortune on staff in a demoted capacity, and paid him three-fourths of the First National bonus during the summer of 1969. Fortune requested the other 25% of the bonus, but his manager told him “to forget about it”. Fortune was finally asked to retire in June 1970, and then fired upon his refusal.

At trial, the jury was asked to render two special verdicts: “1. Did the Defendant act in bad faith . . . when it decided to terminate the Plaintiff’s contract as a salesman by letter dated December 2, 1968, delivered on January 6, 1969? 2. Did the Defendant act in bad faith . . . when the Defendant let the Plaintiff go on June 5, 1970?” The jury answered both questions in the affirmative, and the judge ordered damages of \$45,649.92. The state supreme court affirmed the judgment.

What is interesting about this case is that NCR did not breach the written terms of the agreement, but the court nevertheless allowed a jury to find that they had acted in bad faith in depriving Fortune of bonuses from the transactions he helped procure. *Fortune* can be seen as an efficient outcome in that it reduces employee uncertainty about whether they will be rewarded for their efforts and thereby incentivizes optimal investment in the employment relationship.

2.4. Discussion

The economic model of contract tends to view legal rules as constraints upon individual decision-making, either in terms of increasing transaction costs or imposing constraints upon the wages, hours, and other conditions of employment. In practice, the law is a complex adjudication system that is difficult to describe with an elegant model. Some of the distinctive features of a legal system that are not captured in the economic model of the employment contract include:

³⁵ 373 Mass. 96 (Mass. 1977).

1. Contract terms are not self-enforcing. Enforcement is a privately motivated activity that occurs when a plaintiff brings a case before a court. Even rules that have bureaus dedicated to their enforcement—such as the minimum wage and overtime requirements—rely on information provided by private parties—as well as the volition of agency officials. This demonstrates that enforcement is heterogeneous and a function of employer, employee, and regulator characteristics.
2. When a case is brought to a court, parties cannot rely upon the courts to enforce the agreement as written. Excessive penalties for non-performance are not enforced, for example. Although employment at will is the default rule in the United States, there are several exceptions.
3. Judges do not restrict themselves to contract terms, explicit or otherwise, as relevant legal factors. Courts may collect a large body of evidence regarding the communications and actions of both parties before reaching a decision. Thus, information regarding events not mentioned in the employment contract may nevertheless play a role in adjudicating the dispute.

The fact that courts may overrule contract terms is well-recognized in the legal literature. One of the central issues of this literature is the question of whether or not there is anything we can reasonably call “the law” that allows one to consistently anticipate how courts will rule on a given dispute. There is a related debate regarding how best to think about judicial behavior.³⁶

Within economics, there is a small but growing literature that explores the role of the law in ensuring performance. Johnson et al. (2002) find that even if enforcement is imperfect, the existence of courts can help entrepreneurs enter into new supply contracts. Djankov et al. (2003) construct a database consisting of how costly it is to evict a tenant and collect on a bounced check across a large sample of countries, finding that the cost of collection in civil-law countries is significantly higher than in common-law countries. This result is consistent with subsequent work reported in Djankov et al. (2008) for the problem of debt collection. See La Porta et al. (2008) for a more comprehensive discussion of this literature.

For the most part, this work focuses on the costs of the legal system and assumes that variations in these costs across jurisdictions affect economic performance. Botero et al. (2004), and more recently Djankov and Ramalho (2009), explore the extent to which employment law and regulation affect labor market performance. This work uses cross-country variation in measures of employment-law flexibility to identify the effects of the law upon labor-market performance. These papers suggest that the historical origin of the country’s legal regime—whether common-law or civil-law—is often the decisive factor in the evolution of the country’s employment rules. However, these papers do not *explain* this observation. One possible interpretation, perhaps in need of further

³⁶ See Stephenson (2009) for a nice summary of this debate.

research, is that laws, like organisms, are adapted not just to the environment but to other laws. The various laws in a legal system—of which employment law is a small part—persist at a steady-state equilibrium unless an overwhelming shock—whether political or economic—suffices to move enough laws to another equilibrium to pull the rest of the legal regime along with them. The rarity of such events—Russia’s transition to capitalism is a plausible example—might explain the durable influence of common-law and civil-law institutions on employment laws.

Regardless of this latter conjecture, what is clear from the analysis above is that the law is adaptive, yet existing work does not adequately explain why there is variation in the law. The plausible view taken here is that the law evolves in response to cases brought before the courts. New types of disputes breed new types of law. To understand why these cases arise, we need to understand what exactly is the role of the law in an employment contract. In the next section, we review the literature on the economics of the employment relationship, placing legal rules in the context of the full relationship.

3. THE ECONOMICS OF THE EMPLOYMENT RELATIONSHIP

Economic theories of employment begin with a model of human behavior and choice. The standard assumption in economics models of employment is that the worker is a risk-averse individual who wishes to maximize expected utility adjusted for the utility from doing specific tasks and the work environment.³⁷ One of the lessons of contract theory is that the optimal contract is often a complex function of the technology of production, the characteristics of the prospective employer and employee, and the information available. In order to highlight the empirical implications of the theory, we begin with a discussion of causality. We then discuss economic models of the employment relationship, highlighting their empirical implications.

3.1. Why do we need models?

The purpose of this section is to review the role that economic theory plays in understanding the significance of the law. As discussed in Section 2, even though economic concerns shape the development of the law, economic analysis as developed by the economics profession has played a relatively minor role in explicitly guiding court decisions.³⁸ While there may be no explicit accounting of economic effects, it is safe to say that employment rules established by courts have measurable effects on the economy. Accordingly, the goal of the theory discussed here is to structure empirical tests of the impact of employment policy on economic performance.

³⁷ See Rebitzer and Taylor (2011) for a discussion of recent research in behavioral economics as applied to labor economics.

³⁸ For example, see Justice Breyer (2009)’s observation that economics plays a small role in US Supreme Court decision-making.

The recent empirical work in labor economics has been greatly influenced by the potential-outcomes framework, as beautifully explicated by Holland (1986).³⁹ I shall briefly review this approach when using economic theory to understand the effects of the law on labor-market performance. First, the framework provides guidance on how to best organize and *represent* data. Second, it provides guidance on how to estimate a causal effect. Holland emphasizes that it is *impossible* to establish a causal relationship without some additional hypotheses that themselves can rarely be tested; they must rely on a *model* of how the world works.

Formally, the model proceeds by supposing that we have a universe of units to be treated, denoted by $u \in U$. For the purposes of our discussion, let U denote all potential workers in the economy. In addition, we might also be interested in outcomes at a state or country level. In that case we let $u^s \subset U$ be the subset of individuals living in state $s \in S$, where s could denote a US state among all states or one country among all countries.

This chapter's main concern is labor-market performance, so we restrict our attention to the question of how policy might affect wages and employment. For individual u , let $y^E \in Y^E = \{0, 1\}$ be employment status (with 1 meaning employed), and let her wage per period be given by $y^w \in Y^W = [0, \infty)$. Suppose that these outcomes will be observed in the next period (t). Employment and wages are likely to be affected by employment policies in the next period, denoted by l_t .

Rubin's model was developed in the context of a medical treatment where one asks if a particular drug has an effect. This question is typically answered by randomly dividing a group of individuals into a treatment and a control group. The causal effect of the drug is measured by comparing the outcomes in the two groups. The problem is that this procedure does not identify the effect of the treatment on a particular individual. In some illnesses, individuals become well in the absence of treatment. For others, the illness may be fatal regardless of the treatment. By chance, it is possible that all the former individuals (the false positives) would be assigned to the treatment group, while the latter individuals (the false negatives) would be assigned to the control group. In that case, the experiment would show that the drug had an effect, even though it did not.

The first issue is how to define a causal effect. In the context of our simple model, let $y(u, l^0, t)$ be the outcome under the status-quo law in the next period, and let $y(u, l^1, t)$ be the outcome under the new rule, say an increase in the minimum wage. Let $\Delta = l^1 - l^0$ denote the policy change. Following Holland (1986), we say that the policy change Δ at date t *causes the effect*:

$$D(u, \Delta, t) = y(u, l^1, t) - y(u, l^0, t).$$

This definition is concrete: It is the difference in potential outcomes. In order to measure this "effect", we would have to *observe* the same outcome for two different policies at the

³⁹ See also Angrist et al. (1996) and particularly Imbens and Wooldridge (2009) for an up-to-date discussion.

same time, something that is clearly impossible without time travel. Holland (1986) calls the impossibility of observing a causal effect the *Fundamental Problem of Causal Inference*. His analysis emphasizes the fact that measuring the causal impact of a treatment entails additional hypotheses.

Most solutions to the problem of causal inference rely upon versions of *unit homogeneity* or *time homogeneity*. By unit homogeneity we mean that there is a set of units $U' \subset U$ with the feature that the effect of the change Δ is the same for all $u \in U'$, in which case the effect can be estimated by policy change to unit $u^1 \in U'$ but not to unit $u^0 \in U'$, in which case for $u \in U'$ we have:

$$D(u, \Delta, t) = y(u^1, l^1, t) - y(u^0, l^0, t).$$

By time homogeneity, we mean that the effect of the treatment in different periods is the same. Hence, if we can estimate the effect of a treatment on a unit u by comparing the effect over time:

$$D(u, \Delta, t) = y(u, l^1, t+1) - y(u, l^0, t).$$

The challenge then becomes finding the homogeneous group. Regression discontinuity is an example of a recent popular technique that provides a way to create homogeneous groups that allow for the estimate of the effect of a treatment.⁴⁰ For example, DiNardo and Lee (2004) argue that firms in closely contested unionization drives are almost identical in most respects. Because the outcomes of union certification votes are very close, one can assume that for these firms union status is randomly assigned. Consequently, we can compare the change in firm value for those firms that were unionized to the change for those that were not, and thereby procure a robust measure of the effects of unionization on a firm's productivity.

Lee and McCrary (2005) provide an example of *time homogeneity*. Specifically, they look at the effect on behavior of sanctions against crime. Their study exploits the fact that when a person turns 18, they suddenly become eligible to be tried in adult courts, where they will face more severe sanctions than a juvenile court would impose. On the supposition that a person's characteristics just before and after they turn 18 are the same, observed changes in crime-related behavior can be ascribed to changes in criminal sanctions.

Notice that all the work is being done by the assumption of continuity over time with the same unit, or across units with very similar characteristics. The great benefit of this approach is that, beyond the continuity assumption (which is a strong assumption), this approach is relatively model-free. The problem is that while it may provide a credible

⁴⁰ See Imbens and Lemieux (2008) and Lee and Lemieux (2009) for a discussion of the technique.

measure of the effect of a policy change, the approach says little if one moves away from the point at which the policy change or treatment is applied.

A formal model in this framework has two distinct goals. The first is that it may provide a concise *representation* of a set of facts about the world. It describes the set of measured characteristics that one needs to know in order to capture the effect of a treatment. For unit $u \in U$, let $X(u, t) \in \Theta$ be a set of characteristics. In practice, one may not be able to measure all dimensions of Θ , but let us suppose for the moment that we can. Suppose u is a worker and we are interested in explaining worker wages. Then, we would say that a model that specifies a wage $f(X, l, t)$ for a worker with characteristics X is an *unbiased representation* of the data at date t if for all $u \in U$,

$$\phi_{ut} = y(u, l_u, t) - f(X(u, t), l_u, t)$$

is an iid set of random variables with zero mean.

If our model is linear, we can let $\beta(t, l) = \partial f / \partial X$, in which case we can write our model in the familiar regression form:

$$y_{ut} = \beta(t, l_{ut})^\top X_{ut} + \phi_{ut}.$$

If our model is unbiased, then this is a well-specified model that can be estimated by ordinary least squares. However, even if the model is well-specified, as Holland (1986) emphasizes, the coefficients of the model *cannot* be assumed to represent a *causal* relationship. For example, one of the parameters might be the gender of a worker, say 0 is male and 1 is female. If y_{ut} is the wage, and the coefficient on gender is negative, we cannot say that gender *causes* a wage drop. This is because gender is not a treatment or something that one normally assumes can be varied within a person.

We can use the coefficient on gender to test various theories. For example, human capital theory predicts that a person's wage is a function only of their productive characteristics, such as schooling, ability, and experience. One reason women might be paid less is that they spend more time out of the labor force in child rearing. This reasoning implies that once the full set of characteristics reflecting productivity is included in X , then the coefficient on gender should be zero. If it is not, then we can say there is discrimination in the labor force.

A good theory specifies the set of parameters X that provide all the information necessary to describe wages while preserving *time independence*:

$$y_{ut} = \beta(l_{ut})^\top X_{ut} + \phi_{ut}.$$

Any variation in wages that occurs over time is explained via either changes in the parameters X_{ut} or by changes in the environment l_{ut} . In practice, the econometrician

may not have access to all the relevant information X_{ut} , which leads to the well-known omitted variable bias problem in econometrics. For the present discussion, let us suppose that the relevant data are available and ask how the model can help in measuring the causal impact of a change in law l .

In general, economic theories do not provide precise point predictions; more typically, they make predictions about the sign of an effect. In the context of measuring the effect of the law on outcomes, the variation in treatment typically occurs either across jurisdictions—namely, the experiment assumes that all individuals in a particular jurisdiction $u \in U_s$ face the same legal environment l_{st} , and it is the legal environment that varies across jurisdictions. For example, many countries can be characterized as civil-law or common-law legal systems. We can let U_1 be individuals in common-law countries and U_0 be individuals in civil-law countries, and set $l_s = s$.

In the example of civil- and common-law countries, one could estimate $\beta_s = \beta(l_s)$ for each jurisdiction. In this case, the causal impact of the legal system depends on the distribution of characteristics of individuals in the economy. We would estimate the causal effect of changing from a civil-law system to a common-law system for regions that are currently under civil law in period t by:

$$\text{Effect of common law} = \frac{1}{n_0} \sum_{u \in U_0} (\beta_1 - \beta_0)^\top X_{ut}.$$

In order to estimate the causal effect of a change in the legal system, one needs to use the characteristics of the jurisdiction where the change is to occur. This adjustment is a version of the well-known Oaxaca decomposition, which is widely used in studies of income inequality (see Altonji and Blank (1999)) and union wage differentials. As we discuss in more detail in Section 4, this is not the literature's usual technique. The more common assumption is that the effect of a policy is linearly separable, where for $u \in U_s$ we have:

$$y_{ut} = \beta^\top X_{ut} + \beta_l^\top l_s + \phi_{ut}. \quad (1)$$

Using data for a single period t , then, we can estimate the average effect of the legal system on the wages and employment of individuals by:

$$\beta_l = \frac{1}{n_1} \sum_{u \in U_1} (y_{ut} - \beta^\top X_{ut}) - \frac{1}{n_0} \sum_{u \in U_0} (y_{ut} - \beta^\top X_{ut}).$$

The goal of the theory discussed in this section can be summarized as follows. The theory makes predictions regarding the characteristics X that are needed to represent individual outsources. In particular, it will provide predictions regarding how variations

in individual characteristics relate to variations in outcomes. Theory has predictive power if we can safely assume that the relationship between the X s and the y s is stable over time.

A theory has more predictive power if one can represent outcomes using a smaller set of X s. Given the difficulty of obtaining good measures of individual characteristics, theories with fewer X s are inherently easier to test. On a related note, there is a line of inquiry in statistics that attempts to be model free. This is achieved by supposing that one has a rich set of X variables and that the environment is inherently continuous; as a result, good representations of the data can be used to make predictions on how changes in an individual's X s will affect outcomes. Breiman (2001) suggests that such an approach is sometimes more feasible given present computing resources and the large data sets we have in some domains.

However, representation is not causation. Many individual characteristics are not amenable to experimental treatment. Making causal statements requires that we assume we have a valid representation of the data that allows one to compare outcomes either: (1) across units with similar characteristics but in different treatments, or (2) the same units faced with different treatments over time. In these cases, the theory—in addition to specifying the relevant X variables—also specifies an explicit mechanism by which the law affects the actions of individuals, and hence how one can obtain a valid measure of the causal impact of a change.

3.2. Economics of the employment contract

This section discusses the literature that seeks to explain the form and function of employment contracts. Fundamentally, parties who enter into a contract have agreed to have their behavior constrained in the future. The most basic reason for a contract is to support inter-temporal exchange, something that cannot be avoided in the context of labor services. For example, a day laborer agrees to work eight hours in exchange for a wage at the end of the day. When it comes to paying, the employer may have an incentive to renege or attempt to reduce the agreed-upon wage. One role of the law is to enforce such agreements.

In economics, such simple exchanges are typically assumed to be enforceable. The literature has focused on explaining the form of observed contracts that address one of three more subtle issues:

1. *Risk*. Demand for a worker's services, and hence wages, is likely to change from period to period. Risk-averse workers would like to enter into long-term contracts that would shield them from such shocks.
2. *Authority and Asymmetric Information*. Decision-making and bargaining are costly under asymmetric information. In this case, contract form can affect performance.
3. *Reliance*. Once a worker-firm match has been formed, a contract is needed to ensure that each party makes the appropriate investments into the relationship.

What makes the study of employment contracts difficult is that every relationship has elements of these three ingredients. In particular, teasing out the empirical implications of these models has proven difficult. Nonetheless, much of the structure of the legal rules governing employment can be understood as an attempt to address risk, information asymmetries, or hold-up.

3.2.1. Insurance

We shall illustrate these ideas using a simple three-period employment model. Suppose that in period 0 the firm offers the worker a long-term employment contract C , which the worker can either accept or reject. If the worker does not accept the contract, then in period 1 she will earn w_1^0 and in period 2 she will earn ω_2^0 , which is a random variable with mean m_2^0 and variance σ^2 . We will let the realized value of ω_2^0 be w_2^0 . The utility of the worker in this outside market is given by:

$$U^0 = u(w_1^0) + \delta E\{u(\omega_2^0)\}.$$

The expected lifetime income of the worker is:

$$W^0 = w_1^0 + \delta m_2^0.$$

Given that the worker is risk-averse, a risk-neutral firm who wished to hire this worker for two periods could do so by paying a fixed wage w^* per period at an expected cost of $W^0 - \delta \frac{r}{2} \sigma^2$, where $r = -u''(m)/u'(m)$ is the coefficient of absolute risk aversion for the worker.⁴¹ Let us consider the case in which the owner of the firm is assumed to be able to fully diversify market risk, and hence is able to offer the worker a perfect risk-sharing contract. The firm would be willing to do this because she can offer a wage contract to a risk-averse worker that has a lower expected cost than the worker's market alternative.

Azariadis (1975) introduced the term *implicit contract* to describe the idea that firms voluntarily smooth workers' income over time in order to lower expected labor costs. Azariadis (1975) and Baily (1974) both observe that the enforceability of these contracts depends upon the existence of turnover costs, otherwise under at-will employment wages would necessarily equal the market alternative. Recently, Blanchard and Tirole (2008) have revisited this issue and suggested that mandated severance pay may enhance the risk-sharing properties of labor contracts. They explore the question of how to optimally design minimum wages and severance pay to insure risk-averse workers. There is little role for the law in their model beyond enforcing the agreed-upon severance payments.

⁴¹ The wage is $w^* = \frac{1}{1+\delta} (W^0 - \delta r \sigma^2 / 2)$.

We consider a two-period extension of their model that will allow for a substantive role for the courts. Rather than supposing that the employment contract is implicit, I follow Blanchard and Tirole (2008) and consider the problem of implementing the optimal allocation. Suppose that in period 0 it is efficient for the worker to contract with a firm with the following profit function:

$$\Pi = y_1 - w_1 + E\{\psi_2 - \omega_2\},$$

where y_1 and ψ_2 is firm output in periods 1 and 2, while w_1 and ω_2 is the wage paid to the worker in each period. Again, Greek letters refer to random variables.

We begin with a case that entails no enforcement problems, and characterize the empirical implications of the optimal allocation. In this case, we do not have any explicit treatments—rather, we wish to describe the wage and employment profile of the worker (the y 's of the model) in relation to the worker's outside options, the worker's risk preferences, and the firm's productivity in each period (the explanatory X variables of the model).

The optimal allocation is the solution to:

$$\max_{w_1, \omega_2, e_2} y_1 - w_1 + \delta E\{e_2 \psi_2 - \omega_2\}, \quad (2)$$

subject to:

$$u(w_1^0) + \delta E\{u(e_2 \omega_2 + (1 - e_2) \omega_2^0)\} \geq u(w_1^0) + \delta E\{u(\omega_2^0)\}. \quad (3)$$

In addition to wage payment each period, the optimal allocation must also determine the worker's employment status in period 2, where $e_2 = 1$ if employed at the firm and 0 otherwise. If the worker is not employed, she earns ω_2^0 in the market, which is assumed to be paid to the worker. The next proposition characterizes the first best:

Proposition 1. *The optimal risk sharing allocation has the following properties:*

1. *Employment is ex-post efficient: $e_2^* = 1$ if and only if $\psi_1 \geq \omega_2^0$ (and zero otherwise).*
2. *The worker is fully insured: $w_1^* = w_2^* = w^*$.*
3. *Expected labor cost is equal to the worker's expected future income less a risk premium that increases with worker's aversion to risk (r): $(1 + \delta)w^* \simeq W^* - \delta^r \sigma^2$.*

Observe that this result provides an *institution-free* description of the optimal allocation that links the characteristics of the optimal contract with potentially observable features of the worker, firm, and labor-market alternatives. In practice, there are a large number of institutions that have been created to insure workers, including workman's compensations, unemployment insurance, transfers within the household, and so on.

Rather than delve into the details of these institutions, one may ask the question whether or not institutions are sufficiently rich that something approaching efficient risk-sharing occurs in practice.

Cochrane (1991) works out the implications of the complete risk-sharing model for consumption growth. In our simple model, notice that the wage of a worker (here consumption is assumed equal to wage for simplicity) is independent of whether she works for the current firm or takes up a market alternative (which might mean unemployment). More generally, Cochrane (1991) observes that consumption growth should be independent of idiosyncratic shocks. He finds that full insurance is not rejected for spells of unemployment, loss of work due to strikes, and involuntary moves. However, insurance appears to be incomplete for long illnesses and involuntary job loss. In recent work, von Wachter (2007) finds with German data that the effect of job loss is temporary, with workers returning to their previous earnings in 5 years. Taking a similar approach with data from India, Townsend (1994) rejects the perfect insurance model but does find evidence of significant, albeit imperfect, risk sharing.

This work illustrates the usefulness of the insurance model in organizing consumption data. The work does not test a *causal* relationship, nor does it describe a mechanism that would generate a relationship between wages and risk attitudes. That mechanism could involve, for example, firms setting wage contracts in advance and workers selecting into contracts that are most appropriate for their risk preferences. Alternatively, firms might negotiate contracts directly with the workers and use worker-specific information, such as marital status, to set the wage contract. In addition, the implicit contract model does not explain wage rigidity per se—and certainly not nominal wage rigidity (see Card and Hyslop (1997))—only consumption smoothing. If a worker has access to other insurance opportunities, say via their family, then their actions might appear less risk-averse, implying that there may not be a stable relationship between an individual's risk preference and the wage contract.

Consider the question that Blanchard and Tirole (2008) ask, namely: How can one *implement* the efficient allocation using available legal instruments? The answer to this question can generate some predictions about the effects of changes in the law or in the parameters of social programs such as unemployment insurance. These predictions are causal statements because the choice of law is a treatment; we can ask explicitly what the *causal impact* of a policy change will be.

Insurance contracts require some form of enforcement when employment with the firm in period 2 is efficient. Whenever $\psi_2 < w^*$ the firm would like to dismiss the worker, while the worker would like to quit whenever $\omega_2^0 > w^*$. Given that it is always efficient to perform, the parties would never voluntarily renegotiate the contract price. If the contract were between commercial parties for services, termination of the relationship by either party could be followed by a suit for damages. Under the rule of common law the standard remedy is *expectation damages*—harm caused by the contract

breach. Let us suppose that there is a cost k in pursuing a court case. If the employment contract is adjudicated under standard common-law rules, we would have the following outcomes for contract termination⁴²:

State of the world	Breach decision	Remedy
$\psi_2 < w^* - k$	Firm lays off worker	Worker paid $D = w^* - \omega_2^0$
$\omega_2^0 - k > w^*$	Worker quits	Firm paid $D = \psi_2 - w^*$

Observe that under this rule, if it is efficient for parties to stay together (it is always the case that $\psi_2 - \omega_2^0 > 0$), then there would never be breach. For example, suppose that the worker's outside option is so great that it is worthwhile to quit even while paying legal fees, namely $\omega_2^0 - k > w^*$. Once she pays the damages to the firm, her income would be:

$$w^* - k - (\psi_2 - \omega_2^0) < w^*.$$

Hence, under the standard legal rule of expectation damages, an employment contract that fully insures workers would be enforceable and would implement the efficient contract when turnover is not efficient. In practice, however, these rules are rarely used in employment cases. The common-law rule is employment at will, not expectation damages.

Parties might try to achieve a binding contract by stipulating that each party would pay a large fine F if there is breach. In practice, requiring workers to pay large penalties to leave employment are not enforceable in most legal jurisdictions—this would be akin to a slavery contract. One exception is requiring a worker to pay for training she has received. In the case of professional sports, this goal is achieved by requiring teams to pay a fee to acquire a player. This rule has become controversial, though, and was recently overturned by the European Court of Justice.⁴³

The case of sports teams is the exception. For most employment contracts, employees can leave at will. There is literature, beginning with Harris and Holmström (1982), that explores the optimal wage contract under the assumption that the firm cannot fire the worker, but the worker can leave at will. Under such a rule, the optimal contract is downward rigid. It is fixed in real terms and readjusted upwards each time a worker gets an outside offer. Beaudry and DiNardo (1991) suggest that this model can explain why individuals who are hired during recessions are worse off in the long run than workers

⁴² I assume that the terms of the wage payment are enforceable, so it is only the decision to quit or layoff that is liable to legal recourse. There is a well-known UK case, *Rigby v. Ferodo* [1988] ICR 29, House of Lords, that establishes the enforceability of the wage payment.

⁴³ This is the so-called *Boseman* case. See Feess and Muehlheusser (2002) for a discussion.

hired in boom periods. Chiappori et al. (1999) point out that there may be other reasons for this result, including holdup (which we discuss below).

3.2.2. Asymmetric information and the employment relationship

Consider the following extension of Simon (1951)'s employment model, allowing for task allocation *ex post* in the presence of asymmetric information. Suppose that in period 2 the worker can be assigned to one of two tasks, $x \in \{a, b\}$, and that the productivity of task x is ψ_2^x , which is assumed to be observed only by the firm. In addition, there is a private cost of carrying out task x to the worker given by $c_2^x > 0$, and which is observed only by the worker. If we let $x = 0$ denote the outside option, with $\psi_2^0 = 0$ and $c_2^0 = -\omega$, then the respective payoffs to the firm and worker under task x are:

$$\begin{aligned}\Pi^x &= y_1 - w_1 + \delta I_x \{ \psi_2^x - \omega_2^x \}, \\ U^x &= u(w_1) + \delta u(\omega_2^x - c_2^x),\end{aligned}$$

where $I_x = 1$ if $x \in \{a, b\}$ and 0 otherwise.

Note that regardless of the contract, the optimal task allocation is given by:

$$x^*(\theta_2) = \arg \max_{x \in \{0, a, b\}} \{ \omega_2^0, \psi_2^a - c_2^a, \psi_2^b - c_2^b \}.$$

Let us first suppose that it is efficient for the worker and the firm to stay together both periods. Also suppose that there is no variation in task productivity, but that the worker's cost, c_2^x , can vary. In this case, the efficient solution is to allocate the choice of task to the worker, who will always choose the efficient allocation. More generally, as Milgrom (1988) argues, this effect leads to an organizational structure that limits the authority of the firm, so that within certain task groups individuals are given autonomy.

Given that the worker is risk-averse, the optimal solution entails a fixed wage and an allocation of decision rights to the worker. If $w^* > \psi_2^a$, $\psi_2^b > \omega_2^0 + c_2^{a,b}$, it is efficient for the worker to stay matched with the firm, but the firm would prefer to lay the worker off rather than pay wage w^* . Enforcing the efficient contract requires a stipulated severance pay that is sufficiently large but conditional upon worker's performance. Similarly, the efficient contract should ensure that the worker does not threaten to renegotiate the wage contract in period 2. These points are summarized in the following proposition:

Proposition 2. *If the firm is indifferent over task assignment ($\psi_2^a = \psi_2^b$), then the optimal contract has the following conditions:*

1. *The worker has the right to choose her preferred task.*
2. *The contract wage each period is decreasing with the worker's risk aversion, increasing with minimum cost of effort and expected lifetime market income ($w_1^* = w_2^* - \min\{c_2^a, c_2^b\} \simeq (W^0 - \delta \frac{r}{2} \sigma^2 - \delta \min\{c_2^a, c_2^b\}) / (1 + \delta)$).*

3. If the worker leaves in period 2, then she pays a penalty $P > \omega_2^0 - w_2^* + c$. If the firm dismisses the worker, it pays $f > \max\{\psi_2^a, \psi_2^b\} - w_2^*$ in severance.

This contract is very much like the contract for a tenured academic. The contract asks the professor to carry out teaching duties but typically allows a great deal of discretion over how she teaches and the material she will use. Second, the demand for the services of an academic is stable, and hence there is little benefit from turnover. As a consequence, the academic cannot be fired. The optimal contract also precludes the worker from leaving without paying a penalty, but this sort of provision is not typically observed (aside from the sports contracts mentioned earlier). For academics, the resignation penalty is implicit, consisting in large moving costs, lowering the incentives to leave. Note that a consequence of removing the penalty clause for leaving is that the period 2 wage would be more responsive to the outside market, and as a consequence the first-period wage would *fall*. Conversely, sometimes it is suggested that tenure be abolished. The consequence of abolition would be to lower expected income in period 2, which in turn would *raise* period 1 wages.

In terms of empirical predictions, this result merely links X variables—the risk aversion of the worker and job characteristics—to predicted contract choice. Hence, this proposition does not make any causal claims, but predicts that there should be an association between measured risk aversion of the worker, job characteristics, and turnover. It predicts that certain jobs, such as academic jobs, combine substantial job protection with the freedom to control activities on the job.

In order to introduce a notion of just cause for dismissal, there needs to be a substantive role for the firm in task allocation. The next case supposes that effort costs do not vary with the task, $c_2 = c_2^a = c_2^b$, but the productivity of the tasks vary, $\psi_2^a \neq \psi_2^b$. Let us continue to suppose that it is always efficient to be employed at the same firm for two periods. In this case, it is efficient to provide the worker with a fixed wage contract, $w_1^* = w_2^* - c_2$. It is crucial that the worker and firm *not* negotiate the task allocation. If the wage paid for each task is the same, and the firm has the right to make the task allocation $x \in \{a, b\}$, then it will choose the most productive task. Hence, this contract is *incentive-compatible* in the sense that the firm will make the most efficient choice even though she holds private information. In order to provide the firm with authority over the worker, there must be a penalty associated with not following the firm's instructions. More formally, we have:

Proposition 3. *Suppose that employment with the firm is always efficient. Then the optimal contract consists of a fixed wage each period along with the following conditions:*

1. *The contract wage each period is decreasing with her risk aversion and increasing with cost of effort ($w_1^* = w_2^* - c_2 \simeq (W^0 - \delta_2^r \sigma^2 - \delta c_2)/(1 + \delta)$).*
2. *Should the firm dismiss the worker without cause, the worker is paid $f > \max\{\psi_2^a, \psi_2^b\} - w_2^*$ in severance.*

3. If the worker is indifferent over task assignment ($c_2 = c_2^a = c_2^b$), then she agrees to carry out the task assigned by the firm; otherwise she is dismissed and pays a penalty.

This proposition describes the features of an optimal contract when the worker is risk-averse and separation is not efficient. It is useful because it captures some features of observed default rules in employment law. Notice that the firm has authority because it has the information regarding the best task to carry out. More generally, Aghion and Tirole (1997) have shown that authority should be allocated to the best-informed individual.⁴⁴ Dessein (2002) extends this point to look at the trade-off between communication and delegation, finding that delegation can be more efficient than communication when there is little conflict between the preferences of the worker and firm.

The provision of insurance via wages does create potential conflict. The firm must have the right to penalize workers who do not carry out their assigned tasks. But this power cannot be unchecked. For example, the firm might try to renege upon the wage contract by assigning a worker very unpleasant tasks—formally, those with a high cost of effort c_2 —that would effectively cause the worker to quit. Such a case might lead to litigation where the worker would claim *constructive dismissal*.⁴⁵ Hence, in practice, such a contract may still face significant litigation.

The results above suggest that if labor contracts are incomplete, with parties relying upon the courts to set the default terms, then both propositions predict that a change from at-will to just-cause dismissal will lead to lower wages and higher employment. Higher employment occurs because just-cause dismissal is more efficient in these cases, and hence should increase employment.⁴⁶ Wages are lower because the worker faces less risk.

Note that the employment result does not fundamentally depend on worker's risk aversion. The employment law we have discussed builds upon contract law, where the key issue is ensuring that parties deliver the promised quality. Disputes arise when firms feel that workers have not performed as promised, or workers believe they have performed as promised but the firm has not compensated or continued employment as promised. This set of issues is legally distinct from the body of law that has developed to enforce insurance contracts, and accordingly the doctrines regarding damages in employment law rarely entail an explicit discussion of risk.

Finally, there is literature, beginning with Shapiro and Stiglitz (1984), that views the right of dismissal as a necessary ingredient for effort provision. In their model, the

⁴⁴ See also Chakravarty and MacLeod (2009), who discuss the allocation of authority in the context of contract law. They show that construction contracts carefully allocate authority between the buyer and seller to ensure efficient production.

⁴⁵ This is a legal term of art in English law defined by the UK Employment Rights Act of 1996, Sections 95(1)(c). Even if an employee resigned from her post, she can claim that she was forced to quit due to the employer's action.

⁴⁶ See MacLeod (2005) and MacLeod and Nakavachara (2007) for more details on how employment law may enhance efficiency.

firm offers a high wage and threatens to dismiss the worker should she shirk. In this model, this results in an inefficient allocation due to the high wages offered by the firm. However, as MacLeod and Malcomson (1989) show, the threat of firing is not necessary for effort provision. The firm can use bonus pay, in which enforcement depends upon the firm facing a cost should it renege upon a promised payment. MacLeod (2003) extends this result to the case of a risk-averse worker employed with an imperfect performance measure.⁴⁷ He shows that a necessary condition for the implementation of an efficient contract is the ability to impose a cost upon firms that renege on bonus pay. The good-faith exception to employment at will is one mechanism that may achieve this condition.

3.2.3. *The reliance interest*

In a famous paper, Fuller and Perdue (1936) introduced a conceptual framework that has formed the basis of the modern law-and-economics treatment of contract law. The goal of their paper was to provide a framework for the setting of damages for contract breach. They introduced three ways to measure damages. The first of these, as discussed earlier, is expectation damages. This is the rule that one would use if one wished to enforce an insurance contract because it ensures that each party obtains the desired outcome while ensuring that matching is efficient.

The notion of expectations is not always well defined, particularly in the case where the value of the worker's performance is private information. Another way of measuring damages is the notion of restitution. This damages rule strives to restore the harmed party to the state she was in before the contract was agreed upon. For example, suppose a worker sells her house and moves to a new city in order to take up employment. If the potential employer reneges on the contract, restitution damages would entail paying the harmed worker the costs of relocation so that she may return to her previous state.

A third measure of damages is the *reliance interest*. Take, for example, an employer that spends a significant amount of money training a worker, as in the military providing pilot training. In that case, if the worker were to leave employment early, the employer may ask the worker to repay part of the training expenses.

The early literature on the economics of contract law, notably Rogerson (1984) and Shavell (1984), consider the case in which parties make a sunk investment into a relationship, and then ask, which of this damage rules leads to the most efficient level of investment. This work illustrates an important third motivation for an employment contract: namely, to provide incentives for efficient relationship-specific investments. The early literature assumed that the investment was observable by the courts, and hence could be used to set damages. In an influential paper, Grout (1984) showed that if parties could not write a binding contract, then there would be inefficient investment into the relationship. This model has become the paradigm for the *holdup problem*, a term coined

⁴⁷ See Levin (2003) and Fuchs (2006) for a more detailed analysis of the risk-neutral case in a repeated-game setting.

by Goldberg (1976) to describe situations in which the buyer or seller attempts to change the terms of an agreement after have there been significant sunk investments. Williamson et al. (1975) similarly make the point that relationship-specific investments imply that the employment relationship must be carefully governed to avoid opportunistic behavior by the worker.

These points can be formally illustrated in our model by supposing that the employer makes an investment into capital k in period 1, while the risk-neutral worker makes a similar investment i . In this case, the worker's investment can be any activity that lowers the cost of supplying labor, which might include making friends, investing in a new home, or acquiring skills on the job. Formally, the payoffs of the firm and the worker would be:

$$\begin{aligned}\Pi &= y_1 - w_1 - k + \delta E \{e_2(\psi_2 + y(k)) - \omega_2\}, \\ U &= w_1 - i + \delta E \{\omega_2 + e_2 v(i) + (1 - e_2)\omega_2^0\},\end{aligned}$$

where the notation is as above, except now worker productivity depends on investment k via $y(k)$, and worker utility depends upon her investment i via the $v(i)$. It is assumed that $y(0) = v(0) = 0$, $y', v' > 0$, and $y'', v'' < 0$, and that the efficient levels of investment are characterized by:

$$v'(i^*) = y'(k^*) = 1/\delta\rho_2^*$$

where ρ_2^* is the probability that the worker and firm trade in period 2 under efficient matching (namely $\rho_2^* = \Pr[\psi_2 \geq \omega_2^0]$).

Observe that the level of investment into a relationship depends upon both the discount rate and the expectation that the relationship will continue. This implies that if a worker, for example, overestimates the likelihood that an employment relationship will continue in period 2, this can lead to over-investment, and an increased incentive to litigate discharge should she believe it to be unjustified.

The holdup problem arises when the worker and firm have no binding labor contract but instead negotiate the wage in period 2 *after* the value of their outside options have been realized. Grout (1984) supposes that period 2 wage is given by the Nash bargaining solution, which entails parties dividing evenly the gains from trade to yield a wage:

$$w_2(\psi_2, \omega_2, k, i) = (\psi_2 + y(k) - v(i) - \omega_2^0)/2.$$

When this wage is negative, parties will choose the outside option rather than trade. This rule ensures efficient matching in period 2, but the returns from the specific investments are divided equally between the worker and the firm. In consequence, we have *underinvestment*:

Proposition 4. *In the absence of a binding employment contract, the worker and the firm choose investments to satisfy:*

$$y'(k^{nc}) = v'(i^{nc}) = 1/\delta\rho_2^{nc},$$

where the probability of employment in period 2, $\rho_2^{nc} < \rho_2^*$, is less than the efficient level, and hence investments are less than the first best ($k^{nc} < k^*$ and $i^{nc} < i^*$).

The motivation for Grout's model is the legal rule in the United Kingdom that makes it impossible for unions to enter into binding contracts with employers. The substance of the Trade Disputes Act of 1906 made it impossible for employers to sue unions, and hence to recover damages should a union strike.

To predict the causal impact of such a policy, one needs to work out what would happen if contracts were enforceable. The holdup model supposes that investments are observable by the two parties but cannot be used to set contract terms. Under this assumption, Hart and Moore (1988) show that parties would agree to a contract with a stipulated wage w_2 and severance payment s_2 that would improve upon no contract. In general, however, the contract will not implement the first best. This can only occur in this model if it is always efficient to trade, and there is a contract that, with probability 1:

$$\begin{aligned}\psi_2 + y(k^*) - w_2 &\geq -s_2, \\ w_2 + v(i^*) &\geq \omega_2^0 + s_2.\end{aligned}$$

For this contract to work, one does need legal enforcement. If either the worker or the firm attempts to modify the contract terms, the other party should be able to seek relief in court. This is not to say that parties cannot, if they wish, renegotiate the contract by mutual consent. Given that both parties are better off under the contract than on the outside market, however, the threat not to trade is not credible. Hence, the wage would not be renegotiated in period 2, and both parties receive the full return from any specific investment.

Hart and Moore (1988)'s result that contracts can always improve matters holds only for the case of specific *self-investments*: the investments that affect one's own payoff but not the other party's payoff. Che and Hausch (1999) show that in the case of cooperative investments—that is, when one party's investments affect the payoffs of both parties—and when contract renegotiation cannot be precluded, then there is no benefit from writing any contract. This result depends on the hypothesis that the courts cannot observe the investments. Given the level of litigation in employment, one must conclude that parties do indeed find it useful to write contracts. The issue is how the courts should enforce these contracts.

An interesting feature of the holdup model is that the efficiency of the relationship can be enhanced in some cases with the appropriate allocation of bargaining power. This

idea begins with the so-called property-rights approach of Grossman and Hart (1986). They observe that even though contract may not be explicitly conditioned upon certain events, the law can allocate residual decision rights. The example they explore in detail is property, which in effect is a contract that gives the owner of property the right to carry out any action that is not constrained by other contracts.

We have a similar issue in employment law. That is, under what conditions does the worker or the firm have the right to leave a relationship based on information that may not be observable by the courts? Aghion et al. (1994) show that if one can design contracts to allocate the bargaining power of parties, then one can achieve an efficient allocation in the models of Grout (1984) and Hart and Moore (1988).⁴⁸

These are useful abstract results that delineate conditions under which efficient allocations can be achieved, but they do not specify the legal institutions that would achieve these allocations. MacLeod and Malcomson (1993) explicitly explore the implications of the holdup model on wages over time when the market alternatives are viewed as an outside option in the sense of Shaked and Sutton (1984). The outside option principle has two parts.

First, if at the current, *enforceable* wage both parties are better off than at their next best alternative, then threats to leave/layoff are not credible and hence the wage is *insensitive* to current market conditions. As Howitt (2002) observes, this observation has the potential to provide a theory of rigid wages. Second, when the current wage is worse than, say, the worker's best alternative, then either the wage will be renegotiated to be equal to this alternative, or the worker will leave.

Given these rules, MacLeod and Malcomson (1993) show the following:

1. When investments are *general* and there is a fixed cost to changing jobs or employees, then a fixed wage contract that is renegotiated to match outside offers implements the first best.
2. In the case of two-sided self-investments, if it is possible to index wages so that the outside options are binding only when separation is efficient, then such an indexed wage contract implements the first best.
3. In the case of cooperative relationship-specific investments by the firm, an efficient allocation is implemented with a contract that leaves the worker indifferent between employment and taking up the outside option. Conversely, if the worker is making the investment, then the efficient rule leaves the firm indifferent between hiring the worker and taking up the outside option.

These three cases are not comprehensive, however. For example, Rogerson (1992) shows in a more general, asymmetric-information setup that there is a wide variety of situations where there exist efficient contracts when both parties are risk-neutral. The holdup

⁴⁸ These results build upon the implementation results of Moore and Repullo (1988).

model is attractive because it provides some predictions on contract form when parties approximately satisfy these contracts.

What is particularly interesting about these results is that they are broadly consistent with the doctrine of employment at will. The first case merely requires that the worker and the firm agree to some wage contract that can be periodically renegotiated. In particular, the wage can be in real or nominal terms, and hence, as Howitt (2002) points out, can explain nominal wage rigidity. Though the model also predicts that nominal wages may be renegotiated up or down by arbitrary amounts, depending upon the outside market, a behavior that is consistent with the evidence of Blinder and Choi (1990), McLaughlin (1994) and Card and Hyslop (1997), but not consistent either with menu-cost models or the model with risk-averse agents.

These models do rely upon the legal enforcement of a contract wage that cannot be unilaterally changed by one party, a principle that was affirmed in the United Kingdom by *Rigby v. Ferodo* (1987).⁴⁹ We also observe the use of indexed contracts, particularly union contracts: Cousineau et al. (1983) document the use of indexed contracts by Canadian unions. Notice that the risk-sharing model would predict fixed real wages, with corresponding penalty clauses to enforce the risk-sharing agreement. The fact that penalty clauses are typically not enforceable, especially in the case of employment contracts, leads to the prediction that if parties are going to index, then the indexed contract should approximately follow the market wage, which is what we observe in Canada (and also in the case of long-term supply contracts, as documented by Joskow (1988)).

Crawford (1988) shows that the holdup problem can also be solved when parties sign a series of short-term contracts. Essentially, parties anticipate the future holdup and mitigate its effect by agreeing in the current period to lower wages combined with higher investment. Card et al. (2010) find some evidence in support of this prediction using Italian data. Given that these are unionized firms, this suggests that the firms are able to reach efficient bargains.

The final case is implemented in the absence of any contract, and can help explain the puzzling fact that contracts do not consistently have index terms. As Cousineau et al. (1983) have shown, about 50% of unionized firms in Canada did not index their employment contracts during a period of high inflation. This would imply that unions would have to constantly renegotiate their contracts to match market conditions. Result 3 implies that this contract form provides first-best incentives for the firm to invest in capital and into relationship-specific worker training. Acemoglu and Pischke (1998) also show that if there are significant turnover costs, then firms may also invest into general human capital.

⁴⁹ [1988] ICR 29, [1987] IRLR 516.

3.3. Implementing the efficient employment contract in a market

The literature on the employment contract has identified three broad economic motivations for an employment contract: insuring risk-averse employees, ensuring the revelation of relevant information for decision-making, and encouraging relationship-specific investments. Given these transaction costs, the next issue is: What sort of labor-market institutions ensure efficient matching and trade?

In principle, one could construct a model that includes all the ingredients that have been identified as relevant for understanding employment. At a purely abstract level, Rogerson (1992) and Aghion et al. (1994) have shown that under the appropriate conditions, one can construct an abstract mechanism that implements the efficient allocation in a variety of cases, some of which combine risk aversion and asymmetric information. However, as Tirole (1999) discusses, we still do not know how to relate these abstract results to observed institutions and contract forms. The literature on employment typically explores the implications of regulation for a simple model that has one or at most two transaction costs. The literature on employment protection has for the most part followed the lead of Bentolila and Bertola (1990) and Lazear (1990) in supposing that an increase in labor protection is parsimoniously modeled as an increase in turnover costs.

Lazear (1990) observes that if complete contracts are possible, mandated severance payments can always be undone via the labor contract. In that case, a law mandating severance payments would have no effect on employment but would lower starting wages. Lazear carries out a study of 22 countries over a 29-year period and concludes that increasing severance pay to 3 months' salary for workers with 10 years' experience leads to a 1-percent reduction of the employment-to-population ratio. This is a reduced-form analysis that does not take into account the complex inter-temporal optimization problem faced by firms. This is the goal of Bentolila and Bertola (1990). They find that firing costs create complex inter-temporal incentives that depend upon the state of the business cycle. Specifically, firing costs reduce labor demand in good times but *increase* demand in bad times. Lower starting wages translate into lower firing costs, and hence firms have greater incentives to hire workers during downturns.

The literature has mostly followed the lead of this work and modeled employment protection as a turnover cost. The theoretical contributions have begun with one of the transaction costs (risk, asymmetric information, or holdup) and then explored the implications of employment protection modeled as a turnover cost. This allows one to explore the implications of treating a relationship with a particular policy choice. If we suppose that in different relationships one of the three transaction costs is more important, then this approach generates testable predictions of the effect of a law change for different relationships, which hopefully can be measured and hence form right-hand-side X variables.

I complete the section with a brief discussion of unions. From the perspective of transaction costs, one can view unions as an alternative to employment law. This provides a system for the implementation, enforcement, and arbitration of employment disputes between the firm and unionized employees.

Risk

Since the work of Azariadis (1975) and Baily (1974), the assumption of risk-averse workers has played an important role in the development of labor policy. I showed above that if the main role of the labor contract is to insure workers, then such a contract can be enforced with the use of expectation damages. Moreover, the contract will have the feature that if a relationship is no longer efficient, then the firm has the obligation to “sell” the worker’s contract to another firm. In the absence of bankruptcy constraints and asymmetric information, such an institution would implement the first best.⁵⁰

In practice, we observe contracts with features similar to this in the area of sports, but rarely elsewhere. In the case of athletes, the quality of the player and hence the value of a trade is information that is easily available to the teams in a league. Such conditions are not likely to be satisfied in general, however. There is an active literature in macro-economics that explores the role of turnover taxes and mandated severance pay when complete contracts are not possible. A seminal contribution in this literature is Hopenhayn and Rogerson (1993). They assume that workers have access to complete financial markets and hence can diversify firm risk. Under these assumptions they show that a turnover tax (or severance pay) is equal to one year of wages leads to a 2.5% reduction in employment.

Their model assumes away market incompleteness. Hopenhayn and Nicolini (1997) consider the case in which the firm provides the insurance services for the worker, but the worker is responsible for finding a new job. The point is that the matching process is both costly, and is an important element in labor market performance. They show that there should be a mandated unemployment insurance that is financed out of a re-employment tax. Moreover, the level of insurance (or replacement rate, that is, the fraction of one’s wages that are paid when unemployed) should fall with time. They show that this rule can result in a significant increase in market performance. The result also illustrates one role for government intervention that arises when there is a combination of risk aversion and moral hazard (worker’s search effort is not observed).

Notice that in the presence of fixed unemployment insurance payments, mandated severance pay provides an approximation to such a rule because it provides a high income to the worker early in her unemployment that is lost once the severance pay is spent. In the event of an employment dispute, even if the worker wins the case, in most jurisdictions there is a mandatory rule that parties should mitigate their losses from

⁵⁰ See Dye (1985), who uses this point to build a theory of contract length.

contract breach. In the case of an employee, this means that the employee should make a reasonable effort to find alternative employment. Any damages due to the worker would be based upon lost income given the new job.

Acemoglu and Shimer (1999) introduce a careful model of the matching process that generalizes many of the previous matching process, and then derive the optimal unemployment insurance. If the agent is risk-neutral, then there should be no unemployment insurance. This is equivalent to saying that employment at will is efficient when workers are risk-neutral. However, when workers are risk-averse, the provision of unemployment insurance increases wages, employment, and the capital-labor ratio. Pissarides (2001) explicitly discusses the role of employment protection. He also shows that with search frictions it is optimal to have unemployment insurance, and observes that employment legislation is an (imperfect) substitute for employment protection.

Finally, the recent paper Blanchard and Tirole (2008) builds upon these themes to explore the implementation of an efficient severance pay—unemployment insurance system in the face of a variety of market imperfections. There are cases in which there are limits on insurance and layoff taxes, *ex post* wage bargaining, and *ex ante* heterogeneity of firms or workers. The key insight is that not only do these various cases affect the design of insurance system, but that a third party such as the government is needed in order to implement the second-best optimum. In particular, if the state merely provided a set of courts that enforce private agreements, this would not achieve the first best.

This body of work is carried out using relatively conventional assumptions regarding the operation of the labor market. Together they suggest that arguments by legal scholars—such as Epstein (1984) or Morris (1995–1996)—that the efficiency of free markets implies that there is no role for government intervention into labor markets are not correct as rhetorical statements. However, there are many issues that this literature does not address.

First, these results depend upon workers having stable risk-averse preferences. There is a large body of work that finds that the fine-grained predictions from a model with risk-averse preferences are not consistent with the data. See Rosen (1985) and Hart and Holmström (1987) for important early evaluations of the literature. More recently, Gibbons (1997) has argued that the standard agency model does not adequately explain many features of observed contracts.

Some recent exciting research on the preferences of individuals may provide a way forward. Andreoni and Sprenger (2010a) show that previous research measuring risk and time preferences do not adequately control for the risk inherent in future rewards. In a follow-up paper, Andreoni and Sprenger (2010b) explicitly measure risk and time preferences. They find that individuals cognitively view choices as risky or not, but among risky choices they have relatively flat attitudes toward risk. This is consistent with the fact that individuals want to avoid risk but that there is no stable empirical relationship between attitudes toward risk and the form of the optimal contract. This research is very

new, and this latter point is yet a conjecture. If these results hold, they may allow for much better models of contract and optimal unemployment insurance.

Asymmetric information

Asymmetric information is ubiquitous in the employment relationship, which leads naturally to the question of how employment law and other labor market-institutions should be designed to handle this problem. Section 3.2.2 provided some examples of situations where asymmetric information can help explain both contract form and the allocation of authority within a relationship. The difficulty is how best to mediate the information problems that arise both within the relationship and between different potential matches.

The early literature focuses on the question of how contracts should be designed to ensure optimal matching when the firm has private information on worker productivity and the worker has private information on alternative opportunities. In a classic paper, Diamond and Maskin (1979) compare expectation damages (they call them compensatory damages) to privately negotiated liquidated damages in a buyer-search model. When parties match, they obtain a gain from trade that is split evenly. The issue is how much they should pay should they find a better match—this will affect the incentives to search and whether or not separations are efficient. They find whether or not one damage rule is better than another depends upon the technology of search; hence, in general, it illustrates that in a world with costly search it is difficult to obtain a clear general prediction on the optimal default rule.

In the employment context, workers are rarely asked to pay for damages should they find a better match. Hall and Lazear (1984) begins with this observation and compares three contract forms:

1. *Fixed wage w is set in advance.* Trade occurs only if both parties prefer trade to no trade at this price. This contract leads to inefficient quits and layoffs.
2. *Firm sets wage knowing worker's productivity.* This is essentially the monopsony solution: The firm has an incentive to set wages above the marginal product of labor, so there are inefficient separations whenever the worker's outside option is between the wage and the worker's marginal product.
3. *Workers set wages (monopoly union model).* In this case, the worker sets his wage above his outside option, resulting in the firm inefficiently not employing the worker when marginal product of the worker is greater than his outside option, but below the wage demanded.

If information is symmetric, then in case 1 we should observe renegotiation and efficient trade. However, as Myerson and Satterthwaite (1983) have shown in a general mechanism-design framework, when there is two-sided asymmetric information, efficient trade is impossible. Hence, as Hall and Lazear (1984) observe, there is no simple

contract that implements efficient trade *ex post*. Even in the absence of risk aversion or relation-specific investments, there are limits to efficient trade that no legal rule or contract can overcome.

There are, however, situations under which efficient trade is possible. If a worker's outside option is known, then giving the firm all the bargaining power results in efficient separations, just as it resulted in efficient task assignment in Section 3.2. This is a reasonable assumption when the labor market is thick, as would be the case for, say, casual day labor. In that case, at-will employment is an efficient rule. If variations in outside options are due to variations in worker's ability, then efficiency can still be achieved with the use of piece-rate contracts, as Kanemoto and MacLeod (1992) show. Such a result is consistent with the good-faith exception to employment at will that requires firms to follow through upon promised performance pay.

If it is possible to measure firm productivity, then the efficient rule is to give the worker all the bargaining power. In that case, the worker would offer a wage contract that would make the firm indifferent between acceptance and rejection, with the result that trade would occur if and only if it is efficient. Gibbons (1987), in a paper that complements Kanemoto and MacLeod (1992), shows that if the firm has bargaining power and the workers have information regarding the difficulty of the job, then the resulting contract is inefficient. If the workers were given all the bargaining power in this case, then the first best would be restored.

Though this point follows naturally from the question of how to implement efficient exchange under asymmetric information, it is oddly missing from the literature on unions. Freeman and Medoff (1984) made the point that unions can enhance efficiency via the "voice" mechanism, which can be interpreted as solving the problem of asymmetric information. Beginning with McDonald and Solow (1981), there is a literature that wonders why unions cannot bargain over both wages and employment to achieve an efficient outcome.

One reason is that, in practice, we are typically in a situation with two-sided asymmetric information. In that case, one cannot in general achieve the first best. However, if there are choices whose marginal costs vary with the hidden information, then optimal contracts should incorporate this information. This observation has led to several contributions that explain observed contracts as a solution to this problem. The early work of Grossman and Hart (1981) shows that the employment-wage policies of a firm are designed to reveal underlying productivity, which can result in observed wages that are different from the true marginal product of labor. Moore (1985) has refined this analysis, to show that when there is two-sided asymmetric information with risk-averse workers, the extent to which the optimal employment contract exhibits over- or under-employment depends upon the preferences of the workers, as well as the nature of uncertainty.

Aghion and Hermalin (1990) introduce a contracting model with asymmetric information and signaling. They show that laws mandating employer-provided benefits can enhance efficiency. There are also papers illustrating that several features of union contracts can be viewed as solving an information problem. Kuhn and Robert (1989) show that seniority rules are a form of efficient price discrimination against the firm. Laing (1994) provides a more general analysis of employment contracts with asymmetrically informed agents, finding that a seniority layoff rule may improve efficiency. Levine (1991) suggests that requiring just cause for dismissal risks attracting low-quality workers. Hence, it is argued that mandated just cause rules may enhance efficiency. Kuhn (1992) observes that requiring mandatory notice of plant closings enhances labor-market performance by ensuring that the firms inform workers in a timely fashion and allow them to make more efficient separation decisions.

There is a recent literature that takes a reduced-form approach to employment protection based on the idea that employment protection acts as a turnover tax that interacts with asymmetric information. Kugler (2004b) observes that in the presence of turnover costs, firms favor more skilled employees, and hence try to fill vacancies from currently employed workers. Hence, she finds that increases in employment protection reduces the flow from unemployment. Pries and Rogerson (2005) introduce more structure to the process regarding worker quality. They suppose that the formation of a match is both an inspection and experience good. The former requires firms and workers to engage in explicit search to form matches, while the latter implies that the signal of match quality becomes more precise with tenure. They explore several labor-market policies, including unemployment insurance, a minimum wage, and dismissal costs. They find that dismissal costs lead to higher unemployment, lower turnover, and higher-quality matches. In contrast, Burguet and Caminal (2008) show that if there is contract renegotiation and uncertainty regarding match, then turnover costs can enhance market performance. Guerrieri (2008) introduces a dynamic general equilibrium model with asymmetric information regarding match quality. She shows that in a dynamic economy the first best cannot be achieved without government intervention in the form of a lump sum tax upon all workers. Finally, Matouschek et al. (2009) formally consider the implications of contract renegotiation when there is asymmetric information regarding outside options.

The extent to which asymmetric information might “explain” observed contracts, and justify the existence of unions or additional taxation, depends upon the magnitude of the asymmetric information. Using evidence from layoffs, Gibbons and Katz (1991) find that workers who lose jobs from a plant closing have higher subsequent wages than those who are laid off. This is consistent with later work by Gibbons et al. (2005), where high-skilled workers earn greater returns to their skill. This evidence supports the hypothesis that labor market wages are a first-best approximation for worker ability. If it is true, then

this would suggest that *ex post* efficiency is best achieved with as little market intervention as possible.

This policy choice is no longer ideal if the labor contract must also ensure *ex ante* efficient investment. We turn to this issue next.

Holdup

When worker's productivity is common knowledge, but the cost of labor supply is private to the worker, then it is optimal to allocate all *ex post* bargaining power to the worker. However, this is not in general *ex ante* efficient. When unions and firms are in a long-term relationship, the firm will make investment decisions a function of their expected return from these investments. Grout (1984) has shown that if the union (or worker) has *ex post* bargaining power during contract renegotiation, then there is underinvestment.

Becker (1962), Mincer (1962) and Williamson et al. (1975) have emphasized that the employment relationship typically entails relationship-specific investments. This raises two issues. First, can the employment relationship be governed in such a way that one has efficient investment combined with efficient matching? Second, what are the implications for wages over time? Hashimoto (1981) introduces a model of incomplete wage contracting with relationship-specific investment into worker skill. He shows that if parties cannot condition the wage contract upon the worker's or the firm's alternative opportunities, then there will be inefficient quits (as in Hall and Lazear (1984) discussed above). The result is a wage contract at which the worker and the firm share the rents from firm-specific investments, as hypothesized by Becker (1962) and Mincer (1962).

This result assumes that the worker and the firm can commit to a fixed wage contract. If that is not possible, then the current wage is always set by *ex post* renegotiation, which can be expected to lead to an *ex post* efficient allocation when information is symmetric. Grout (1984) has shown that this leads to the worker's capturing a positive fraction of the rent created by the firm's investments, which in turn leads to lower investment by the firm and slower employment growth. Grout's goal is to model the implications of a UK law making it impossible to enforce wage agreements between a union and a firm. Even when such commitment is possible, Hart and Moore (1988) show that the fact that the worker and the firm can voluntarily renegotiate a contract in the face of new information implies that in general the first best cannot be achieved with a binding agreement.

However, Carmichael and MacLeod (2003) show that if parties can agree upon a *fair division rule* that divides the gains to trade in proportion to the investments made by each party, then the first best can be achieved even in the absence of a binding complete contract. Such a rule requires one party to penalize the other should the agreement be perceived as unfair. In the context of union-firm bargaining, there is some recent evidence that unions do retaliate when it is perceived that the wage bargain is unfair. Krueger and Mas (2004) show that a dispute between Firestone and their union led to lower-quality tires. Mas (2006) finds that the resale value of Caterpillar products fell for

equipment built during a labor dispute, suggesting again that labor unrest resulted in lower product quality. Finally, Mas (2008) finds that when police unions in New Jersey got adverse rulings in arbitration that led to lower wages, the police reduced their effort as measured by arrest rates.

These results fit in with an extensive literature, beginning with Akerlof (1982), that the extent to which a worker believes that treatment is fair affects productivity (see in particular the work by Bewley (1999) and Fehr and Schmidt (1999)). The holdup model provides an elegant explanation of why fairness is so important. The economic model predicts that investment into a relationship is a function of the return from such an investment. Note that after an investment has been made, however, it is a sunk cost, and hence rational choice theory would predict that any agreement made *after* investments have been made should be independent of these investments. Consequently, the only way compensation can be linked to investments *ex post* is for parties to follow a social norm that links them—in other words, parties should believe and enforce a norm of fairness that results in parties who invest more receiving more compensation. Carmichael and MacLeod (2003) provide a general proof of the existence of such norms. See also the work of Hart and Moore (2004), who argue that contracts can act as efficiency-enhancing focal points.

Such models have a potential to provide an efficiency-based explanation of why unions with bargaining power can enhance firm productivity. The next section discusses some of the empirical evidence in this regard. This perspective may also provide some insights into the decline in unionization that has occurred in the private sector in the United States (see Farber and Western (2001)). The issue is whether or not there exist alternatives to unions that enhance the productivity of the employment relationship. First, there is the possibility that employment law is a substitute for union protection. At the moment, we simply do not have any studies that explore this idea. Acemoglu and Pischke (1998) show that firm-sponsored investment into training can be enhanced by the fact that firms have superior knowledge regarding worker productivity. In Acemoglu and Pischke (1999), they also argue that increased employment protection enhances firm-sponsored training. They suggest that firm-sponsored training is higher in Europe than in the United States due to higher employment protection in Europe. We do know if the increase in employment protection in the US has led to more training. MacLeod and Nakavachara (2007) introduce a model with investment and asymmetric information, and show that increased employment regulation in the US would lead to a more productive relationship for highly skilled workers.

For lower-skilled workers, Autor (2003) has documented the fact that there has been an increase in temporary help agencies in the United States. He finds that this is in part explained by the increase in employment protection. However, Autor (2001) also documents that as an organizational form, temporary help agencies play a significant role in screening and training of workers.

3.4. Summary

In this section, we have reviewed the literature on the theory of the employment relationship from the perspective of transaction-costs economics and contract theory. From this research, we learn that an optimal employment contract is shaped by many factors in addition to the demand and supply of factors of production. The need to provide insurance to workers underlies many of the contributions, in part because this model is quite elegant and can in many situations deliver clear predictions. One prediction that it does not deliver is a theory of why employment relations can entail conflict, and why the allocation of bargaining power has important efficiency consequences.

Models of asymmetric information naturally deliver a theory of conflict, and can explain why parties for whom trade is efficient may fail to reach an agreement (see Crawford (1982)). The vast majority of work on employment focuses on the case in which the asymmetric information concerns the outside options for the worker and firm. Many of the employment disputes discussed in the previous section deal with disagreements concerning what happened within the relationship. The model in Simon (1951) is a useful starting point for thinking about these issues, but currently there is little work on the role of the courts in finding facts in employment cases.⁵¹ The main message here, consistent with the work of Milgrom (1988), is that the efficient contract for task assignments is to give the informed party decision rights. Chakravarty and MacLeod (2009) discuss how the law can achieve this goal in the context of construction contracts. The way the law achieves the formal allocation of authority within an employment relationship has not been explored in detail, however. Williamson (1991) makes the point that for the most part courts do not intervene in the day-to-day management of employees. The question of how various employment law doctrines affect this authority relationship is still an open question.

The provision of incentives to make relationship-specific investments gives the third motivation for entering into a binding contract. What is interesting is that the contractual instruments here—specifically, the allocation of authority and bargaining power—also play a central role in achieving efficient investment. There is a need for work that helps us understand which, if any, of these theories provide the most useful way to think about the employment contract. The theories by themselves are not typically framed in terms of making causal inferences. Rather, they make predictions regarding how variations in match characteristics (X variables) are related to observed features of the relationship (wages, employment, bonus or severance pay).

There is some recent work by Cahuc et al. (2006) and Postel-Vinay and Robin (2002) that uses the holdup model of wage determination developed in MacLeod and Malcomson (1993) to empirically estimate models of wage and employment

⁵¹ See MacLeod and Nakavachara (2007) and Stahler (2008) for a start.

determination. Cahuc et al. (2006) find that only highly skilled workers have significant bargaining power, while low-skilled workers have none. Postel-Vinay and Robin (2002) find in a panel with French data that personal characteristics tend to be more important for highly skilled workers. These are not causal exercises, but they do suggest that employment law is more likely to be important for highly skilled workers.

4. THE EVIDENCE

The theory of transaction costs provides an economic rationale for intervention into labor markets. Each of the models we have discussed capture some features of the employment relationship that seem empirically plausible. The next step is to see whether changes in employment law do improve matters. We discuss two sets of empirical results that illustrate a range of approaches. First, we review the literature on the employment contract suggesting that there are likely to be mechanisms by which employment law and unions affect the efficiency of labor markets. Second, we review the literature that asks to what extent unions enhance productivity.

4.1. Employment law

In terms of Rubin's model, the unit of analysis for employment laws are typically governments. The most common outcome variables are employment per unit of population, wages, the unemployment rate, and GDP growth. The issue then is how does a change in employment law change these outcome variables? This is a very difficult question because many events occur along with changes in the law that make it difficult to identify a causal effect. Surprisingly, the results tend to be relatively consistent. The majority of studies find either no effect or a negative effect of increasing employment protection. The results of this work are summarized in Table 2. We do not discuss all of these papers, but note that they can be divided into three broad classes: cross-sectional country studies, cross-sectional country studies with time, and the studies from the United States and India that use within-country state variation.

A good starting point are the cross-country studies. Botero et al. (2004) gather data from a several sources to construct measures of various types of employment regulation, including hours restrictions and dismissal procedures. They find that stricter employment protection is associated with countries whose law originated in the civil law tradition. For these countries, it is found that labor force participation is lower and unemployment is higher. Of course, these statements are not causal, but rather say that there is a co-variation between legal origins and employment performance. Djankov et al. (2003) show that civil codes make it more costly to use the courts for contract enforcement, but even this evidence is not necessarily causal. For example, if judges in civil law countries are more corrupt, then the more bureaucratic rules may be a response to this corruption. These basic results have recently been replicated by Djankov and Ramalho (2009), and hence there appears to be a relationship. The question is, why?

Table 2 Employment law summary.

Author, year and journal	Title	Sample	EP measure	Finding on employment
Addison and Grosso (2000)	The effect of dismissals protection on employment: more on a vexed theme	As Lazear (1990)	As Lazear (1990).	Employment protection is statistically insignificant.
Aghion (2008)	The unequal effects of liberalization: evidence from dismantling the License Raj in India	Indian states	Following, Besley and Burgess (2004), indicator of whether state amendment to Industrial Disputes Act of 1947 is neutral, pro-worker, or pro-employer.	Pro-worker states experience less employment growth relative to pro-employer states following delicensing.
Ahsan and Pages (2009)	Are all labor regulations equal? Evidence from Indian manufacturing	Indian states	Employment Protection Legislation according to Besley and Burgess (2004) methodology.	6.5% reduction in employment.
Amin (2009)	Labor regulation and employment in India's retail stores	Retail stores in India	Regulation index from Besley and Burgess (2004) and Regulation index based on stores' perceptions about severity of labor laws.	Flexible labor regulations increase employment by 22% for the average stores.

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Autor et al. (2006)	The costs of wrongful-discharge laws	United States	Implied-contract, public-policy, and good-faith exceptions.	Implied-contract: 0.8%–1.7% reduction in state employment rates. No effect for public-policy and good-faith.
Autor et al. (2007)	Does employment protection reduce productivity? Evidence from US states	United States, establishment-level data from the Census Bureau	Adoption of wrongful-discharge protection by state courts in the US from 1970 to 1999.	Wrongful-discharge protection reduces employment flows and firm entry rates. Moreover, plants engage in capital deepening and experience a decline in total factor productivity, indicative of altered production techniques. Evidence of strong contemporaneous growth in employment, however, leads us to view the findings as suggestive but tentative.
de Barros and Corseuil (2004)	The impact of regulations on Brazilian labor market performance	Brazilian establishment-level data; 1985–1998	New constitution of 1988, which caused many changes in labor codes, including increasing compensation for dismissals without just cause fourfold.	Effect of 1988 constitutional change on lagged employment is statistically insignificant.

(continued on next page)

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Besley and Burgess (2004)	Can labor regulation hinder economic performance? Evidence from India	Indian states	Indicator of whether amendment is neutral, pro-worker, or pro-employer.	States that amended in a pro-worker direction experienced lowered employment. 7.2% decrease in registered manufacturing employment; 28.5% decrease in daily employment in registered manufacturing.
Bird and Knopf (2009)	Do wrongful- discharge laws impair firm performance?	US data from approximately 18,000 commercial banks	Wrongful-discharge protections	The authors estimate the effects of wrongful-discharge protections adopted by US state courts during 1977-99. After controlling for local state economic conditions, the authors find evidence of a relationship between the adoption of the implied contract exception and the increase in labor expenses. In addition, adoption of the implied contract exception is found to have a significant and negative effect on overall profitability. The study corroborates previous findings that wrongful-discharge laws place increased costs on employers.

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Blanchard and Wolfers (2000)	The role of shocks and institutions in the rise of European unemployment: the aggregate evidence	20 OECD countries; 1960-1999; 8 five-year averages of data	Static and time-varying measures taken from Nickell (1997); time-varying measure taken from Lazear (1990) and updated.	Looking at unemployment, they find that shock-EP interaction terms point to amplification of the effects of adverse shocks. Essentially the same is true for the remaining institutional variables with two exceptions. The exceptions are coordination of collective bargaining and active labor market policies, which ameliorate the effects of adverse shocks. In general, much weaker interaction effects and poorer fit when static EP (and UI) measures are replaced by their time-varying counterparts.
Botero et al. (2004)	The regulation of labor	85 countries	Employment laws index composed measuring alternative employment contracts, cost of increasing hours worked, cost of firing workers, and dismissal procedures.	No significant effect on employment in the unofficial economy, negative effect on male LFP, positive effect on UE rate.

(continued on next page)

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Caballero et al. (2004)	Effective labor regulation and microeconomic flexibility	New sectoral panel for 60 countries	Job security regulation.	The job security regulation clearly hampers the creative-destruction process, especially in countries where regulations are likely to be enforced. Moving from the 20th to the 80th percentile in job security, in countries with strong rule of law, cuts the annual speed of adjustment to shocks by a third while shaving off about one percent from annual productivity growth. The same movement has negligible effects in countries with weak rule of law.
Di Tella and MacCulloch (2005)	The consequences of labor market flexibility: panel evidence based on survey data	21 OECD countries: 1984-1990	World Competitiveness Report data; indicator of flexibility (see text); Time-varying measure with five data points.	Statistically significant positive association between flexibility indicator and overall employment population ratio across all specifications. By demographic group this effect is much stronger for females than for males. Parallel results are obtained for the participation rate. Some evidence that flexibility increases average hours worked. The association between flexibility and the unemployment rate is negative throughout but not always statistically significant. The results for long-term unemployment are less precisely estimated.

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Djankov and Ramalho (2009)	Employment laws in developing countries	The Doing Business database, the World Development Indicators, informality data from the Global Competitiveness Report	Rigidity of employment index published by Doing Business.	Developing countries with rigid employment laws tend to have larger informal sectors and higher unemployment, especially among young workers. A number of countries, especially in Eastern Europe and West Africa, have recently undergone significant reforms to make employment laws more flexible. Conversely, several countries in Latin America have made employment laws more rigid. These reforms are larger in magnitude than any reforms in developed countries and their study can produce new insights on the benefits of labor regulation.
Downes (2004)	Labor market regulation and employment in the Caribbean	Barbados, Trinidad, and Jamaica aggregate data	Categorical measure of severance payment regime.	Statistically insignificant effect.
Elmeskov (1998)	Key lessons from labor market reforms: evidence from OECD countries' experience	19 OECD countries; 1983-1995	OECD (1994, Table 6.7, panel B, col. 2) ranking, but modified to take account of changes since late 1980s; two observations, time-varying indicator.	Employment protection raises structural unemployment but interaction effects are important. Employment protection not statistically significant in either highly centralized/coordinated or decentralized bargaining regimes.

(continued on next page)

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Eslava (2004)	The effects of structural reforms on productivity and profitability enhancing reallocation: evidence from Colombia	Plant-level longitudinal dataset for Colombia for the period 1982–1998	Market reforms in 1990s	Market reforms are associated with rising overall productivity that is largely driven by reallocation away from low- and towards high-productivity businesses. In addition, the allocation of activity across businesses is less driven by demand factors after reforms. The increase in aggregate productivity post-reform is entirely accounted for by the improved allocation of activity.
Feldmann (2008)	Business regulation and labor market performance around the world	74 countries; 2000–2003	Labor market regulations ratings from Economic Freedom of the World index.	Labor market regulations increase unemployment and reduce employment.
Feldmann (2009)	The unemployment effects of labor regulation around the world	73 countries; 2000–2003	Labor market regulations from the EFW index as well as the component indicators.	Stricter regulation generally appears to increase unemployment.

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Freeman (2002)	Institutional differences and economic performance among OECD countries	23 + countries; 1970-1990	Fraser Institute index of economic freedom (see text); time-varying measure with 6 data points.	Countries with a high degree of economic freedom have high employment-population rates and low unemployment—at least in terms of levels. Only for unemployment do the results survive the inclusion of country fixed effects.
Garibaldi (1999)	Deconstructing job creation	21 OECD countries; 1980-1998	OECD (1994, Table 6.5 panel B, col. 5) ranking; Moment-in-time measure.	There is a strong negative association between EP measure and employment growth in cross section (for 24 out of 27 cases), but in panel regressions the association is less precisely estimated and is statistically significant in one of five specifications only.
Grubb and Wells (1993)	Employment regulation and patterns of work in EC countries	11 EU countries; 1989	Authors' own indicators of restrictions on overall employee work (ORDW), dismissal of regular workers (RDSM), fixed-term contracts (RFTC), and temporary work agencies (RTWA).	ORDW reduces employment, increases self-employment, and reduces part-time work. RDSM (RFTC) increases (decreases) temporary work. RTWA but not RDSM reduces temporary agency work.

(continued on next page)

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Heckman and Pagés (2000)	The cost of job security regulation: evidence from Latin American labor markets	41 countries from LAC and OECD; 1980–1997 (max)	Authors' own cardinal measure based on severance pay, notice interval, and compensation for unfair dismissal (see text). Two period time-varying measure.	EP effect is negative and statistically significant for total employment for each estimating procedure. Similar results obtained for males and youth, but not females, the impact of EP on male employment being half the total employment effect and the youth effect is almost double the average effect. EP effects for females and self employment vary widely across estimating procedure. The results for unemployment depend on methodology and there is no statistically significant effect of EP on longer-term unemployment. Disaggregation by broad national grouping reveals that the employment effects of EP by demographic group are negative and mostly statistically significant. The exception is females in the Latin-American grouping. The effects of EP on unemployment are nearly always positive and stronger for the OECD grouping.

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Kahn (2010)	Employment protection reforms, employment and the incidence of temporary jobs in Europe: 1996–2001	Belgium, Finland, France, German, Italy, Netherlands, Portugal, Spain, United Kingdom, 1996–2001.	European Community Household Panel (ECHP)	A robust finding is that policies making it easier to create temporary jobs on average raise the likelihood that wage and salary workers will be in temporary jobs. This effect is felt primarily when the regional unemployment rate is relatively high. However, there is no evidence that such reforms raise employment. Thus, these reforms, while touted as a way of jump-starting individuals' careers in the job market, appear rather to encourage a substitution of temporary for permanent work.
Kaplan (2009)	Job creation and labor reform in Latin America	Firm level data from 14 Latin American countries	Index from the Economic Freedom of the World report as indicator of labor market rigidity.	More flexible labor regulations would lead to an average net increase of 2.08% in total employment. Firms with fewer than 20 employees would benefit the most, with average gains in net employment of 4.27%.
Kugler (1999)	The impact of firing costs on turnover and unemployment: evidence from the Colombian labor market reform	Repeated cross-sections from the Colombian National Household Survey (NHS)	Colombian Labor Market Reform of 1990, which substantially reduced dismissal costs.	The exit hazard rates into and out of unemployment increased after the reform by over 1% for formal workers (covered by the legislation) relative to informal workers (uncovered). The increase of the hazards implies a net decrease in unemployment of a third of a percentage point, which accounts for about one quarter of the fall in unemployment during the period of study.

(continued on next page)

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Kugler (2004a)	The Effect of job security regulations on labor market flexibility: evidence from the Colombian labor market reform	Colombian household data on employment; 1988, 1992, 1996	Colombia labor market reform of 1990 reduced severance payments, widened the definition of “just” dismissals, extended the use of temporary contracts, and sped up the process of mass dismissals.	Reform contributed to about 10% of the reduction in the unemployment rate between the pre and post reform periods.
Kugler (2004b)	How do firing costs affect worker flows in a world with adverse selection?	National Longitudinal Survey of Youth	Unjust-dismissal provisions in US states.	As firing costs increase, firms increasingly prefer hiring employed workers, who are less likely to be lemons. Unjust-dismissal provisions in US states reduce the re-employment probabilities of unemployed workers relative to employed workers. The relative effects of unjust-dismissal provisions on the unemployed are generally smaller for union workers and those who lost their previous jobs due to the end of a contract.
Lazear (1990)	Job security provisions and employment	20 countries; 1956-1984	Severance pay due blue-collar workers with 10 years of service; time-varying measure.	Employment protection raises unemployment and reduces employment participation and hours.

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
MacLeod and Nakavachara (2007)	Can wrongful discharge law enhance employment?	United States	Implied-contract exception.	Implied-contract: 3.6% reduction in employment of medium-investment group under school training criteria; 3.0% reduction in employment of medium-investment group under informal training criteria; other results are insignificant, but can reject null hypothesis that all adoption variables are insignificant for any training (11.7% level) and school training (0.0% level). Public-policy: No effect on employment of low and high investment groups across all criteria of investment. Good-faith: 7.1%-9.6% reduction in employment of low-investment group, 0%-3.4% increase in employment of medium-investment group, and 6.5%-14.5% increase in employment of high-investment group under any training, school training, and formal training criteria. No significant effects under informal-training criterion.
Miles (2000)	Common law exceptions to employment at will and US labor markets	US panel of state labor market aggregates from 1965 to 1994	The exceptions to employment at will that limit the circumstances of worker dismissal.	The exceptions to employment are seen to have no effect on aggregate employment nor unemployment. However, temporary employment is observed to increase by a statistically significant 15% following the adoption of an exception. The implied contract exception, in particular, appears to be responsible for this increase.

(continued on next page)

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Mondino and Montoya (2004)	The effects of labor market regulations on employment decisions by firms: empirical evidence for Argentina	Argentinian panel of manufacturing firms	Measure of regulation, composed of taxes and expected severance payment.	Negative effect of rigid labor regulation on employment.
Montenegro and Pagés (2004)	Who benefits from labor market regulations? Chile, 1960-1998	Chilean household survey data; 1960-1998	Job security measure derived in Pagés and Montenegro (2007) which captures the expected cost, at the time a worker is hired, of dismissing the worker in the future.	Job security is associated with lower employment rates for young workers, female and unskilled workers, and higher employment for older and skilled workers.
Nickell (1997)	Unemployment and labor market rigidities: Europe versus North America	20 OECD countries; 1983-1988 and 1989-1994	OECD (1996, Table 6.7, panel B, col. 5) ranking; also use of labor standards measure covering in addition to employment protection working time, minimum wages, and employee representation rights (OECD, 1994, Table 4.8, col. 6).	Employment protection reduces overall employment rate but not that of prime-age males. For unemployment, employment protection effect is negative but statistically insignificant. Employment protection reduces short-term unemployment and increases long-term unemployment. Coefficient estimate for worker labor standards variable is statistically insignificant in unemployment regression.

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Nicoletti et al. (2003)	Interactions between product and labor market regulations: do they affect unemployment?	20 OECD countries; 1982-1998	Two indicators of the stringency of the regulatory apparatus. The first is EP per se, and is based on the time varying OECD (1999, Table 2.5) measure. The second is a measure of the degree of product model regulation and is both static (based on Nicoletti et al., 1999) and time varying (based on the authors' evaluation of regulation and market conditions in 7 energy and service industries, 1970-1998).	In initial fixed effects specification, EP is associated with a statistically significant reduction in employment. When EP enters in interaction with the coordination of collective bargaining dummies, its effects are negative and statistically significant for both intermediate and high coordination. The same results obtain for the random effects and second stage regressions. In each case, the negative effect on employment is stronger in countries with an intermediate degree of coordination. The effect of the static product market regulation variable is statistically significant and negative. Finally, for the fixed effect panel regressions, EP is negative and statistically significant in the basic specification. In interactive form, however, the negative coefficient estimate for EP is only statistically significant for the intermediate coordination measure. In interaction with the coordination measure, the product market regulation variable is negative throughout, but is statistically significant for low and intermediate coordination.
OECD (1993)		19 OECD countries; 1979-1991	Severance pay and notice periods combined across blue- and white-collar workers; moment-in-time indicator.	Employment protection has positive effects on jobless duration, especially in southern Europe.

(continued on next page)

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
OECD (1999)	Economic protection and labor market performance	19 OECD countries; 1985–1990, 1992–1997	OECD (1999, Table 2.5) measures for late 1980s and late 1990s; single overall indicator and also separate indicators for regular employment, temporary employment and collective dismissals; in some specifications further disaggregations for regular and temporary employment.	Irrespective of the form of the indicator, EP coefficient estimate is statistically insignificant for overall unemployment. It is positive and statistically significant for prime-age male unemployment (overall indicator only). For all other demographic groups EP is statistically insignificant. Further, changes in EP do not affect changes in unemployment for other than prime-age females, where the effect is negative and statistically significant (strictness of EP with respect to regular employment). For employment, the coefficient estimates for EP are negative but statistically insignificant for overall, prime-age female, youth, and temporary employment. Otherwise they are positive and in the case of self-employment statistically significant (overall EP measure and its regular employment variant). Further, changes in EP have statistically insignificant effects for overall employment and for all demographic groups. For self-employment and the share of temporary employment, some statistically significant negative effects are observed.

Table 2 (continued)

Author, year and journal	Title	Sample	EP measure	Finding on employment
Pagés and Montenegro (2007)	Job security and the age-composition of employment: evidence from Chile	Chilean annual household surveys; 1960-1998	Index combining information on notice periods, compensation for dismissal, the likelihood that a firms' difficulties be considered as justified cause of dismissal, and severance pay due to that event. This index measures the expected cost, at the time the worker is hired, of dismissing a worker in the future.	Job security is associated with a substantial decline in the wage employment-to-population rate of young workers. No decline in young self-employment rates or in the wage employment rates of older workers. Adverse effect of job security on youth employment is driven by the link between severance payments and tenure. Job security does not have a significant impact on overall aggregate employment, participation or unemployment rates.
Saavedra and Torero (2004)	Labor market reforms and their impact over formal labor demand and job market turnover: the case of Peru	Peruvian firm and sector level data; 1986-1997.	Expected severance payment	Reduction in firing costs (as measured by expected severance payment) has positive effect on employment level.
Scarpetta (1996)	Assessing the role of labor market policies and institutional settings on unemployment: a cross-country study	17 OECD countries; 1983-1993	OECD strictness ranking for regulation of dismissal averaged over regular and fixed-term contracts (OECD, 1994, Table 6.7, panel B, col. 2).	Employment protection raises structural unemployment, with stronger effects for youth and long-term unemployment. Employment protection increases nonemployment rate.

Table 3 Employment law for a selected set of countries.

Country	Min. work age	Holidays	OT Prem.	Sev. Pay	Ret. Ben. Yrs.	Unemp. Wait	Matern. Leave
Switzerland	15	9	1.25	0.0	1	5	4.5
United States	16	0	1.5	0.0	20	3	1.35
Singapore	12	11	1.5	12.9	5	n.a.	3.0
Finland	16	11	2.0	0.0	0	7	4.2
Italy	15	11	1.1	0.0	19	7	1.35
New Zealand	16	11	1.0	0.0	0	70	1.5
Portugal	16	12	1.5	12.9	15	0	2.0
Uruguay	15	5	2.0	12.9	35	0	4.1
Malaysia	14	10	1.5	2.14	20	n.a.	4.05
Mexico	14	7	2.0	18.04	10	n.a.	3.0
Lebanon	13	13	1.5	12.9	20	n.a.	3.5
Russia	16	9	2.0	8.6	25	0	12.0
Lithuania	14	11	1.0	8.6	30	8	2.0
Jordan	16	12	1.25	12.9	10	7	2.0
Morocco	12	11	1.25	0.0	9	n.a.	3.0
Indonesia	12	12	1.5	25.8	20	n.a.	2.25
Zimbabwe	17	11	1.0	0.0	10	n.a.	3.0
Armenia	16	13	1.5	6.45	25	0	6.0
India	14	5	2.0	6.43	10	n.a.	4.0
Vietnam	18	5	1.5	12.9	20	n.a.	3.0
Madagascar	14	2	1.3	4.2	15	n.a.	3.0
Mozambique	18	9	1.5	25.8	10	n.a.	0.0

Employment Rules in Sample of OECD Countries. The 85 countries studied in Botero et al. (2004) were ranked by per capita income, and every fourth country was selected. The table presents measurements for the following employment rules. *Min. Work Age.* The minimum legal age for attaining full-time employment. *Holidays.* Number of legally mandated paid holidays per year. *OT Prem.* The premium for working overtime, as a multiple of normal-time wages. *Sev. Pay.* Legally mandated severance payment for terminated workers, in week's pay. *Ret. Ben. Yrs.* Years of work required before a worker is eligible for retirement benefits through the country's social security program. Obtained by multiplying the relevant index by 45 (the maximum observation in years) and subtracting that number from 45. *Unemp. Wait.* Waiting period in days before a worker becomes eligible for unemployment benefits. Obtained by multiplying the relevant index by 70 (the maximum observation in days), and subtracting that number from 70. "n.a." indicates that the country does not offer unemployment benefits. *Matern. Leave.* Number of months of legally mandated maternity leave.

A natural way to control for cross-country variations is to use variation over time in laws. Early studies along these lines include the influential work of the OECD (1993). They found that employment protection increased jobless duration. In subsequent work, the OECD (2003) finds that employment protection is often insignificant, but is associated with increased unemployment for prime-age males. These, like the cross-section studies, must construct measures of employment protection that are meaningful in different countries. The difficulty, as we see in Table 3, is that there are a large number of possible laws, each of which get implemented in an idiosyncratic fashion. One way to deal with this complexity is to have a more narrow study of a law change. A series of studies by Kugler (1999, 2004a, 2005) uses the 1990 market-based reforms in Columbia.

Kugler finds that these reforms generally lead to more flexible labor markets and lower unemployment.

Given that these changes occurred at a single point in time, this implies a true causal effect under the hypothesis that other secular changes would not have produced this effect. One way to satisfy this assumption is to narrow the analysis to a set of units that are more similar, but face changes at different times. Besley and Burgess (2004) pioneered this approach using variations in employment law across states in India. Given that all the units (states) are in the same country, this controls for legal origin. Using data from 1958-1992 on employment law legislation, Besley and Burgess (2004) find that pro-worker legislation lowers economic growth. Similar results have been replicated by Aghion (2008) and Ahsan and Pages (2009). However, given the large cultural diversity in India, one might question the extent to which the law is exogenous to other events in society.

Possibly, the most convincing studies on the effect of the law exploit the fact that in the United States employment law is under state jurisdiction. One can then estimate the effect that state level changes in exceptions to employment at will have on the labor market. These law changes for the 1983-95 period are illustrated in Fig. 1. As we can see, a large number of states adopted the implied-contract and public-policy exceptions during this time period. In this case, the identifying assumption is that US states are sufficiently similar that one can assume that the impact of the law in each state is similar. One then uses a generalization of model (1):

$$y_{ut} = \beta^\top X_{ut} + \beta_l^\top l_{st} + \phi_{ut},$$

where u denotes state. In most papers, the Current Population Survey (CPS) is used to measure employment and wages by state. Bertrand et al. (2004) show that one cannot assume that the error term ϕ_{ut} is iid, and one must allow for correlation over time. In practice this is achieved by adding state-specific time trends and computing standard errors with clustering of the errors at the state level (in effect allowing arbitrary covariance over time within states, but assuming independent errors across states).

Miles (2000) is an early study by a legal scholar who carefully ensured that employment law is correctly coded. This can be difficult in the United States because law is created by both the courts and the states. He finds very little effect on employment from any of the law variables, but finds that the implied contract exception leads to an increase in temporary employment, consistent with the later work of Autor (2003). Autor et al. (2004, 2006) further refine the law variables used by Miles (2000), finding that the implied-contract exception leads to a robust reduction in state employment, but that the public-policy and good-faith exceptions have no effect.

The theory discussed above predicts that the effect of the law depends upon the characteristics of the employment relationship. In particular, if hold-up is the transaction

cost most responsible for the creation of employment law, then we would expect the law to have more effect on matches with higher levels of relationship-specific investment. This idea motivates the work of MacLeod and Nakavachara (2007). They match the CPS job training supplement that measures the amount of skill in a job with an occupation code to divide occupations in the low, middle, and high skill. They then use the data created by Autor et al. (2004) to explore the effect of relating skill requirement with the law. They find that the negative effects of the implied contract and good faith exceptions tend to be concentrated in low-skill (high-turnover) jobs. In fact, the good faith exception has a *positive* effect on employment for high-skill workers.

Though this appears to be at odds with the previous literature, it should be noted that the earlier literature focuses on the average effects, which obscures the effect on different subgroups. One of the messages from the recent research on contract theory is that optimal contract form *should* be sensitive to the characteristics of the job, particularly for high-performance jobs. Consequently, we cannot obtain a complete understanding of how the law works without taking into account its impact on different types of relationships. Both the theory and the early work of Freeman and Medoff (1984) predict that if transaction costs are a significant source of inefficiency in employment relationships, then there is a role for unions to enhance performance.

4.2. The effect of unions

Table 4 provides a list of studies that explore the effect of unions upon firm productivity. The early studies by Brown and Medoff (1978) and Clark (1980b) explicitly recognize that unions may enhance performance by increasing investments in firm-specific human capital, improving worker morale and other organizational parameters. Clark (1980b) also observes that union contracts have many terms that address issues in the workplace other than compensation. Brown and Medoff (1978) estimate a Cobb-Douglas production function using data from the May CPS merged with the 1972 Census of Manufacturers (COM). They find that labor productivity is consistently higher at unionized firms. They are careful to point out that this result can be explained by several factors, and while consistent with the hypothesis that unions enhance productivity, they do not establish a causal link.

Clark (1980b) refines this approach by using a time series on US cement plants, which allows one to explore the effect of unions upon plants in the same industry, with precise control on capital equipment. Like Brown and Medoff (1978), Clark finds that unions enhance productivity. One of the interesting findings is that if one compares only new plants, then unionized firms have 3.6% higher productivity, but a lower capital/labor ratio. This is consistent with the holdup model of Grout (1984), which predicts that firms will invest less in physical capital in response to union bargaining power. Abowd (1989) uses stock market data to conclude that unions and firms maximize total wealth.

Table 4 Unions and productivity.

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Abowd (1989)	The effect of wage bargains on the stock market value of the firm	Contract settlements, 1976–1982	Change in value of common stock	Unexpected change in collectively bargained labor costs	The change in the value of common stock resulting from an unexpected change in collectively bargained labor costs is estimated. Using bargaining unit wage data and NYSE stock returns, it is shown that there is a dollar-for-dollar tradeoff between these variables. This result is consistent with stock valuations based on present value maximizing managerial decisions (Hotelling's lemma). The analysis provides support for the hypothesis that collective bargains maximize the sum of shareholders' and union members' wealth.
Allen (1984)	Unionized construction workers are more productive	US manufacturing industries, 1972	Value added	Fraction unionized	This paper presents evidence on the effect of unionism on productivity in construction. The linkages are distinct from those studied previously in industrial settings. Apprenticeship training and hiring halls probably raise union productivity, while jurisdictional disputes and restrictive work rules lower it. Using Brown and Medoff's methodology, union productivity, measured by value added per employee, is 45% to 52% higher than nonunion. The estimate declines to 17% to 22% when estimates of inter-area construction price differences are used to deflate value added. Occupational mix differences and, possibly, apprenticeship training account for 15% to 27% of this difference.

(continued on next page)

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Allen (1986)	The effect of unionism on productivity in privately and publicly owned hospitals and nursing-homes	Publicly and privately owned hospitals and nursing homes, 1976	Value added, square feet	Union dummy, union dummy *public ownership	This paper examines the effect of unions on productivity within a sample of publicly and privately owned hospitals and nursing homes to determine whether public ownership influences union behavior. The results show that the productivity of union contractors is much greater in private than in public projects. Within the sample of private projects, the estimates of the union-nonunion productivity difference are generally positive but very imprecise.
Allen (1986)	Unionization and productivity in office building and school construction	Commercial office buildings completed in 1973-74, elementary schools completed in 1972	Value added, square feet	Union dummy	This study examines the difference in productivity between union and nonunion contractors in the construction industry within two samples, one of 83 commercial office buildings completed in 1973-74 and the other of 68 elementary and secondary schools completed in 1972. An analysis that includes controls for differences in capital-labor ratios, observable labor quality, region, and building characteristics shows that union productivity in the office building projects was at least 30% higher than nonunion productivity, measured in terms of square feet of floor space completed per hour worked; and from zero to 20% higher in school projects, measured in physical units and value added, respectively.

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Becker and Olson (1990)	The effects of the NLRA on stockholder wealth in the 1930s	Firms at risk of being unionized in the 1930s	Cumulative average excess return	NLRA and related events	Whereas most recent research examining the effects of the NLRA on labor-management relations has focused on the impact of specific provisions of the law, this study provides an estimate of the impact of the passage of the NLRA on 75 firms that were at great risk of being unionized in the 1930s. Taking changes in shareholder wealth as a measure of the shift in the balance of power caused by the NLRA, the authors find that the passage of the NLRA caused a statistically and economically significant decline in shareholder wealth for the sampled firms. Specifically, by April 1937, stockholder wealth in the firms was 15.9% lower than would have been expected had the NLRA not been enacted.
Becker and Olson (1992)	Unions and firm profits	US manufacturing firms, 1977	Excess value, price-cost margin	Fraction unionized	This study examines the effect of unionization on manufacturing firm profits, extending earlier research by combining industry-level and firm-level measures of unionization. Using several profit measures, we find that quasi-rents from firm investments in intangible assets are a relatively greater source of union profit effects than product market concentration and that union profit effects occur largely in the first 10% of firm coverage, suggesting spillover effects on nonunion employees.

(continued on next page)

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Bemmels (1987)	How unions affect productivity in manufacturing plant	Manufacturing plants, 1982	Value added	Fraction unionized, fraction unionized squared, fraction unionized *participative management, fraction unionized *performance appraisal, fraction unionized *incentive pay	This study investigates several hypotheses as to how unions may affect productivity. Analysis of data from 46 manufacturing plants for 1982 indicates a negative union impact on productivity. The author finds evidence suggesting that unions reduce the effectiveness of some managerial practices undertaken to increase productivity, and that poor labor-management relations climate also reduce productivity. These two factors account for almost 50% of the negative union impact.

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Black and Lynch (2001)	How to compete: the impact of workplace practices and information technology on productivity	Manufacturing establishments, 1987-1993	Sales	Union dummy, union dummy ★ TQM (total quality management system), union dummy ★ profit sharing for non-managerial workers	Using data from a unique, nationally representative, sample of businesses, the authors examine the impact of workplace practices, information technology, and human capital investments on productivity. The authors estimate an augmented Cobb-Douglas production function with both cross section and panel data covering the period of 1987-1993, using both within and GMM estimators. It is found that it is not whether an employer adopts a particular work practice but rather how that work practice is actually implemented within the establishment that is associated with higher productivity. Unionized establishments that have adopted human resource practices that promote joint decision making coupled with incentive-based compensation have higher productivity than other similar nonunion plants, whereas unionized businesses that maintain more traditional labor management relations have lower productivity. Finally, plant productivity is higher in businesses with more-educated workers or greater computer usage by non-managerial employees.

(continued on next page)

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Bronars and Deere (1990)	Union representation elections and firm profitability	US union elections, 1962-1980	Petition date excess return; certification date excess return	Firm and industry characteristics	Union representation elections are associated with significant declines in firm profitability. In addition to the significant mean effect of union elections on the equity value of firms, there exists substantial variation in the magnitude of equity losses across individual election events. Cross-sectional variation in shareholder equity losses can be explained by the labor intensity of the firm, the size of the union wage premium and fraction of workers organized in the firm's industry, the presence or absence of right-to-work laws in the state where the election is held, the number of workers covered in the representation election, and the number of previous union representation elections in the firm. The empirical results indicate that equity losses are the greatest in industries where union wage gains are the highest and unionization rates are the largest, and in the most labor-intensive firms, independent of the size of the bargaining unit involved in the election. The latter result indicates the presence of union spillover effects.

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Bronars and Deere (1994)	Unionization and profitability— evidence of spillover effects	US union elections, 1962-1980	Excess return	Petition date	Ruback and Zimmerman's (1984) event study, which showed that formal union-organizing activity significantly lowers a firm's market value, provides compelling evidence that unionization reduces a firm's profitability. Union-organizing activity may also have sizable cross-firm or spillover effects on the performance of neighboring firms. Thus the total impact of union organizing on the wealth of firm owners may be substantially understated by focusing only on the firm in which the union-organizing activity occurs. This paper presents an alternative approach for estimating union spillover effects on profitability using a data set and methodology similar to Ruback and Zimmerman's. Analysis of the impact of petitions for NLRB representation elections at one firm on the share prices of other firms in the same narrowly defined industry suggests that the total negative effects of unionization on profits, after cross-firm or spillover effects are included, are nearly three times as large as the own-firm effects reported by Ruback and Zimmerman.

(continued on next page)

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Brown and Medoff (1978)	Trade unions in the production process	US manufacturing industries, 1972	Value added	Fraction unionized	In order to estimate the effects of unions on worker productivity, a Cobb-Douglas production function is modified so that unionization is included as a variable. The resulting functional form is similar to that used to isolate the effect of worker quality in previous studies. Using state by two-digit SIC observations for US manufacturing, unionization is found to have a substantial positive effect on output per worker. However, this result depends on two important assumptions which we cannot verify directly; attempts to relax these assumptions are not conclusive.

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Clark (1980a)	The impact of unionization on productivity: a case study	Six establishments in US cement industry, 1953–1976	Tons per man-hour	Union dummy	This study examines the effect of unionization on productivity through the use of time-series data on selected establishments in the US cement industry. The analysis combines statistical estimation of the union impact and interviews with union and management officials to forge a link between econometric estimation and the traditional institutional analysis of union policy and management practice. The econometric analysis deals primarily with the problem of controlling for interfirm differences in variables such as the quality of management and also for the possible union impact on labor quality. The case studies are designed to show the specific ways in which unionization affects productivity. The empirical results indicate that unionization leads to productivity gains, deriving in large part from a series of extensive changes in management personnel and procedures.

(continued on next page)

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Clark (1980b)	Unionization and productivity: micro- econometric evidence	US cement industry, 1973–1976	Tons per man-hour	Union dummy	It is widely agreed that unionization affects the rules and procedures governing the employment relation in organized establishments. The effect of these changes on establishment productivity, however, is unclear. This issue is examined using establishment level data from the US cement industry. A positive union effect on the order of 6%–8% is found in both cross-section and time series data. Although some caution is in order in interpreting the results, the evidence suggests that unionization can lead to productive changes in the operation of the enterprise.

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Clark (1984)	Unionization and firm performance: the impact on profits, growth, and productivity	Product-line businesses, 1970-1980	Sales, value added, return on capital, return on sales, rate of growth of deflated sales	Union dummy	<p>The history of collective bargaining in the US has been marked by dramatic episodes of confrontation that underscore the change unionization brings to the operation of an enterprise. In spite of the diffusion of many practices associated with collective bargaining recent research has revealed the continuing existence of important differences between union and nonunion establishments in policies governing compensation, exit and entry, dispute resolution and internal promotion.</p> <p>Unionization works through more than one mediating factor and the impact of the union on a given measure of firm performance depends on the particular context in which bargaining and production take place. This article uses microeconomic data on over 900 product-line businesses to gauge the impact of the union on economic performance. A clear implication of the theoretical analysis is the need to examine several indicators of firm behavior in order to draw inferences about the operation and consequences of collective bargaining.</p>

(continued on next page)

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Cooke (1994)	Employee participation programs, group-based incentives, and company performance: a union-nonunion comparison	Michigan manufacturing firms, 1989	Value added	Union dummy interacted with work teams and profit/gain sharing plan	This analysis examines whether union representation positively or negatively influences the effectiveness of employee participation programs and group-based incentives in improving firm performance. Examined at the firm level, a model of the independent and interaction effects of participation, profit and gain sharing, and union representation is estimated against data on 841 manufacturing firms in Michigan in 1989. The evidence indicates that employee participation programs contributed substantially more to performance in unionized firms than in nonunion firms, whereas profit and gain sharing programs contributed substantially more to performance in nonunion firms than in unionized firms.

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
DiNardo and Lee (2004)	Economic impacts of new unionization on private sector employers: 1984–2001	US union elections, 1984–1999	Total value of shipments	Vote share	Economic impacts of unionization on employers are difficult to estimate in the absence of large, representative data on establishments with union status information. Estimates are also confounded by selection bias, because unions could organize at highly profitable enterprises that are more likely to grow and pay higher wages. Using multiple establishment-level data sets that represent establishments that faced organizing drives in the United States during 1984–1999, this paper uses a regression discontinuity design to estimate the impact of unionization on business survival, employment, output, productivity, and wages. Essentially, outcomes for employers where unions barely won the election (e.g., by one vote) are compared with those where the unions barely lost. The analysis finds small impacts on all outcomes that we examine; estimates for wages are close to zero. The evidence suggests that—at least in recent decades—the legal mandate that requires the employer to bargain with a certified union has had little economic impact on employers, because unions have been somewhat unsuccessful at securing significant wage gains.

(continued on next page)

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Eberts and Stone (1987)	Teacher unions and the productivity of public schools	Educational programs, late 1970s	Standardized test scores	Union dummy	Do teacher unions affect the productivity of public schools? The authors examine this question using individual student data from the Sustaining Effects Survey sponsored by the US Department of Education. Holding resources constant and using achievement gains on standardized tests as the measure of output, they find that union districts are seven percent more productive for average students. For the minority of students who are significantly above or below average, however, nonunion districts are more productive by about the same margin, apparently because teacher unions reduce the use of specialized instructional techniques. This result is consistent with the view that unions tend to standardize the workplace. Across all students, the average union productivity advantage is three percent.
Ehrenberg et al. (1983)	Unions and productivity in the public sector: a study of municipal libraries	Municipal libraries, 1977	Number of information requests, number of borrowers, number of interlibrary loans, total circulation, book and periodical circulation	Union dummy	This paper develops and illustrates the use of two methodologies to analyze the effect of unions on productivity in the public sector. Although the methodologies are applicable to a wide variety of public sector functions, the focus of the paper is on municipal libraries because of the availability of relevant data. The empirical analysis, which uses 1977 cross-section data on 260 libraries, suggests that collective bargaining coverage has not significantly affected productivity in municipal libraries.

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Hirsch (1991)	Union coverage and profitability among United States firms	US manufacturing firms, 1972-1980	Tobin's q, rate of return on capital	Fraction unionized	This paper utilizes unique survey data on labor union coverage at the firm level to examine union effects on the profitability of 705 US companies during the 1970s. Market value and earnings are estimated to be about 10%-15% lower in an average unionized company than in a nonunion company, following extensive control for firm and industry characteristics. Deleterious union effects on firm profitability are sizable throughout the 1972-80 period, but vary considerably across industries. The relatively poor profit performance of unionized companies may help explain the recent decline in US union membership.
Hirsch and Connolly (1987)	Do unions capture monopoly profits?	US manufacturing firms, 1977	Tobin's q, rate of return on sales	Fraction unionized	This paper challenges the conclusion reached in recent studies that unions reduce profits exclusively in highly concentrated industries. From their review of previous studies and their analysis of 1977 data on 367 Fortune 500 firms, the authors conclude that there is no convincing evidence that concentration produces monopoly profits for unions to capture. Moreover, they find that the union wage effect is not greater in concentrated industries, as suggested by the hypothesis that unions capture concentration-related profits. Evidence is found suggesting that a firm's market share, its expenditures on research and development, and its protection from foreign competition provide more likely sources for union rents than does industry concentration.

(continued on next page)

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Lee and Mas (2009)	Long-run impacts of unions on firms: new evidence from financial markets, 1961-1999	US union elections, 1961-1999	Abnormal return	Election	The authors estimate the effect of new unionization on firms' equity value over the 1961-1999 period using a newly assembled sample of National Labor Relations Board (NLRB) representation elections matched to stock market data. Event-study estimates show an average union effect on the equity value of the firm equivalent to a cost of at least \$40,500 per unionized worker. At the same time, point estimates from a regression-discontinuity design—comparing the stock market impact of close union election wins to close losses—are considerably smaller and close to zero. The analysis indicates a negative relationship between the cumulative abnormal returns and the vote share in support of the union, which allows to reconcile these seemingly contradictory findings. Using the magnitudes from the analysis, the authors calibrate a structural “median voter” model of endogenous union determination in order to conduct counterfactual policy simulations of policies that would marginally increase the ease of unionization.

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Lovell et al. (1988)	The effect of unionization on labor productivity: some additional evidence	US Economy, 1948-1973	GDP	Fraction unionized	Following Brown and Medoff (1978), a number of studies have investigated the effect of unionization on labor productivity using a log-linear, Cobb-Douglas model of technology. To derive this model, a first-order Taylor-series approximation to the intrinsically nonlinear unionization variable is made; the resulting linear equation is estimated with generalized least-squares (GLS) techniques. The authors demonstrate that this approximation introduces a bias that necessarily results in an overstatement of the absolute value of the exact union productivity effect. The authors illustrate the magnitude of this bias by comparing GLS estimates of the linear Brown-Medoff model with GLS estimates of the exact, nonlinear relationship, using aggregate time-series data from the private domestic sector of the US economy.
Machin (1991)	The productivity effects of unionization and firm size in British engineering firms	52 British engineering firms over the period 1978-82	Value-added	Several indicators of union presence	Average union-nonunion effects on labor productivity estimated using this measure, or using a dummy variable indicating the presence of closed-shop arrangements, are found to be statistically insignificant. However, there is some variation around this average, and the union impact on value added per employee is found to depend significantly on firm size, the estimated effects being more negative in larger firms.

(continued on next page)

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Mitchell and Stone (1992)	Union effects on productivity: evidence from western US sawmills	Western US sawmills, 1986	Lumber (million board feet)	Union dummy	The authors conjecture that previous studies have tended to overestimate the productivity of union firms relative to nonunion firms due to inadequate controls for output quality and input usage—important omissions if the higher cost of unionized labor leads to less labor-intensive products and techniques. To avoid those problems, this study examines a fairly standardized commodity, lumber, and controls for detailed product attributes and inputs. An analysis of data from a survey administered by the authors shows that unionized sawmills were between 12% and 21% less productive than nonunionized mills in fiscal year 1986. As predicted, when product quality and raw material usage are not included in the analysis, the estimate of union productivity is biased upward.
Ruback and Zim- merman (1984)	Unionization and profitability: evidence from the capital market	US union elections, 1962–1980	Abnormal return	Petition and certification dates	This paper examines the effect of unionization on the profitability of firms. Abnormal monthly common stock returns for a sample of 253 NYSE-listed firms are estimated for the month in which the union petitions for an election and for the month in which the National Labor Relations Board certifies the election outcome. The results suggest that unionization, on average, is associated with a reduction in equity value. When unions win an election, the average loss associated with the unionization drive is 3.8% of equity value. When unions lose an election, there is an average net reduction of 1.3% in the equity value of the firm.

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Salinger (1984)	Tobin's q, unionization, and the concentration- profits relationship	Manufacturing firms, 1979	Tobin's q, profit rate	Fraction unionized	This article uses Tobin's q, the ratio of the market value of a firm to the replacement value of its physical assets, to measure monopoly power and to examine the relationship between market structure and profitability. Tobin's q is a better measure of monopoly profits than indices of single-period profitability because it measures long-run monopoly power. In additions, it is subject to less measurement error and it contains an adjustment for risk. The relationship between q and long-run monopoly power is established. Provided that all inputs are supplied competitively, q should be highly sensitive to even small amounts of monopoly power. Since the level of q is generally not high in the American economy, the result suggests either that monopoly power is absent or that unions manage to capture monopoly rents. Empirical tests of the relationship between Tobin's q and measures of market structure and unionization provide evidence that unions do capture most monopoly rents.

(continued on next page)

Table 4 (continued)

Author, year and journal	Title	Sample	Main dependent variables	Main independent variables	Finding on productivity
Voos and Mishel (1986)	The union impact on profits: evidence from industry price-cost margin data	US manufacturing firms, 1968-1972	Price-cost margin	Fraction unionized	This paper uses industry price-cost margin data to estimate the extent to which unions reduce profits. Estimates allowing for the endogeneity of union status are contrasted with estimates that assume union status is exogenous and not determined in part by either profitability or industry structure. Endogeneity is found to be an important consideration in estimating the union impact on profits: two-stage estimates are considerably larger than OLS estimates. The final section explores the total estimated redistribution from capital to labor in the manufacturing sector. An important conclusion is that unions raise prices less than was previously believed.

Recently, developments have begun to focus on workplace organization.⁵² Black and Lynch (2001) find that there is an interaction between workplace practice and unionization. When firms have more consensual decision-making in the workplace, then unionization is associated with higher productivity. Conversely, unions are associated with lower productivity in firms that use “traditional” management practices. Doucouliagos and Laroche (2003) carry out a meta-analysis of a large number of studies on the productivity impact of unions. They find that the evidence is broadly consistent with a positive-productivity effect in the United States and a negative effect in the United Kingdom. This is an interesting observation given that the law is quite different in the two jurisdictions. In the United Kingdom, unions cannot commit to a binding agreement, whereas this is possible in the United States. This difference in productivities is consistent with the evidence, though we are far from having convincing causal evidence.

Even if unions enhance firm productivity, this does not necessarily translate into an increase in profits. This depends on the bargaining power of unions, and the extent to which firms earn rents.⁵³ A common strategy is to merge National Labor Relations Data on union certification with firm stock-market data to look at the impact of certification on profits. Ruback and Zimmerman (1984) find that unionization causes a 3.8% fall in equity value. Abowd (1989) finds that unionization can also reduce profits, but that this is an efficient redistribution from firms to workers. These and similar studies must deal with the fact that unionization is endogenous, and hence it is very difficult to estimate the causal impact of a union, independent from other factors. DiNardo and Lee (2004) use a regression discontinuity design in which they compare the outcomes in firms where the unions barely won certification, to ones where there was a close lost. Under the hypothesis that the groups have similar characteristics, then differences can be attributed to union status. Their finding is that the effect is essentially zero.

Lee and Mas (2009) use a similar approach, but obtain a better measure of abnormal stock-market returns. They find that unionization causes a 10% decline in abnormal returns. What is particularly interesting (especially in the light of the efficient-markets hypothesis) is that the negative effect is largest a year after the vote for unionization. This result is consistent with the body of research that looks at the impact of unions on firm productivity (see Doucouliagos and Laroche (2009)). Even if unions truly enhance productivity, if they lower profitability then we should expect firms to reallocate resources to jurisdictions with less union penetration. Kuhn (1998) suggests that there seems to be little evidence of this in Canada, while Machin (2000) finds that new firms in the United Kingdom tend not be unionized.

⁵² See Ichinowski and Shaw (2003), and the Handbook chapters by Bloom and Van Reenen (2011) and Oyer and Schaefer (2011).

⁵³ See Ashenfelter and Johnson (1969) for a classic discussion of union-firm bargaining and associated empirical implications.

As Farber and Western (2001) documents for the United States, and Machin (2000) for the United Kingdom, there has been a large decline in unionization that is consistent with the hypothesis of excessive rent extract by unions in the face of alternative, nonunion, investment opportunities for firms. Yet, there are still some areas with significant union presence. For example, workers in the entertainment industry are unionized. This is an industry where highly skilled actors, writers, and musicians must move from job to job, and where for each job there is a large number of potential candidates. In the absence of a union contract, the wage would be set at the talent's opportunity cost, which is likely to be far below the return necessary to make it profitable to invest in his or her particular skill. In this case, a union has the potential to increase the talent pool, though this is a hypothesis that has not been carefully tested.

The other area with significant union presence is the public sector. Ehrenberg et al. (1983) find that unions do not significantly affect productivity for municipal librarians. Byrne et al. (1996) find that unionized police are less effective in dealing with crime. Eberts and Stone (1987) explore the effect of unions upon teacher productivity, finding that on average unionized teachers increase test scores by about 3%. However, their impact is more homogeneous, and they do not do as well with students with above- or below-average ability. In a recent study, Lovenheim (2009) finds that teacher unions raise costs by about 15% while having no impact on school performance.⁵⁴

In summary, the evidence on unions is consistent with the hypothesis that they do have some bargaining power with respect to the firm. As we discussed in the previous section, bargaining power may allow parties to implement more efficient arrangements. The extent to which this is possible in practice is controversial. It is clear that over the past century we have witnessed a large rise and fall in unionization rates in the United States. The transaction-cost perspective suggests that unions can be viewed as substitutes for legal enforcement of contracts and other forms of labor-market regulation. The fact that employment law in the United States has become stronger in the last 30 years—in the sense of providing more protection for disadvantaged groups⁵⁵ and stronger employment protection—suggests that private law may be providing a substitute for unionization.

In the case of the public sector, the skills acquired by workers are likely to be job-specific, and moreover the demand for these skills are stable. This suggests that the optimal contract is of a long duration, which may explain why public-sector unions are so prevalent. However, it is extremely difficult to measure public-sector productivity, and hence to evaluate properly the available labor institutions for these relationships. Moreover, it is difficult to argue that public-sector unions are purely rent-seeking organizations. As Blank (1994) documents, the public/private sector wage ratio has been

⁵⁴ See Eberts (2007) for a general discussion of the role of teacher unions in education.

⁵⁵ See Chay (1998), Oyer and Schaefer (2002), and Jolls and Prescott (2004) for studies of the labor-market impact of legislation to protect civil rights and disabled individuals.

falling in the United States, even while private-sector unionization has been falling. If unions were the main source of wage growth for workers, then we should observe the opposite. We do not have definite answers to any of these questions, and hence there is much room for further research.

5. DISCUSSION

There is a remarkable consensus that increased employment protection law tends to reduce employment for individuals with less attachment to the labor market. Increases in employment protection law tend to adversely affect matches at the margin. That being said, the economics research uses a relatively crude representation of the law. We know virtually nothing about how specific legal rules interact with different types of worker-firm matches.

At a policy level, employment protection entails changes to specific rules, such as the number of days' notice for a dismissal, mandatory dismissal payments, and specification of the conditions under which a protected employee may be dismissed. At the moment, policymakers have little guidance on how to set these parameters, aside from the blanket recommendation to reduce them all.

Our discussion of the law illustrates that rules evolve in response to specific issues that need to be addressed in the labor market. In the case of common law rules, as in the United States or United Kingdom, a judge may create law in response to a specific set of facts, yet this new law affects all matches. This process is not well understood. The benefit of common-law rule making is that it usually restricts itself to the bounds of particular cases, and therefore is not speculative. However, though judges are aware and are certainly concerned with the broad impact that a decision may have, there is no systematic feedback mechanism for evaluating the consequences of these rule changes.⁵⁶

In a global context, we see a great deal of competition between different legal systems. Firms may opt out of the courts completely by relying upon mandatory arbitration. However, arbitration courts are increasingly looking like public courts, where there is a long process of deposing witnesses and presentation of volumes of evidence before a decision is reached. More generally, all adjudication systems consist of evidence-collection and decision-making that complement the employment practices used by a firm. We need to better understand the substantive role that these courts play in the complex problem of managing human resources.⁵⁷ The literature on employment law has focused on turnover costs. Yet, the discussions of both the law and the literature on transaction costs suggest that information costs are key to understanding the role that the courts play in handling disputes.

⁵⁶ See Krueger (1991), who suggests that the most important feature of the law for private parties is predictability.

⁵⁷ See the chapters by Rebitzer and Taylor (2011) and Van Reenen (2010) on human resource management.

In addition, the focus on turnover costs fails to deal with the selective nature of court decisions. The courts are a venue of last resort for a party who feels that another party has breached a duty. This implies that the law is not applied equally to all individuals. All contracts in the United States are subject to the rule of good faith and fair dealing, which is meant to protect individuals from others, such as managers who blatantly breach their obligations. An issue that is rarely addressed in the economics literature is the extent to which we need courts to protect individuals who face poor treatment from bad managers.⁵⁸ The law, and employment law in particular, exists to deal with specific extreme cases, and not the average employment relationship. In contrast, empirical research on employment law is focused on the average effect.

In this chapter, unions and labor law are discussed as alternative governance structures that may enhance the management of the employment relationship. In theory, allocating more power to workers can be efficiency-enhancing when they have private information that can impact match quality. The difficulty is that such power also results in more rent extraction by workers. The existing evidence is consistent with the hypothesis that union power leads to lower profits for firms. There is also mixed evidence regarding whether or not unions enhance match quality. Unions, like managers, are likely to vary in their ability to strike efficient agreements. Hence, we should not be surprised that there is mixed evidence regarding their effects on match quality. The theory suggests that the rise and fall of unions in the private sector can be explained by the extent to which unions enhance the productive efficiency of firms. At the moment we simply do not have any evidence that explains the observed pattern.⁵⁹

Finally, I have characterized elsewhere the literature on employment law as consisting of *three solitudes* (see MacLeod (2007b)), namely, that the law, the economic theory of contract, and empirical labor economics each has its own aesthetic and group of scholars that have developed for the most part independently of each other. One reason for this is that the law evolved to deal with the pragmatic issue of how to govern the exchange relationship before the analytic tools were in place to study these phenomena. We can view the existing structure of the law as evidence that the central ingredients of contract theory—the insurance motive, asymmetric information and incomplete contracts—are empirically relevant concepts, though we do not know to what extent they provide the best unifying framework.

Ensuring that the economic theory of contract is empirically relevant is not helped by the fact that the theory evolved out of the need to extend the reach of general equilibrium theory, which itself developed from the mathematically sophisticated models of Arrow (1951) and Debreu (1959). The important work of Hart (1975) illustrates a fundamental shortcoming of this approach and shows that when markets are incomplete we cannot

⁵⁸ For example, Sullivan and von Wachter (2009) show that job loss leads to measurable declines in health; hence, there are real costs for workers who invest in a long-term match with a poor employer.

⁵⁹ See Farber and Western (2001) for an explanation for the decline.

expect competitive equilibria to achieve an efficient outcome. This work illustrates that simply extending the model of general equilibrium to deal with incomplete markets is not likely to be a fruitful path.

As discussed in Section 3, the subsequent literature on game theory and mechanism design developed a theory that is much more specific to the nature of individual transactions, and allows one to link transactions costs to the observed structure of employment contracts.⁶⁰ The fact that the law has a long and rich history with its own mode of thought and language has made it difficult to link these abstract models of contract theory to legal practice. The work of Williamson (1975) appeared before much of modern contract theory was developed, and hence his analysis has roots in the legal tradition. As a consequence, his work develops a language that has been influential in introducing the notion of a transaction cost to law and economics, though at the cost of making it very difficult for the scholar schooled in modern theory to tease out the empirical implications of the theory.

In contrast, the third solitude of empirical labor economics uses the model of a competitive market as developed by Marshall (1948) and Samuelson (1947) to successfully organize vast quantities of data on employment and wages, as we can see in the first three volumes of the *Handbook of Labor Economics*. The goal of the empirical literature reviewed in Section 4 is to establish a causal link between changes in law and the union status of workers and the change in employment and wages. Such identification is more credible the closer it is to approximating the treatment-control paradigm that is widely used in science. Hence, there is a bias in this literature towards studying changes that occur over fairly short time periods.⁶¹ Yet the economic theory of contract highlights the fact that individuals enter into agreements because they expect that these relationships will be rewarding in the future, sometimes the distant future. The temporal distance between cause and effect makes it very difficult to explore the implications of these models.

The essence of a contract is that we voluntarily give up freedom of action in the expectation that this reduction will make us better off. It is a fact of life that these great expectations are often dashed, requiring changes to our plans. The history of employment and labor law can be viewed as a sequence of changes that were brought about in the expectation of improving the lives of workers in the long run. We certainly need better ways to measure and evaluate these expectations, so that we may find the optimal trade-off between opportunities of the moment and those that require durable investment and commitment. There is a growing literature that is beginning to explore these issues and the interplay between law and long term outcomes. Harrison and Scorse (2010) explore the effect of anti-sweatshop campaigns and find that they lead to higher

⁶⁰ See in particular the recent books by Laffont and Martimort (2002) and Bolton and Dewatripont (2005).

⁶¹ See Lemieux and MacLeod (2000) for a study where we show that previous studies could not properly identify the effect of unemployment insurance parameters because they ignore learning effects that take many years to work their way through the system.

wages, with little employment lost. Fiess et al. (2010) explore the impact of informal labor markets upon growth.

Finally, Sullivan and von Wachter (2009) show that job loss leads to higher worker mortality. This research is very important because it illustrates why the pecuniary costs and benefits of job loss cannot fully capture the effect of job loss upon individual well being. This can help explain why employment and labor law have for centuries played such an important role in civil society, and why we need more research on the interplay between law, employment contracts, and labor market performance.

REFERENCES

- Abowd, J.M., 1989. The effect of wage bargains on the stock-market value of the firm. *American Economic Review* 79, 774–800.
- Acemoglu, D., Pischke, J.-S., 1998. Why do firms train? Theory and evidence. *Quarterly Journal of Economics* 113, 79–119.
- Acemoglu, D., Pischke, J.-S., 1999. The structure of wages and investment in general training. *Journal of Political Economy* 107, 539–572.
- Acemoglu, D., Shimer, R., 1999. Efficient unemployment insurance. *Journal of Political Economy* 107, 893–928.
- Addison, Teixeira, Grosso, 2000. The effect of dismissals protection on employment: more on a vexed theme. *Southern Economic Journal* 67, 105–122.
- Aghion, E.A., 2008. The unequal effects of liberalization: evidence from dismantling the license Raj in India. *American Economic Review* 98, 1397–1412.
- Aghion, P., Dewatripont, M., Rey, P., 1994. Renegotiation design with unverifiable information. *Econometrica* 62, 257–282.
- Aghion, P., Hermalin, B., 1990. Legal restrictions on private contracts can enhance efficiency. *Journal of Law, Economics, and Organization* 6, 381–409.
- Aghion, P., Tirole, J., 1997. Formal and real authority in organizations. *Journal of Political Economy* 105, 1–29.
- Ahsan, Pages, 2009. Are all labor regulations equal? Evidence from Indian manufacturing. *Journal of Comparative Economics* 37, 62–75.
- Akerlof, G.A., 1982. Labor exchange as partial gift exchange. *Quarterly Journal of Economics* 97, 543–569.
- Alchian, A., Demsetz, H., 1972. Production, information costs, and economic organization. *American Economic Review* 62, 777–795.
- Allen, 1984. Unionized construction workers are more productive. *Quarterly Journal of Economics* 99, 251–274.
- Allen, S.G., 1986. The effect of unionism on productivity in privately and publicly owned hospitals and nursing-homes. *Journal of Labor Research* 7, 59–68.
- Altonji, J.G., Blank, R.M., 1999. Race and gender in the labor market. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*, vol. 3C. North-Holland, Amsterdam, pp. 3143–3260.
- Amin, 2009. Labor regulation and employment in India's retail stores. *Journal of Comparative Economics* 37, 47–61.
- Andreoni, J., Sprenger, C., 2010a. Estimating time preferences from convex budgets. Tech. Rep. University of California, San Diego.
- Andreoni, J., Sprenger, C., 2010b. Risk preferences are not time preferences. Tech. Rep. University of California, San Diego.
- Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 444–455.
- Aoki, M., 2001. *Towards a Comparative Institutional Analysis*. MIT Press, Cambridge, MA.

- Arrow, K.J., 1951. An extension of the basic welfare theorems of classical welfare economics. In: Neyman, J. (Ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, California, pp. 507–532.
- Ashenfelter, O., Johnson, G.E., 1969. Bargaining theory, trade unions, and industrial strike activity. *American Economic Review* 59, 35–49.
- Autor, D.H., 2001. Why do temporary help firms provide free general skills training? *Quarterly Journal of Economics* 116, 1409–1448.
- Autor, D.H., 2003. Outsourcing at will: the contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics* 21, 1–42.
- Autor, D.H., Donohue, J.J., Schwab, S.J., 2004. The employment consequences of wrongful-discharge laws: large, small, or none at all? *American Economic Review* 94, 440–446.
- Autor, D.H., Donohue, J.J., Schwab, S.J., 2006. The costs of wrongful-discharge laws. *Forthcoming Review of Economics and Statistics* 88, 211–231.
- Autor, D.H., Kerr, W.R., Kugler, A.D., 2007. Does employment protection reduce productivity? Evidence from US States. *The Economic Journal* 117, F189–F213.
- Azariadis, C., 1975. Implicit contracts and underemployment equilibria. *Journal of Political Economy* 83, 1183–1202.
- Baily, M.N., 1974. Wages and employment under uncertain demand. *Review of Economic Studies* 41, 37–50.
- Baker, T., 2003. *Insurance Law and Policy*. Aspen Publishers.
- Beaudry, P., DiNardo, J., 1991. The effect of implicit contracts on the movement of wages over the business cycle: evidence from micro data. *Journal of Political Economy* 99, 665–688.
- Becker, Olson, 1990. The effects of the NLRA on stockholder wealth in the 1930s. *Industrial & Labor Relations Review* 44, 116–129.
- Becker, Olson, 1992. Unions and firm profits. *Industrial Relations* 31, 395–416.
- Becker, G., 1962. Investment in human capital: a theoretical analysis. *Journal of Political Economy* 70, 9–49.
- Bemmels, B., 1987. How unions affect productivity in manufacturing plant. *Industrial & Labor Relations Review* 40, 241–253.
- Bentolila, S., Bertola, G., 1990. Firing costs and labour demand: how bad is eurosclerosis? *Review of Economic Studies* 57, 381–402.
- Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119, 249–275.
- Besley, T., Burgess, R., 2004. Can labor regulation hinder economic performance? Evidence from India. *Quarterly Journal of Economics* 119, 91–134.
- Bewley, T.F., 1999. *Why Wages don't Fall during a Recession*. Harvard University Press, Cambridge and London.
- Bird, R.C., Knopf, J.D., 2009. Do wrongful-discharge laws impair firm performance? *Journal of Law & Economics* 52, 197–222.
- Black, S.E., Lynch, L.M., 2001. How to compete: the impact of workplace practices and information technology on productivity. *Review of Economics and Statistics* 83, 434–445.
- Blanchard, Wolfers, 2000. The role of shocks and institutions in the rise of European unemployment: the aggregate evidence. *Economic Journal* 110, 1339–1382.
- Blanchard, O.J., Tirole, J., 2008. The joint design of unemployment insurance and employment protection: a first pass. *Journal of the European Economic Association* 6, 45–77.
- Blank, R.M., 1994. Public sector growth and labor market flexibility: the United States vs. the United Kingdom. In: *Social Protection Versus Economic Flexibility: Is there a Trade-off?* University of Chicago Press.
- Blanpain, R., 2003. *European Labour Law*, 9th and rev. ed., Kluwer Law International, The Hague, Boston.
- Blau, F., Kahn, L., 1999. Institutions and laws in the labor market. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*, vol. 3A. North-Holland, Amsterdam, pp. 1399–1461.
- Blinder, A.S., Choi, D.H., 1990. A shred of evidence on theories of wage stickiness. *Quarterly Journal of Economics* 105, 1003–1015.
- Bloch, M., 1961. *Feudal Society*, Routledge & Kegan Paul.

- Bloom, Nicholas, Van Reenen, John, 2011. Human resource management and productivity. In: Ashenfelter, Orley, Card, David (Eds.), *Handbook of Labor Economics*, vol. 4b. Elsevier, pp. 1697–1767.
- Bolton, P., Dewatripont, M., 2005. *Contract Theory*. MIT Press, Cambridge, MA.
- Botero, J.C., Djankov, S., La Porta, R., Lopez-de Silanes, F., Shleifer, A., 2004. The regulation of labor. *Quarterly Journal of Economics* 119, 1339–1382.
- Breiman, L., 2001. Statistical modeling: the two cultures. *Statistical Science* 16, 199–215.
- Breyer, S., 2009. Economic reasoning and judicial review. *Economic Journal* 119, F135–F215.
- Bronars, S., Deere, D., 1994. Unionization and profitability — evidence of spillover effects. *Journal of Political Economy* 102, 1281–1287.
- Bronars, S.G., Deere, D.R., 1990. Union representation elections and firm profitability. *Industrial Relations* 29, 15–37.
- Brown, C., Medoff, J.L., 1978. Trade unions in the production process. *Journal of Political Economy* 86, 355–378.
- Burguet, R., Caminal, R., 2008. Does the market provide sufficient employment protection? *Labour Economics* 15, 406–422.
- Byrne, D., Dezhbakhsh, H., King, R., 1996. Unions and police productivity: an econometric investigation. *Industrial Relations* 35, 566–584.
- Caballero, R.J., Cowan, K.N., Engel, E.M., Micco, A., 2004. *Effective Labor Regulation and Microeconomic Flexibility*. NBER Working Paper No. 10744.
- Cahuc, P., Postel-Vinay, F., Robin, J., 2006. Wage bargaining with on-the-job search: theory and evidence. *Econometrica* 74, 323–364.
- Card, D., 1986. Efficient contracts with costly adjustment: short-run employment determination for airline mechanics. *American Economic Review* 76, 1045–1071.
- Card, D., Hyslop, D., 1997. Does inflation ‘grease the wheels of the labor market’? In: Romer, C.D., Romer, D.H. (Eds.), *Reducing Inflation: Motivation and Strategy*. University of Chicago Press, Chicago, IL.
- Card, D., Devicienti, F., Maida, A., 2010. Rent-sharing, holdup, and wages: evidence from matched panel data. NBER Working Paper 16192.
- Carmichael, H.L., MacLeod, W.B., 2003. Caring about sunk costs: a behavioral solution to hold-up problems with small stakes. *Journal of Law, Economics and Organization* 19, 106–118.
- Chakravarty, S., MacLeod, W.B., 2009. Contracting in the shadow of the law. *Rand Journal of Economics* 40, 533–557.
- Chay, K.Y., 1998. The impact of federal civil rights policy on black economic progress: evidence from the equal employment opportunity act of 1972. *Industrial and Labor Relations Review* 51, 608–632.
- Che, Y.-K., Hausch, D.B., 1999. Cooperative investments and the value of contracting. *American Economic Review* 89, 125–147.
- Chiappori, P.-A., Salanié, B., Valentin, J., 1999. Early starters versus late beginners. *Journal of Political Economy* 107, 731–760.
- Clark, K.B., 1980a. The impact of unionization on productivity — a case-study. *Industrial & Labor Relations Review* 33, 451–469.
- Clark, K.B., 1980b. Unionization and productivity — micro-econometric evidence. *Quarterly Journal of Economics* 95, 613–639.
- Clark, K.B., 1984. Unionization and firm performance — the impact on profits, growth, and productivity. *American Economic Review* 74, 893–919.
- Coase, R., 1937. The nature of the firm. *Economica* 4, 386–405.
- Cochrane, J.H., 1991. A simple test of consumption insurance. *The Journal of Political Economy* 99, 957–976.
- Cooke, W.N., 1994. Employee participation programs, group-based incentives, company performance — a union-nonunion comparison. *Industrial & Labor Relations Review* 47, 594–609.
- Cousineau, J.-M., Lacroix, R., Bilodeau, D., 1983. The determination of escalator clauses in collective agreements. *Review of Economics and Statistics* 65, 196–202.
- Crawford, V.P., 1982. A theory of disagreement in bargaining. *Econometrica* 50, 607–637.
- Crawford, V.P., 1988. Long-term relationships governed by short-term contracts. *The American Economic Review* 78 (3), 485–499.

- de Barros, P., Corseuil, 2004. The impact of regulations on Brazilian labor market performance. In: *Law and Employment: Lessons from Latin America and the Caribbean*. The University of Chicago Press.
- Debreu, G., 1959. *Theory of Value*. Yale University Press, New Haven, CT.
- Dessein, W., 2002. Authority and communication in organizations. *Review of Economic Studies* 69, 811–838.
- Di Tella, R., MacCulloch, R., 2005. The consequences of labor market flexibility: panel evidence based on survey data. *European Economic Review* 49, 1225–1259.
- Diamond, P.A., Maskin, E., 1979. An equilibrium analysis of search and breach of contract, I: steady states. *Bell Journal of Economics* 10, 282–316.
- DiNardo, J., Lee, D.S., 2004. Economic impacts of new unionization on private sector employers: 1984–2001. *Quarterly Journal of Economics* 119, 1383–1441.
- Djankov, S., Hart, O., McLiesh, C., Shleifer, A., 2008. Debt enforcement around the world. *Journal of Political Economy* 116, 1105–1150.
- Djankov, S., La Porta, R., Lopez-de Silanes, F., Shleifer, A., 2003. Courts. *Quarterly Journal of Economics* 118, 453–517.
- Djankov, S., Ramalho, R., 2009. Employment laws in developing countries. *Journal of Comparative Economics* 37, 3–13.
- Doucoulagos, C., Laroche, P., 2003. What do Unions do to productivity? A meta-analysis. *Industrial Relations* 42, 650–691.
- Doucoulagos, H., Laroche, P., 2009. Unions and profits: a meta-regression analysis. *Industrial Relations* 48, 146–184.
- Downes, E.A., 2004. Labor market regulation and employment in the Caribbean. In: *Law and Employment: Lessons from Latin America and the Caribbean*. The University of Chicago Press.
- Dunford, B.B., Devin, D.J., 1998. Employment at-will and employee discharge: a justive perspective on legal action following termination. *Personnel Psychology* 51, 903–934.
- Dye, R.A., 1985. Optimal length of labor contracts. *International Economic Review* 26, 251–270.
- Eberts, R., Stone, J.A., 1987. Teacher unions and the productivity of public schools. *Industrial & Labor Relations Review* 40, 354–363.
- Eberts, R.W., 2007. Teachers unions and student performance: help or hindrance? *Future of Children* 17, 175–200.
- Ehrenberg, R.G., Sherman, D., Schwartz, J., 1983. Unions and productivity in the public sector: a study of municipal libraries. *Industrial & Labor Relations Review* 36, 199–213.
- Ehrlich, I., Posner, R.A., 1974. An economic analysis of legal rulemaking. *The Journal of Legal Studies* 3, 257–286.
- Elmeskov, Martin S., 1998. Key lessons for labour market reforms: evidence from OECD countries' experience. *Swedish Economic Policy Review* 5.
- Epstein, R.A., 1984. In defense of the contract at will. *The University of Chicago Law Review* 51, 947–982.
- Eslava, E.A., 2004. The effects of structural reforms on productivity- and profitability-enhancing reallocation: evidence from Colombia. *Journal of Development Economics* 75, 333–371.
- Farber, H.S., Western, B., 2001. Accounting for the decline of unions in the private sector, 1973–98. *Journal of Labor Research* 22, 459–485.
- Feess, E., Muehlheusser, G., 2002. Economic consequences of transfer fee regulations in European football. *European Journal of Law and Economics* 13, 221–237.
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114, 817–868.
- Feinman, J.M., 1978. The development of the employment at will rule. *American Journal of Legal History* 20, 118–135.
- Feldmann, H., 2008. Business regulation and labor market performance around the world. *Journal of Regulatory Economics* 33, 201–235.
- Feldmann, H., 2009. The unemployment effects of labor regulation around the world. *Journal of Comparative Economics* 37, 76–90.
- Fiess, N.M., Fugazza, M., Maloney, W.F., 2010. Informal self-employment and macroeconomic fluctuations. *Journal of Development Economics* 91, 211–226.

- Freeman, 2002. Institutional Differences and Economic Performance Among OECD Countries. CEP Discussion Papers.
- Freeman, R.D., Medoff, J.L., 1984. What do Unions do? Basic Books, New York, NY.
- Friedman, M., 1962. Capitalism and Freedom. University of Chicago Press, Chicago, with the assistance of Rose D. Friedman. 23 cm.
- Fuchs, W., 2006. Contracting with repeated moral hazard and private evaluations. Tech. Rep. University of Chicago, Chicago.
- Fuller, L.L., Perdue Jr., William R., 1936. The reliance interest in contract damages: 1. The Yale Law Journal 46, 52–96.
- Garibaldi, M., 1999. Deconstructing job creation. IMF Working Paper No. 99/109.
- Gibbons, R., 1987. Piece rate incentive schemes. Journal of Labor Economics 5, 413–429.
- Gibbons, R., 1997. Incentives and careers in organizations. In: Kreps, D.M., Wallis, K.F. (Eds.), Advances in Economics and Econometrics: Theory and Applications. Cambridge University Press, Cambridge, UK, pp. 1–37.
- Gibbons, R., Katz, L.F., 1991. Layoffs and lemons. Journal of Labor Economics 9, 351–380.
- Gibbons, R., Katz, L.F., Lemieux, T., Parent, D., 2005. Comparative advantage, learning, and sectoral wage determination. Journal of Labor Economics 23, 681–723.
- Goetz, C.J., Scott, R.E., 1980. Enforcing promises: an examination of the basis of contract. Yale Law Journal 89, 1261–1322.
- Goldberg, V.P., 1976. Regulation and administered contracts. Bell Journal of Economics 7, 426–448.
- Gould IV, W.B., 2004. A Primer on American Labor Law. MIT Press.
- Greif, A., 1989. Reputation and coalitions in medieval trade — evidence on the Maghribi traders. Journal of Economic History 49, 857–882. times Cited: 79.
- Grossman, S.J., Hart, O.D., 1981. Implicit contracts, moral hazard, unemployment. American Economic Review 71, 301–307.
- Grossman, S.J., Hart, O.D., 1986. The costs and benefits of ownership: a theory of vertical and lateral integration. Journal of Political Economy 94, 691–719.
- Grout, P., 1984. Investment and wages in the absence of binding contracts: a Nash bargaining approach. Econometrica 52, 449–460.
- Grubb, Wells, 1993. Employment regulation and patterns of work in EC countries. OECD Economic Studies 21.
- Guerrieri, V., 2008. Inefficient unemployment dynamics under asymmetric information. Journal of Political Economy 116, 667–708.
- Hall, R.E., Lazear, E.P., 1984. The excess sensitivity of layoffs and quits to demand. Journal of Labor Economics 2, 233–257.
- Hardin, G., 1968. Tragedy of commons. Science 162, 1243–1248.
- Harris, M., Holmström, B., 1982. A theory of wage dynamics. Review of Economic Studies 49, 315–333.
- Harrison, A., Scorse, J., 2010. Multinationals and anti-sweatshop activism. American Economic Review.
- Hart, O., 1975. On the optimality of equilibrium when the market structure is incomplete. Journal of Economics Theory 11, 418–443.
- Hart, O.D., 1995. Firms, Contracts and Financial Structure. Oxford University Press, Oxford, UK.
- Hart, O.D., Holmström, B., 1987. The theory of contracts. In: Bewley, T. (Ed.), Advances in Economic Theory: Fifth World Congress. Cambridge University Press, Cambridge, UK, pp. 71–155.
- Hart, O.D., Moore, J., 1988. Incomplete contracts and renegotiation. Econometrica 56, 755–785.
- Hart, O.D., Moore, J., 2004. Agreeing now to agree later: contracts that rule out but do not rule in. Tech. Rep. 10397. NBER, Cambridge, MA.
- Hashimoto, M., 1981. Firm specific capital as a shared investment. American Economic Review 71, 475–482.
- Heckman, J.J., Pagés, C., 2000. The cost of job security regulation: evidence from Latin American labor markets. Economia 1, 109–154.
- Hirsch, B.T., 1991. Union coverage and profitability among US firms. Review of Economics and Statistics 73, 69–77.
- Hirsch, B.T., Connolly, R., 1987. Do unions capture monopoly profits? Industrial & Labor Relations Review 41, 118–136.

- Holland, P.W., 1986. Statistics and causal inference. *Journal of the American Statistical Association* 81, 945–960.
- Holmes Jr., O.W., 1881. *The Common Law*. Little, Brown, Boston, MA.
- Hopenhayn, H., Nicolini, J., 1997. Optimal unemployment insurance. *Journal of Political Economy* 105, 412–438.
- Hopenhayn, H., Rogerson, R., 1993. Job turnover and policy evaluation: a general equilibrium analysis. *Journal of Political Economy* 101, 915–938.
- Howitt, P., 2002. Looking inside the labor market: a review article. *Journal of Economic Literature* 40, 125–138. tY - JOUR Accession Number: 0603937. Geographic Descriptors: US. Publication Type: Journal Article. Update Code: 200205.
- Ichinowski, C., Shaw, K., 2003. Beyond incentive pay: insiders' estimates of the value of complementary human resource management practices. *Journal of Economic Perspectives* 17, 155–180.
- Imbens, G.W., Lemieux, T., 2008. Regression discontinuity designs: a guide to practice. *Journal of Econometrics* 142, 615–635.
- Imbens, G.W., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47, 5–86.
- Jackson, M.O., 2001. Mechanism theory. In: *The Encyclopedia of Life Support Systems*.
- Jensen, M., Meckling, W., 1976. Theory of the firm: managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3, 305–360.
- Johnson, S., McMillan, J., Woodruff, C., 2002. Courts and relational contracts. *Journal of Law, Economics, and Organization* 18, 221–277.
- Jolls, C., 2007. Employment law. In: Polinsky, A.M., Shavell, S. (Eds.), *Handbook of Law and Economics*, vol. 2. Elsevier, Amsterdam, The Netherlands (chapter 17).
- Jolls, C., Prescott, J.J., 2004. Disaggregating employment protection: the case of disability discrimination. Tech. Rep. National Bureau of Economic Research, Cambridge, MA.
- Joskow, P.L., 1988. Price adjustment in long-term-contracts—the case of coal. *Journal of Law & Economics* 31, 47–83.
- Kahn, L.M., 2010. Employment protection reforms, employment and the incidence of temporary jobs in Europe: 1996–2001. *Labour Economics* 17, 1–15.
- Kanemoto, Y., MacLeod, W.B., 1992. The ratchet effect and the market for second hand workers. *Journal of Labor Economics* 10, 85–92.
- Kaplan, D., 2009. Job creation and labor reform in Latin America. *Journal of Comparative Economics* 37, 91–105.
- Kessler, D.P., Rubinfeld, D.L., 2007. Empirical study of the civil justice system. In: Polinsky, A.M., Shavell, S. (Eds.), *Handbook of Law and Economics*, vol. I. Elsevier B. V, Oxford, UK.
- Kornhauser, L., MacLeod, W.B., 2010. Contracts between legal persons. In: Gibbons, R., Roberts, J. (Eds.), *Handbook of Organizational Economics*. Princeton University Press.
- Krueger, A., Mas, A., 2004. Strikes, scabs, and tread separations: labor strife and the production of defective Bridgestone/Firestone tires. *Journal of Political Economy* 112, 253–289.
- Krueger, A.B., 1991. The evolution of unjust-dismissal legislation in the United States. *Industrial and Labor Relations Review* 44, 644–660.
- Kugler, A., 2004a. The effect of job security regulations on labor market flexibility: evidence from the Colombian labor market reform. In: Heckman, J.J., Pags, C. (Eds.), *Law and Employment: Lessons from Latin American and the Caribbean*. The University of Chicago Press, Chicago.
- Kugler, A., 2005. Wage-shifting effects of severance payments savings accounts in Colombia. *Journal of Public Economics* 89, 487–500.
- Kugler, A.D., 1999. The impact of firing costs on turnover and unemployment: evidence from the Colombian labour market reform. *International Tax and Public Finance* 6, 389–410.
- Kugler, A.D., 2004b. How do firing costs affect worker flows in a world with adverse selection? *Journal of Labor Economics* 22, 553–584.
- Kuhn, P., 1992. Mandatory notice. *Journal of Labor Economics* 10, 117–137.
- Kuhn, P., 1998. Unions and the economy: what we know; what we should know. *Canadian Journal of Economics-Revue Canadienne D'Economie* 31, 1033–1056.

- Kuhn, P., Robert, J., 1989. Seniority and distribution in a 2-worker trade union. *Quarterly Journal of Economics* 104, 485–505.
- La Porta, R., Lopez-de Silanes, F., Shleifer, A., 2008. The economic consequences of legal origins. *Journal of Economics Literature* 46, 285–332.
- Laffont, J.-J., Martimort, D., 2002. *The Theory of Incentives*. Princeton University Press, Princeton, NJ.
- Laing, D., 1994. Involuntary layoffs in a model with asymmetric information concerning worker ability. *Review of Economic Studies* 61, 375–392.
- Lazear, E.P., 1990. Job security provisions and employment. *Quarterly Journal of Economics* 105, 699–726.
- Lee, D.S., Lemieux, T., 2009. Regression discontinuity designs in economics. Tech. Rep. 14723. NBER, Cambridge, MA.
- Lee, D.S., Mas, A., 2009. Long-run impacts of unions on firms: new evidence from financial markets, 1961–1991. Tech. Rep. 14709. Princeton University, Cambridge, MA.
- Lee, D.S., McCrary, J., 2005. Crime, punishment, and myopia. Tech. Rep. 11491. NBER, Cambridge, MA.
- Lemieux, T., MacLeod, W.B., 2000. Supply side hysteresis: the case of the Canadian unemployment insurance system. *Journal of Public Economics* 78, 139–170.
- Levin, J., 2003. Relational incentive contracts. *American Economic Review* 93, 835–857.
- Levine, D.I., 1991. Just-cause employment policies in the presence of worker adverse selection. *Journal of Labor Economics* 9, 294–305.
- Lovell, Sickles, Warren, 1988. The effect of unionization on labor productivity: some additional evidence. *Journal of Labor Research* 9, 55–63.
- Lovenheim, M.F., 2009. The effect of teachers' unions on education production: evidence from union election certifications in three midwestern states. *Journal of Labor Economics* 27, 525–587.
- Machin, S., 1991. The productivity effects of unionization and firm size in British engineering firms. *Economica* 58, 479–490.
- Machin, S., 2000. Union decline in Britain. *British Journal of Industrial Relations* 38, 631–645.
- MacLeod, W.B., 2003. Optimal contracting with subjective evaluation. *American Economic Review* 93, 216–240.
- MacLeod, W.B., 2005. Regulation or markets? The case of employment contracts. *Cesifo Economic Studies* 51, 1–46.
- MacLeod, W.B., 2007a. Reputations, relationships and contract enforcement. *Journal of Economics Literature* XLV, 597–630.
- MacLeod, W.B., 2007b. Three solitudes in contract: law, data, and theory. *Scottish Journal of Political Economy* 54, 606–616.
- MacLeod, W.B., Malcomson, J.M., 1989. Implicit contracts, incentive compatibility, and involuntary unemployment. *Econometrica* 57, 447–480.
- MacLeod, W.B., Malcomson, J.M., 1993. Investments, holdup, and the form of market contracts. *American Economic Review* 83, 811–837.
- MacLeod, W.B., Nakavachara, V., 2007. Legal default rules: the case of wrongful discharge laws. *Economic Journal* 117, F1–F62.
- MacNeil, I.R., 1974. The many futures of contracts. *Southern California Law Review* 47, 691–816.
- Marshall, A., 1948. *The Principles of Economics*. Macmillan, New York, NY.
- Mas, A., 2006. Pay, reference points, and police performance. *Quarterly Journal of Economics* 121, 783–821.
- Mas, A., 2008. Labour unrest and the quality of production: evidence from the construction equipment resale market. *Review of Economic Studies* 75, 229–258.
- Matouschek, N., Ramezzana, P., Robert-Nicoud, F., 2009. Labor market reforms, job instability, and the flexibility of the employment relationship. *European Economic Review* 53, 19–36.
- McDonald, I.M., Solow, R.M., 1981. Wage bargaining and employment. *American Economic Review* 71, 896–908.
- McLaughlin, K., 1994. Rigid wages. *Journal of Monetary Economics* 34, 383–414.
- Miles, T.J., 2000. Common law exceptions to employment at will and US labor markets. *Journal of Law Economics & Organization* 16, 74–101.
- Milgrom, P., 1988. Employment contracts, influence activities, and efficient organizational design. *Journal of Political Economy* 96, 42–60.

- Mincer, J., 1962. On-the-job training: cost, returns and some implications. *Journal of Political Economy* 70, 50–79.
- Mitchell, M.W., Stone, J.A., 1992. Union effects on productivity — evidence from western United States sawmills. *Industrial & Labor Relations Review* 46, 135–145.
- Mondino, Montoya, 2004. The effects of labor market regulations on employment decisions by firms: empirical evidence for Argentina. In: *Law and Employment: Lessons from Latin America and the Caribbean*. The University of Chicago Press.
- Montenegro, C.E., Pagés, C., 2004. Who benefits from labor market regulations? Chile 1960–1998. In: Heckman, James J., Pagés, Carmen (Eds.), *Law and Employment: Lessons from Latin America and the Caribbean*. The University of Chicago Press, Chicago, IL, pp. 401–434.
- Moore, J., Repullo, R., 1988. Subgame perfect implementation. *Econometrica* 56, 1191–1220.
- Moore, J.H., 1985. Optimal labour contracts when workers have a variety of privately observed reservation wages. *Review of Economic Studies* 52, 37–67.
- Morris, A.P., 1995–1996. Bad data, bad economics, and bad policy: time to fire wrongful discharge law. *University of Texas Law Review* 74, 1901.
- Myerson, R.B., Satterthwaite, M.A., 1983. Efficient mechanisms for bilateral trading. *Journal of Economic Theory* 29, 265–281.
- Naidu, S., Yuchtman, N., 2009. How Green was my Valley? Coercive contract enforcement in 19th century industrial Britain. Tech. Rep. Harvard University.
- Nickell, S., 1997. Unemployment and labor market rigidities: Europe versus North America. *Journal of Economic Perspectives* 11 (3), 55–74.
- Nicoletti, G., Scarpetta, S., Lane, P., 2003. Regulation, productivity and growth—OECD evidence. *Economic Policy*.
- OECD, 1993. *Employment Outlook*. OECD.
- OECD, 1994. *The Jobs Study: Evidence and Explanations*. OECD, Paris, France.
- OECD, 1996. *Employment Outlook*. OECD, Paris, France.
- OECD, 1999. *Employment Outlook*. OECD, Paris, France.
- OECD, 2003. *Employment Outlook*. OECD, Paris, France.
- Ostrom, E., 2000. Collective action and the evolution of social norms. *Journal of Economic Perspectives* 14, 137–158.
- Oyer, P., Schaefer, S., 2002. Litigation costs and returns to experience. *American Economic Review* 92, 683–705.
- Oyer, P., Schaefer, S., 2011. Personnel economics: hiring and incentives. In: Ashenfelter, Orley, Card, David (Eds.), *Handbook of Labor Economics*, vol. 4b. Elsevier, pp. 1769–1823.
- Pagés, C., Montenegro, C.E., 2007. Job security and the age-composition of employment: evidence from Chile. *Estudios de Economía* 34 (2), 109–139.
- Pauly, M.V., 1968. The economics of moral hazard. *American Economics Review* 58, 531–537.
- Peierls, R.E., 1960. Wolfgang Ernst Pauli. 1900–1958. *Biographical Memoirs of Fellows of the Royal Society* 5, 175–192.
- Pissarides, C.A., 2001. Employment protection. *Labour Economics* 8, 131–159.
- Posner, R., 2008. *How Judges Think*. Harvard University Press.
- Posner, R.A., 1974. Theories of economic regulation. *The Bell Journal of Economics and Management Science* 5, 335–358. fLA 00058556 American Telephone and Telegraph Company Copyright 1970 American Telephone and Telegraph Company.
- Posner, R.A., 2003. *Economic Analysis of Law*, 6th ed., Little, Brown and Company, Boston, USA.
- Postel-Vinay, F., Robin, J., 2002. Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica* 70, 2295–2350.
- Pries, M., Rogerson, R., 2005. Hiring policies, labor market institutions, and labor market flows. *Journal of Political Economy* 113, 811–839.
- Rebitzer, J.B., Taylor, L.J., 2011. Extrinsic rewards and intrinsic motives: standard and behavioral approaches to agency and labor markets. In: Ashenfelter, Orley, Card, David (Eds.), *Handbook of Labor Economics*, vol. 4a. Elsevier, pp. 701–772.
- Rogerson, R., Shimer, R., Wright, R., 2005. Search-theoretic models of the labor market: a survey. *Journal of Economic Literature* 43, 959–988.

- Rogerson, W.P., 1984. Efficient reliance and damage measures for breach of contract. *RAND Journal of Economics* 15, 39–53.
- Rogerson, W.P., 1992. Contractual solutions to the hold-up problem. *Review of Economic Studies* 59, 777–793.
- Rosen, S., 1985. Implicit contracts: a survey. *Journal of Economic Literature* 23, 1144–1175.
- Rothstein, M.A., Liebman, L., 2003. *Employment Law*. Thomson West, New York, NY.
- Ruback, R.S., Zimmerman, M.B., 1984. Unionization and profitability: evidence from the capital market. *Journal of Political Economy* 92, 1134–1157.
- Saavedra, J., Torero, M., 2004. Labor market reforms and their impact on formal labor demand and job market turnover: the case of Peru. In: *Law and Employment: Lessons from Latin America and the Caribbean*. The University of Chicago Press.
- Salinger, 1984. Tobin's q , unionization, and the concentration–profits relationship. *The RAND Journal of Economics* 15, 159–170.
- Samuelson, P., 1947. *Foundations of Economic Analysis*. Harvard University Press.
- Scarpetta, S., 1996. Assessing the role of labour market policies and institutional settings on unemployment: a cross-country study. *OECD Economic Studies*.
- Schwartz, A., Scott, R.E., 2003. Contract theory and the limits of contract law. *The Yale Law Journal* 113, 541–619.
- Shaked, A., Sutton, J., 1984. Involuntary unemployment as a perfect equilibrium in a bargaining model. *Econometrica* 52, 1351–1364.
- Shapiro, C., Stiglitz, J.E., 1984. Equilibrium unemployment as a worker discipline device. *American Economic Review* 74, 433–444.
- Shavell, S., 1984. The design of contracts and remedies for breach. *Quarterly Journal of Economics* 99, 121–148.
- Simon, H., 1957. *Administrative Behavior*. The MacMillan Company, New York.
- Simon, H.A., 1951. A formal theory of the employment relationship. *Econometrica* 19, 293–305.
- Stahler, N., 2008. Firing costs, severance payments, judicial mistakes and unemployment. *Labour Economics* 15, 1162–1178.
- Stephenson, M.C., 2009. Legal realism for economists. *Journal of Economic Perspectives* 23, 191–211.
- Sullivan, D., von Wachter, T., 2009. Job displacement and mortality: an analysis using administrative data. *Quarterly Journal Of Economics* 124, 1265–1306.
- Tirole, J., 1999. Incomplete contracts: where do we stand? *Econometrica* 67, 741–782.
- Townsend, R.M., 1994. Risk and insurance in village India. *Econometrica* 62, 539–591.
- von Wachter, T., 2007. In the right place at the wrong time: the role of firms and luck in young workers' careers. *American Economic Review*.
- Voos, Mishel, 1986. The union impact on profits: evidence from industry price–cost margin data. *Journal of Labor Economics* 4, 105–133.
- Williamson, O.E., 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. The Free Press, New York.
- Williamson, O.E., 1991. Comparative economic organization: the analysis of discrete structural alternatives. *Administrative Science Quarterly* 36, 269–296.
- Williamson, O.E., Wachter, M.L., Harris, J.E., 1975. Understanding the employment relation: the analysis of idiosyncratic exchange. *Bell Journal of Economics* 6, 250–278.
- Wood, H.G., 1877. *A Treatise on the Law of Master and Servant*. J.D. Parsons, Albany, NY.