Essays on the Economics of Crime and Econometric Methodology

by

Sean Michael May

B.Sc. (Honours), Mathematics
Queen's University, 1996

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2000

Author ....
....
Department of Economics
May 5, 2000

Certified by...
....
Joshua D. Angrist
Professor of Economics
Thesis Supervisor

Certified by...
....
K. Daron Acemoglu
Professor of Economics
Thesis Supervisor

Accepted by..
....
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Departmental Committee on Graduate Students

Essays on the Economics of Crime and Econometric Methodology

by

Sean Michael May

Submitted to the Department of Economics
on May 5, 2000, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

The research presented in this thesis covers two topics: the economics of crime and econometric methodology. The first chapter addresses the question of whether higher wages reduce teenage crime rates. I exploit exogenous variation in the wages of teenagers resulting from federal minimum wage legislation. Instrumental variables estimates show a strong negative relationship between wages and arrest rates for burglary, larceny, motor vehicle theft, vandalism, and robbery. Wage elasticities of property crime arrest rates range between -1 and -2. In contrast with the results for property crime, wages do not have a strong impact on arrest rates for most violent crime. The second chapter examines the effect of crime on the labor market outcomes of victims. I use longitudinal data from the National Crime Victimization Survey to estimate the employment-related costs of crime. Estimates suggest that being the victim of a violent crime causes a transitory decline in the employment rates and household income of victims of 2 to 3 percent. Victims of property crime do not show a significant decline in employment rates or household income as a result of the crime. For victims of violent crime, average lost earnings are roughly $700, a figure close to estimates of the total property loss and medical costs suffered by victims of violent crime. The third chapter contains work, joint with Bryan Brown and Whitney Newey, that describes a relatively efficient moment-restricted bootstrap for generalized method of moments estimators. We show the bootstrap improves on the standard asymptotic approximation and illustrate that the bootstrap improvement can be large, as evidenced by Monte Carlo simulations and an empirical example in dynamic panel data models.

Thesis Supervisor: Joshua D. Angrist
Title: Professor of Economics

Thesis Supervisor: K. Daron Acemoglu
Title: Professor of Economics

Acknowledgments

I have benefited from the suggestions and assistance of my thesis advisors: Josh Angrist, Daron Acemoglu, and Whitney Newey. David Autor, Steve Pischke, and Jim Poterba also provided helpful comments. The Social Science and Humanities Research Council of Canada provided financial support.

Thanks to my friends Jon Guryan, John Johnson, and Melissa Schettini for their thoughts and advice. I spent much of my time at MIT in the company of Adam Ashcraft, Campe Goodman, and Salma Qarnain; I am grateful for their friendship and their help.

Special thanks to my family.

Contents

List of Figures

List of Tables

Introduction

The research presented in this thesis covers two topics: the economics of crime and econometric methodology. The first chapter uses changes in federal minimum wage legislation to demonstrate that teenagers commit less property crime as their earnings opportunities in legal sectors of the economy improve. The second chapter examines the effect of crime on the labor market outcomes of victims. Empirical results show that victims of violent crime experience transitory declines in employment rates and household income. The third chapter contains work, joint with Bryan Brown and Whitney Newey, that describes a relatively efficient bootstrap technique for generalized method of moments estimators. We demonstrate that this methodology improves on the standard asymptotic approximations to the distribution of the estimator.

Chapter 1 addresses the question of whether higher wages reduce teenage crime rates. The recent dramatic fall in crime rates combined with the remarkable growth of the U.S. economy has focused attention on whether improvements in the labor market prospects of young workers can explain this trend in crime. To gauge the importance of this link, I exploit exogenous variation in the wages of teenagers resulting from federal minimum wage legislation. Instrumental variables estimates show a strong negative relationship between wages and arrest rates for burglary, larceny, motor vehicle theft, vandalism, and robbery. Wage elasticities of property crime arrest rates range between -1 and -2. In contrast with the results for property crime, wages do not have a strong impact on arrest rates for most violent crime. These estimates suggest that rising wages may account for as much as 30 percent of the recent decline in teenage arrest rates. A rough calculation suggests that the marginal social benefit of reduced crime is approximately equal to the costs of the teenage wage subsidy required to attain this reduction. This calculation is probably a conservative estimate of the benefit of a policy that targets at-risk groups.

Chapter 2 focuses on the labor force outcomes of victims of violent and property crime. Because crime is so widespread—in 1998, roughly one in ten Americans was the victim of a crime—it is important to understand the consequences of crime for the victims. Conventional estimates of the cost of crime have included the cost of stolen or damaged property, medical costs, and the value of the reduction in victims' quality of life. These estimates, however, exclude earnings losses caused by victimization. Previous work by psychologists has documented both high rates of incidence of posttraumatic stress disorder (PTSD) among victims of crime and lower rates of employment among sufferers of PTSD. I use longitudinal data from the National Crime Victimization Survey to estimate the employment-related costs of crime. Estimates suggest that being the victim of a violent crime causes a transitory decline in the employment rates of victims of roughly 2 percent, but that victims also experience a decline in employment rates prior to victimization. Results using household income as an outcome show that violent crime is associated with a transitory 2 to 3 percent decline in income. Victims of property crime do not show a significant decline in employment rates or household income as a result of the crime. For victims of violent crime, average lost earnings are roughly \$700, a figure close to estimates of the total property loss and medical costs suffered by victims of violent crime.

Chapter 3, co-authored with Bryan Brown and Whitney Newey, describes a relatively efficient bootstrap method for generalized method of moments (GMM) estimators that uses a moment-restricted distribution. This approach is motivated by the concern that the usual asymptotic theory can be a poor approximation to the distribution of the estimator, particularly when there are many overidentifying restrictions or the parameters of interest are not well identified. The bootstrap provides one approach to improvements in this approximation. We show that our moment-restricted distribution estimator attains the semiparametric efficiency bound for estimation of the distribution of a single observation, and that the bootstrap improves on the standard asymptotic approximation. We also illustrate that the bootstrap improvement can be large, as evidenced by Monte Carlo simulations and an empirical example in dynamic panel data models.

Chapter 1

Wages and Youth Arrests

1.1 Introduction

Economists, criminologists, and policy makers have long speculated on the relationship between economic conditions and crime. Becker (1968) proposes that "a rise in the income available in legal activities ... would reduce the incentive to enter illegal activities and reduce the numbers of offenses." Others have noted that the relationship between the economy and crime also runs in the opposite direction: high crime rates may inhibit economic growth. For example, arrest and incarceration may lower the future earnings and employment prospects of the offender, as suggested by previous studies. In addition, high levels of crime may encourage employers to change location or discourage new business formation, contributing to slower economic growth and a decline in regional employment prospects. The economy and crime may be linked through other channels as well. In particular, property crime rates have been shown to increase with economic growth. Criminologists speculate that a strong economy increases the quantity and value of consumer goods that can be stolen, thereby raising the returns to property crime.[1]

Renewed interest in this relationship has been generated by the low levels of crime in

[1]Grogger (1995) finds transitory effects of arrests on earnings and employment that are moderate in magnitude. Kling (1999) finds small effects of incarceration on long-term employment rates, but larger effects on future earnings. Bound and Freeman (1992) use the NLSY and find large effects of incarceration on future employment prospects. Cullen and Levitt (1996) find that rising crime rates in cities causes urban flight. Willis (1997) finds that crime has a negative impact on new business creation. Schwartz and Exter (1990) and Norström (1988) document a positive relationship between economic growth and property crime.

recent years. The National Crime Victimization Survey esimates that there were approximately 31 million criminal victimizations in 1998, the fewest number recorded since 1973. This low level of crime is not the result of a long, gradual decrease in crime rates; it is the result of a recent, dramatic reduction in crime. From 1993 to 1998, violent crime victimization rates have fallen by roughly 27 percent and property crime victimization rates have seen an even sharper decrease of 32 percent. A variety of explanations have been advanced to explain this trend in crime rates: increases in the number of police officers per capita and changes in the law enforcement strategies used by the police; growth in the number of criminals incarcerated; declines in the drug trade; rising private expenditures on security guards and protection; and improvements in the labor market prospects of young workers.[2]

The remarkable growth of the U.S. economy in recent years has focused attention on this last explanation. Although many studies have examined the relationship between crime and the economy, these studies have, for the most part, measured the effect of unemployment on the level of crime. Such studies tend to find that lower rates of unemployment are associated with lower levels of crime, but that the effect is moderate in magnitude (Freeman, 1999). In addition, even the largest estimate of the effect of unemployment on crime is too small to explain much of the variation in crime. The uncertain conclusions of this empirical evidence on the relationship between economic conditions and crime may be attributed, in part, to two factors. First, unemployment rates are only one measure of the labor market prospects faced by criminals; potential earnings may also play a role in crime participation decisions. Second, previous studies may fail to take into account that high crime rates inhibit economic growth. This failure may bias estimates of the responsiveness of crime to economic conditions.

This chapter exploits exogenous variation in teenage wages to address the question of whether youth crime is responsive to economic conditions. I compare state-level changes in the mean log wage of teenagers and the change in teenage arrest rates between 1989 and 1992. I employ an instrumental variables strategy that uses the 1990 and 1991 increases in the federal minimum wage as a source of exogenous variation in the wages of teenagers. Two-

[2]Many academic and press articles have commented on the decline in crime rates. Volume 88 of the *Journal of Criminal Law and Criminology* contains a sample of academic views.

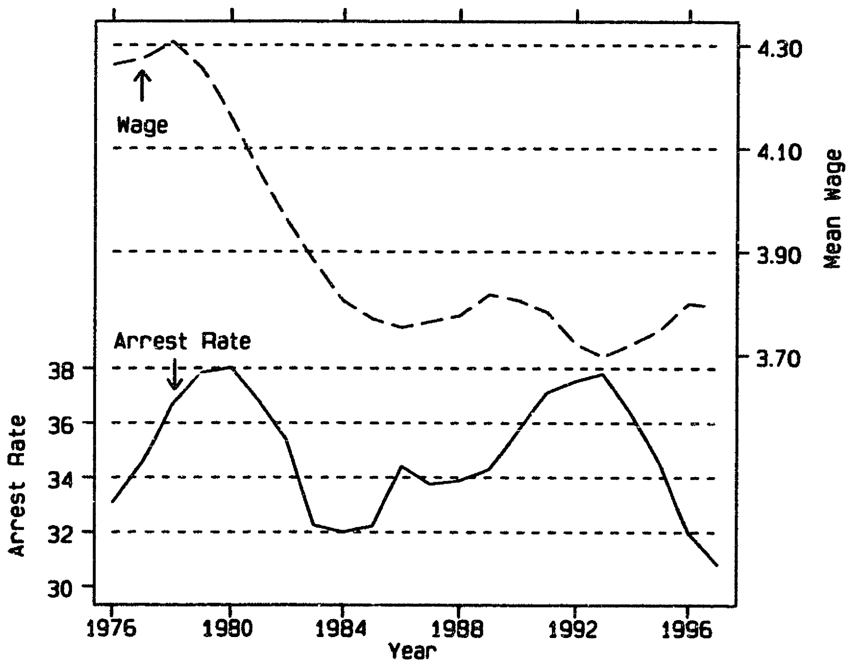

Figure 1-1: This figure shows the time-series of arrest rates and average real wages for 16–19 year olds in the United States between 1976–1997. The dotted line is the real wage series, the solid line is the series for arrest rates. Arrest data are taken from the Uniform Crime Reports; wage data are taken from March CPS.

stage least squares estimates show that while arrest rates for violent crimes do not respond strongly to changes in wages, participation in burglary, motor vehicle theft, vandalism, and robbery is negatively related to market wages. Using these estimates, elasticities of arrest rates with respect to market wages are between −1 and −2 for property crime. These results suggest that rising wages may account for as much as 30 percent of the fall in youth arrest rates in recent years.

The chapter is organized as follows. First, background information on previous work relating economic conditions to crime is presented. Also discussed is the decline in real wages and employment opportunities that characterized the labor market for low-skilled workers during the 1980s, followed by an overview of the legislation that increased the federal minimum wage from $3.35 per hour in 1989 to $4.25 per hour in 1991. The third section outlines a simple model that describes three mechanisms through which wages might affect criminal activity. The fourth section contains a brief description of the data sources used in this chapter and a reduced form analysis of the effect of the minimum wage change on arrest rates. Two-stage least squares estimates of the effect of teenage wages on arrest rates are reported in the fifth section. Calculations that estimate the social costs (in increased wage bills) and benefits (in reduced crime) of a wage increase for teenagers are also presented. The sixth section examines the robustness of the instrumental variables estimates. This section focuses on state-specific trends in arrest rates and possible disemployment effects of a minimum wage increase. The conclusion follows.

1.2 Background

1.2.1 Market Work and Youth Crime

Empirical studies show that many young criminals consider possible work opportunities in both the legal and illegal sectors of the economy. Grogger (1998) analyzes data from a sample of young men from the NLSY who reported income from crime in the previous year. Almost all worked. Moreover, an almost identical fraction of criminals and non-criminals reported positive weeks worked in the previous year. Fagan and Freeman (1997) review studies documenting the patterns of employment among criminals. These studies suggest that criminals actively allocate their time between overtime work and crime, youths regularly shift between crime and work, and that employment has only a small effect on the crime participation decisions of teenagers. Freeman (1999) concludes "one interpretation of the porous boundary between crime and legitimate work is that young offenders are engaged in an active process of income optimization, taking advantage of economic opportunities that present themselves."

Since a number of studies have measured relatively modest economic returns from crime, it is not surprising that many young criminals are also employed. The National Bureau of Economic Research (NBER) conducted two such surveys: the 1980 Survey of Inner City Youths in Boston, Chicago, and Philadelphia (ICY); and the 1989 Boston Youth Survey (BYS). Freeman (1991) found that individuals who reported committing criminal acts in the previous year made an average of $1607 [$2423 in 1989 dollars] from illegal activities in 1979–1980. The BYS shows roughly comparable figures for Boston youths in 1989. Respondents who reported earnings from crime made on average about $3000 per year from these illegal activities. Average hourly pay from crime was $19, but for those who reported committing crime on a weekly basis the figure was only $9.75 per hour. Respondents reported earnings from legal sources of $7.50 per hour.

A study similar to the NBER work was conducted by the RAND Corporation. Reuter et al. (1990) interviewed a sample of men between the ages of 18 and 40 resident in the District of Columbia. For those respondents who identify themselves as making daily sales of drugs, the median monthly net income from drug sales was $2000. For respondents who reported lower frequencies of drug sales, median monthly net income was correspondingly lower: $830 for those who made sales two or more days a week, and $50 for those reporting sales one day per week or less. The median wage from legal employment was $7 per hour in the full sample and $5 per hour in a sample of 18–24 year olds.

A straight comparison of the returns to crime and wages, however, ignores two distinguishing features of earnings from crime: income from illegal sources escapes taxation, but there are hazards associated with criminal activity that do not play a role in labor market activities.[3] Furthermore, although crime may have a higher return, arrests and prison terms will reduce the lifetime earnings of an offender.[4]

Most previous work on the relationship between economic conditions and crime focused on the effects of employment on crime. Two studies which have attempted to measure the responsiveness of crime to earnings are Schmidt and Witte (1984) and Grogger (1998). Schmidt and Witte (1984) find negative but insignificant effects of wages on arrest rates for three types of offense. Contrary to these findings, Grogger (1998) concludes that criminal activity is responsive to market wages. He employs a structural model of time allocation between leisure, market work, and criminal activities to identify the effect of wage incentives on criminal participation. Estimates of the wage elasticity of crime participation range from -0.95 to -1.20.

[3]In the study by Reuter et al. (1990), a survey of adolescents classified as "frequent dealers" finds that 38 percent see arrest as being very likely in a year of drug dealing, 25 percent see a prison sentence as being very likely, and 50 percent see severe injury or death as being a very likely outcome.

[4]Kling (1999) examines the employment and earnings of people convicted of committing serious crimes. He finds small negative effects of incarceration on long-term employment rates, and larger negative effects on long-term earnings of $200 to $600 per quarter (this corresponds to 10–30 percent of mean earnings prior to criminal charges).

These studies suggest that crime is highly responsive to wages; however, none of them exploit exogenous variation in wages. Consequently, they may be biased because of the negative relationship between arrest and incarceration and future employment prospects and earnings, or because high crime rates have been shown to contribute to declines in economic conditions.

1.2.2 Federal Minimum Wage Legislation

The decline in the real wages of low skilled men from the mid-1970s through the 1980s is well documented. Juhn, Murphy, and Pierce (1993), for example, estimate that the real wage of workers in the tenth percentile of the wage distribution fell by 25 percent from 1970 through 1989. Accompanying the drop in pay, the same period saw an erosion in the hours worked by those in the lower tail of the wage distribution (Juhn, Murphy, and Topel, 1991). The correlation between falling real wages and rising levels of crime in the 1980s has not gone unnoticed by economists. Freeman (1996) comments: "labor market incentives influence the supply of men to crime, and, in particular, that the collapse of the job market for less skilled men during the 1980s and 1990s may have contributed to their increased rate of criminal activity." Similarly, Bound and Freeman (1992) examine the widening differentials in black-white earnings and employment for young men from the mid-1970s through the 1980s. They speculate that, among other things, the deteriorating job market for the less skilled may have made crime more attractive by reducing the returns to legitimate employment.

Amidst this climate of falling wages for low skilled workers in the 1980s, the federal minimum wage was pegged at $3.35 per hour for most of the decade, a level at which it had been set in early 1981. The declining real value of the federal minimum wage prompted a number of state legislatures to pass legislation which guaranteed higher wages to its workers than the federal standard. The first of these states, Maine, enacted legislation in January of 1985. By 1989, 16 states and the District of Columbia had passed legislation that set a state-specific minimum wage above the federal minimum wage of $3.35 per hour.

By 1989 inflationary pressures had reduced the real value of the minimum wage to its lowest level since the 1950s. Attempts by the Democratic Congress to raise the federal

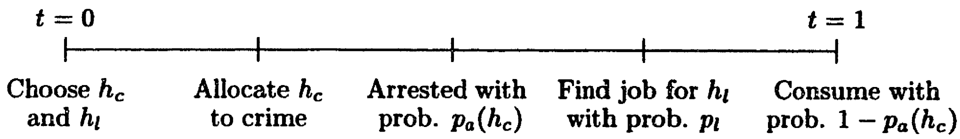

Figure 1-2: Timing of the Model.

minimum wage succeeded in late 1989. The new law raised the minimum wage in two steps: effective April 1, 1990 the minimum wage would increase to \$3.80 per hour, and on April 1, 1991 it rose a further \$0.45 to \$4.25 per hour.

This policy change forms the basis of the subsequent analysis. Using 1989 as a pre-treatment year and 1992 as a post-treatment year, the relationship between the impact of the minimum wage increase and arrest rates is estimated. The impact of the minimum wage change is used as an instrument for the change in the logarithm of the wages of teenagers between 1989 and 1992.

1.3 Theoretical Framework

The following model illustrates three important factors in the relationship between wages and criminal activity: a wage increase makes crime less attractive by increasing the returns to legitimate employment; a wage increase may increase the unemployment rate and subsequently making it more difficult for criminals to find employment; and a wage increase has a deterrent effect as it increases the opportunity costs of incarceration.

The model has a single time period in which an individual allocates his time between market work, h_l, and criminal activity, h_c. He maximizes his expected income subject to the time constraint that $h_l + h_c = T$, where T is his total endowment of time. He takes as given his wage, w. Returns to crime are described by the function $r(h_c)$, where it is assumed that $r' > 0$ and $r'' < 0$. Uncertainty in the model arises from two sources. The individual is arrested with probability $p_a(h_c)$, where $p_a' > 0$ and $p_a'' > 0$. In the event that the individual is arrested, he can neither retain the earnings from his criminal activity, nor is he able to participate in the labor market. The individual also faces uncertainty in the labor market. Conditional on not having been arrested, the probability of finding a job is a function of the market wage, and is denoted $p_l(w)$.

The problem of the individual can then be written as

$$\max_{h_l, h_c} \ (1 - p_a(h_c)) \left[p_l(w) w h_l + r(h_c)\right] \tag{1.1}$$

subject to the constraint

$$h_l + h_c = T. \tag{1.2}$$

The first order condition for the maximization problem is

$$(1 - p_a(h_c))[r'(h_c) - p_l(w)w] = p_a'(h_c)[p_l(w)w(T - h_c) + r(h_c)]. \tag{1.3}$$

To simplify matters, first examine the case in which the probability of arrest is constant. Under this assumption, $p_a' \equiv 0$ and the first order condition reduces to

$$r'(h_c) = p_l(w)w. \tag{1.4}$$

(Here it is assumed that $r'(0)$ is large enough to eliminate corner solutions in which the individual allocates all his time to market work and none to criminal activity.) The first order condition for the maximization problem simply equates the return from criminal activity to the expected return from market work.

From equation (1.4), the response of an individual to an exogenous change in the wage rate—such as that caused by an increase in the minimum wage—is:

$$\frac{\partial h_c}{\partial w} = \frac{p_l'(w)w + p_l(w)}{r''(h_c)}. \tag{1.5}$$

Ignoring the dependence of the probability of finding a job on the wage (formally, this corresponds to the case where $p_l(w)$ is constant and $p_l' \equiv 0$), time allocated to crime declines in response to the wage increase because of diminishing marginal returns to criminal activity ($r'' < 0$).

Returning to the case in which $p_l(w)$ is non-constant, neoclassical models of labor markets predict that an increase in the market wage rate—due to a minimum wage increase, say—may have disemployment effects. This relationship between employment and wages can be modeled as $p_l' < 0$. The response of the individual to a change in wage is now ambiguous. The increase in wage is offset by a decreased probability of finding employment; the individual may increase or decrease h_c. Models of monopsony employment or models where search or other frictions give firms market power provide different predictions of the effect of a minimum wage increase on employment. In such models, an increase in the minimum wage may have positive employment effects: a rising minimum wage increases both $p_l(w)$ and w, leading to an unambiguous reduction in time allocated to crime.

This simple example serves to illustrate the first two important points concerning the effect of an exogenous wage increase on the level of crime. Rising wages have a direct effect on the level of crime by increasing the return to market work. Counteracting this effect is a decrease in the probability of finding a job, making criminal pursuits more lucrative.

Returning to the model in which the probability of arrest is a function of the amount of time allocated to crime, the first order condition characterizing the optimal choice of h_c is

$$\underbrace{(1 - p_a(h_c))r'(h_c)}_{1} = \underbrace{(1 - p_a(h_c))p_l(w)w}_{2} + \underbrace{p_a'(h_c)I}_{3}. \tag{1.6}$$

In equation (1.6), $I \equiv p_l(w)w(T - h_c) + r(h_c)$ is the expected income of the individual conditional on not being arrested, term 1 is the expected marginal return to crime, term 2 is the expected marginal cost of crime due to a reduction in the amount of time devoted to market work, and term 3 is the expected marginal cost of crime due to an increase in the probability of being arrested.

Assuming that there is an exogenous increase in the market wage rate, w, the two effects discussed previously remain. Wages increase the return to market work, but they also affect the probability of finding employment. Now, however, there is a third effect. An increase in wages makes crime more costly by increasing the amount of income the individual loses in the event that he is arrested. As a result, reductions in crime may result from a wage increase due to a rise in the opportunity cost of being arrested.

The model presented here does not incorporate leisure: an individual choses to allocate his time solely to work or crime. With leisure in the model, there is a fourth channel through which an increase in wages may affect crime. Certain types of crime such as vandalism or motor vehicle theft may be motivated not by economic concerns, but instead may be committed because they provide some entertainment to the criminal. An increase in market wages may decrease time allocated to leisure activities, resulting in a reduction in recreational crimes.

1.4 Data Sources and Reduced Form Estimates

1.4.1 Wage Data

Wage data were compiled from the monthly files of the Current Population Survey (CPS). Each month, the outgoing rotation groups (the structure of the CPS includes one-quarter of the households in the full monthly sample in each outgoing rotation group) in the CPS are asked to provide supplementary information on their weekly hours and earnings. The sample of wage-earners in the data set was limited to those between the ages of 16 and 19. There are three reasons for this restriction. Teenagers typically earn low wages and are more likely to be affected by minimum wage legislation. Teenagers, as a group, exhibit a high propensity to commit crime. The third reason for the age restriction is that the attachment of adult criminals to illegal activities may be stronger than that of younger criminals. Consequently, we expect that the elasticity of criminal activity with respect to wages is higher for teenage criminals.

An hourly wage series for teenager workers is constructed as follows: For those respondents who identify themselves as being paid on a hourly basis, it is simply reported hourly wage. For those respondents who identify themselves as being salaried workers, the hourly wage is imputed as the ratio of usual weekly earnings to usual weekly hours. All of the analysis contained in this chapter was performed at the state level. Accordingly, the mean logarithm of hourly wages in 1989 (before the federal minimum wage increase) and in 1992 (after the federal minimum wage increase) was constructed for each state.

The fraction of teenager workers in 1989 affected by the law change was constructed as a measure of the impact of the legislation. More specifically, the measure of the state-level impact used was the fraction of teenage workers for whom the hourly wage was between the state minimum wage and $4.25 an hour. In all that follows, state minimum wage is used to mean the greater of the federal minimum wage in 1989 [$3.35 per hour] and the minimum wage rate set by the state legislature. This fraction ranged from a minimum of zero percent in California, Connecticut, the District of Columbia, and Rhode Island (states that, for the greater part of 1989, had a state minimum wage of at least $4.25 per hour) to a high of 74.4 percent in Kentucky. The mean for all states was 40.1 percent with a standard deviation of 22.0 percent.

1.4.2 Data on Criminal Activity

The Uniform Crime Reports (UCR) contain data on arrests and crimes reported to police; they are tabulated by the Federal Bureau of Investigation (FBI) using the records of local law enforcement agencies that are submitted to the FBI on a voluntary basis.[5] Although the UCRs contain data on arrests and crime reported to police, I limit my attention to data on arrests.[6] Characteristics of offenders such as age are not known for all crimes reported to the police, making it difficult to directly analyze the effects of a wage increase on the crimes committed by an age cohort.

The UCR data used in this chapter includes state-level counts of arrests by age and offense category. Arrest data for three types of property crime—burglary, larceny-theft, and motor vehicle theft—and four types of violent crime—murder, forcible rape, robbery, and assault—is taken from this source. Arrests for a fourth category of property crime, vandalism, are available only by two broad age categories: juvenile (under the age of 18) and adult. To match arrest data with wage data from the CPS, I impute state arrests of 16–19 year olds for vandalism by multiplying juvenile arrests by the national ratio of teenage to juvenile arrests for vandalism.

Florida is omitted from the subsequent analysis because of the lack of arrest data in

[5]Although participation in the program is voluntary, in 1983 the participating law enforcement agencies represented over 98 percent of the population of the United States living in SMSAs and 97 percent of the total population of the United States.

[6]A problem with this approach is that arrests are a function of both the level of crime and of the enforcement efforts of police agencies. Levitt (1998), for example, investigates the observed negative relationship between arrest rates and crime. He attributes the negative relationship to a deterrence effect: As arrest rates increase for one type of offense, criminals substitute towards other types of offenses.

Table 1.1: Descriptive Statistics.

	1989	1992	Change from 1989 to 1992
Fraction of Affected Teenagers	0.4014	—	—
	(0.2204)		
Log Hourly Wages	1.452	1.560	0.109
	(0.111)	(0.070)	(0.063)
Burglary Arrest Rate	6.738	6.694	-0.0807
	(2.464)	(2.931)	(1.647)
Larceny Arrest Rate	18.46	19.38	0.7103
	(6.788)	(7.179)	(4.168)
Motor Vehicle Theft Arrest Rate	4.382	4.312	-0.0519
	(2.290)	(2.846)	(1.087)
Vandalism Arrest Rate	4.562	6.148	1.574
	(2.435)	(3.269)	(1.269)
Murder Arrest Rate	0.2704	0.3785	0.1107
	(0.2298)	(0.2316)	(0.0950)
Rape Arrest Rate	0.3260	0.3785	0.0521
	(0.1150)	(0.1351)	(0.1101)
Robbery Arrest Rate	2.321	2.999	0.7222
	(1.988)	(2.181)	(0.5908)
Assault Arrest Rate	3.781	4.925	1.189
	(1.803)	(2.207)	(1.266)

Notes: Means with standard deviations in parentheses are shown. In all cases, the reported arrest rates refer to the arrest rates for persons aged 16–19 per 1000. Weights equal to 1989 state teenage population were used in the column labeled *1989* and *Change from 1989 to 1992*. Weights equal to 1992 state teenage population were used in the column labeled *1992*.

1989.

1.4.3 Reduced Form Estimates

The reduced form relationship between the minimum wage legislation and arrest rates for teenagers is shown graphically in Figure 1-3. The figure graphs the times series of arrest rates for 16–19 year olds between 1984 and 1992, grouped by the fraction of teenagers in 1989 earning between the state minimum wage and $4.25 per hour. *Low Impact*, *Medium Impact*, and *High Impact* states are states in which the fraction of affected teenagers was, respectively, between 0 and 20 percent, between 20 and 40 percent, and over 40 percent.

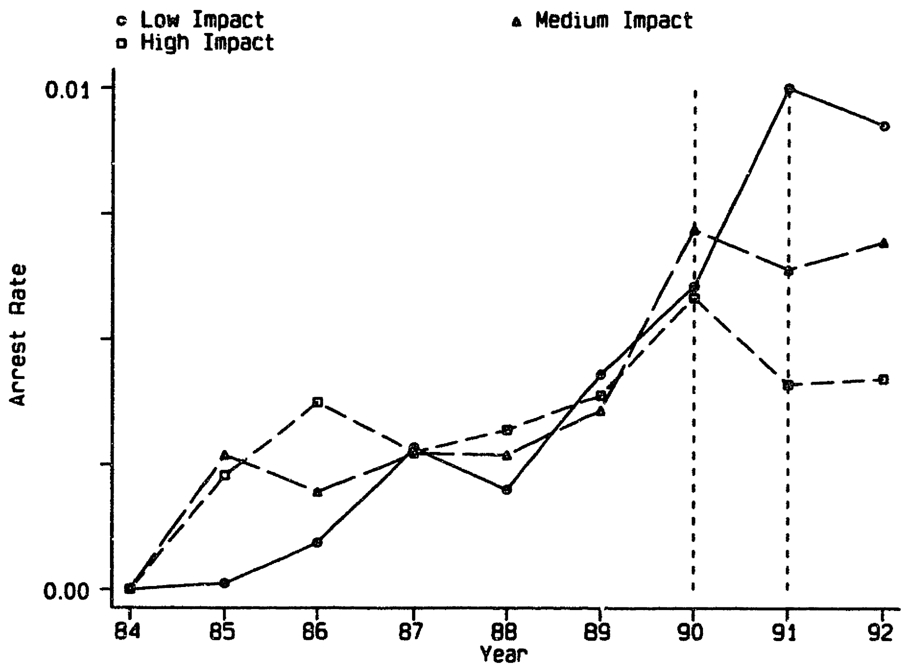

Figure 1-3: This figure shows the time-series of arrest rates for 16–19 year olds in the United States between 1984–1992 by fraction of affected teenagers. *Low-Impact*, *Medium-Impact*, and *High Impact* states are those in which the fraction of affected teenagers in 1989 was less than 20 percent, between 20 and 40 percent, and over 40 percent. Arrest rates in 1984 were normalized to be zero. Vertical lines are drawn in 1990 and 1991, the years of the federal minimum wage increase.

To eliminate level differences in arrest rates, arrest rates are normalized to be zero in 1984. Arrests appear to follow similar trends before 1990, especially in the years immediately preceding the federal mimimum wage changes. In 1990 and 1991 (the two years marked by vertical lines), however, there is a divergence in arrest rates between the three groups. Arrest rates in low-impact states continue to trend upwards, arrest rates in medium-impact states break from the pre-existing trend and decrease, and arrest rates in high-impact states decrease by the largest amount. This figure suggests that states which experienced the largest increases in the wages of teenagers also experienced greater declines in teenage arrests.

This observation can be formalized in a regression framework as follows. Denote the change in teenage arrest rates between 1989 and 1992 in state i and offense type c by Δy_{ic}. In what follows, f_i is the fraction of teenagers in 1989 earning between the state minimum wage and \$4.25 per hour in state i, and Δa_{ic} is the change in adult arrest rates between

1989 and 1992 for offense type c in state i. The first specification is

$$\Delta y_{ic} = \alpha_c + \beta_c f_i + u_{ic}. \tag{1.7}$$

For each value of c—corresponding to the eight offense categories: larceny-theft, burglary, motor vehicle theft, vandalism, murder, forcible rape, robbery, and assault—equation (1.7) was estimated using least squares. Results are shown in Table 1.2.

A second specification adds the change in arrest rates for 25–44 year olds as a regressor. Addition of the change in adult arrest rates attempts to control for state-level changes in arrests caused, for example, by changes in law enforcement or policing. Using the previous notation, this specification can be written

$$\Delta y_{ic} = \alpha_c + \beta_c f_i + \gamma_c \Delta a_{ic} + u_{ic}. \tag{1.8}$$

This equation was estimated separately for each value of c; the results appear in Table 1.2.

Arrests are a function of both the level of criminal activity and of enforcement, so controlling for changes in law enforcement policies may give a clearer picture of the effect of the minimum wage change on youth crime. Introducing the change in adult arrest rates as a regressor, however, possibly confuses a control variable with an outcome of the policy change. Suppose, for example, adult criminals are low-skilled workers and, as such, their wages are appreciably affected by minimum wage legislation. Wage increases then increase the earnings of adult criminals, possibly resulting in changes in adult arrest rates.

On the other hand, if the wages of adult criminals are unaffected by minimum wage legislation, using changes in adult arrest rates is a valid control strategy. This assumption is supported by wage data from the 1989 CPS: only 2.7 percent of men aged 25–44 earn wages low enough to be directly affected by minimum wage legislation. A regression of the change in mean log wages of adults between 1989 and 1992 on the fraction of affected teenagers in 1989 produces a coefficient of 0.0281 with a standard error of 0.0347. States with more teenagers affected by the minimum wage legislation did not experience significantly different changes in the wages of 25–44 year olds between 1989 and 1992.[7]

[7] It is also possible to measure whether states experiencing greater movements in the wage distribution

Table 1.2: Reduced-Form Estimates for Changes in State Arrest Rates, 1989–1992.

A. Property Crimes	Burglary		Larceny		Auto Theft		Vandalism	
Mean of Dependent Variable in 1992	6.694		19.38		4.312		6.148	
Fraction of Affected Teenagers	-1.773 (1.048)	-2.714 (0.6929)	-0.0306 (2.730)	-4.515 (1.482)	0.3172 (0.7107)	-1.472 (0.6363)	-1.667 (0.7957)	-2.681 (0.7337)
Change in Adult Arrest Rate	—	3.512 (0.4323)	—	2.750 (0.2428)	—	5.821 (1.024)	—	3.396 (0.8301)

B. Violent Crimes	Murder		Rape		Robbery		Assault	
Mean of Dependent Variable in 1992	0.3785		0.3785		2.999		4.925	
Fraction of Affected Teenagers	0.0128 (0.0622)	-0.0198 (0.0568)	0.0762 (0.0712)	-0.0071 (0.0551)	-1.126 (0.3512)	-1.330 (0.3019)	-0.2371 (0.8284)	-0.5277 (0.3647)
Change in Adult Arrest Rate	—	1.521 (0.4360)	—	1.344 (0.2174)	—	2.180 (0.4938)	—	1.333 (0.0939)

Notes: Standard errors are shown in parentheses under parameter estimates. All specifications include a constant. The dependent variables are changes in state arrest rates for burglary, larceny-theft, motor vehicle theft, vandalism, murder, forcible rape, robbery, and assault per 1000 for 16–19 year olds from 1989 to 1992. *Fraction of Affected Teenagers* is the fraction of teenagers earning between the state minimum wage and $4.25 per hour in 1989. *Change in Adult Arrest Rate* is the change in adult arrest rates for the corresponding offense type between 1989 and 1992. All estimates were calculated using weighted least squares, with weight equal to the number of teenagers in the state in 1989.

In what follows, the fraction of teenagers earning between the state minimum wage and $4.25 per hour in 1989, f_i, serves as a measure of the impact of the minimum wage change. States that have a higher fraction of affected teenagers should see a larger increase in teenage wages as a result of the federal minimum wage legislation. Correspondingly, we expect to see greater declines in the arrest rates for states in which the wages of teenagers increased the most. For crimes that provide a source of income, economic theory predicts that β_c should be less than zero. This reduced-form interpretation of the minimum wage change is supplemented in the next section, where f_i is used to instrument for change in mean log wages between 1989 and 1992.

of teenagers also experienced larger declines in the arrest rates of adults. A regression of the fraction of affected teenagers in 1989 on the change in adult violent and property crime rates between 1989 and 1992 yields coefficients and standard errors of 1.243 and 1.524 for property crime, and −0.5642 and 1.225 for violent crime.

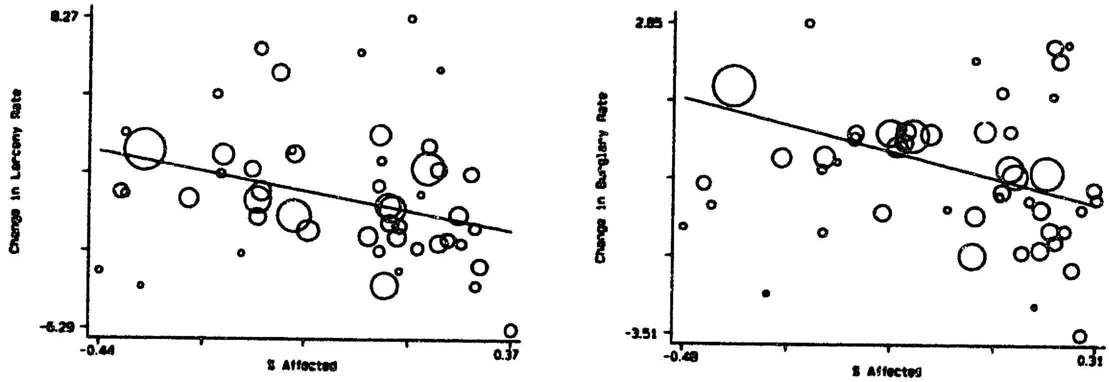

Figure 1-4: These figures show the reduced-form relationship between the fraction of teenagers affected by the minimum wage change and the change in arrest rates for two types of property crime: burglary and larceny. Larger circles correspond to states with higher populations. Also shown is the estimated regression line. The slope estimates and standard errors are −2.714 and 0.6929 for burglary, and −4.515 and 1.482 for larceny. Complete regression results are presented in panel A of Table 1.2.

Estimation results for property crimes are presented in panel A of Table 1.2. These results show a stronger correlation between minimum wage impact and a reduction in arrest rates. Controlling for changes in the adult arrest rate, burglary, larceny, and motor vehicle theft show negative and statistically significant coefficients on the fraction of affected teenagers, confirming the hypothesis about the relationship between wages and crime. The estimate of β_c for vandalism is also negative and statistically significant. While vandalism is not a source of income, a reduction in vandalism may signal an increase in social order.[8] It may also signal a decrease in the amount of time allocated by teenagers to recreational crimes.

Estimation results for violent crimes are shown in panel B of Table 1.2. With the exception of robbery, the effect of the wage change on violent crime is statistically insignificant. The incidents of murder and forcible rape that occur are most likely not acts with economic motivations (the possibility does exist, of course, of murders committed by drug traffickers, or rapes accompanied by robbery). Assault shares the feature with murder and forcible

[8] Among other things, the mayor of New York city has recently targeted vandals as part of well-publicized city-wide crack down on "quality of life offenses." It is believed by some that a policy of zero-tolerance for minor offenses eliminates a potential breeding ground for violence and other types of serious crime.

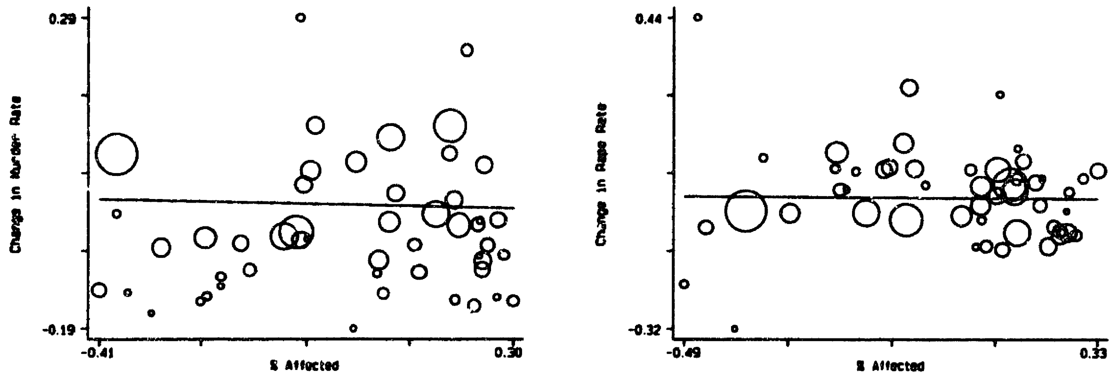

Figure 1-5: These figures show the reduced-form relationship between the fraction of teenagers affected by the minimum wage change and the change in arrest rates for two types of violent crime: murder and rape. Larger circles correspond to states with higher populations. Also shown is the estimated regression line. The slope estimates and standard errors are −0.0198 and 0.0568 for murder, and −0.0071 and 0.0551 for rape. Complete regression results are presented in panel B of Table 1.2.

rape that it is a crime that provides no income to the criminal.[9] Confirming this view, the estimated coefficient for assault is small in magnitude and statistically insignificant.

Robbery differs fundamentally from the three other types of violent crime analyzed in this chapter. It is defined by the FBI as "the taking or attempting to take anything of value from the care, custody, or control of a person or persons by force or threat of force or violence and/or by putting the victim in fear." Because robbery is a source of income, it provides criminals with an alternative to employment; as such, robbery is expected to decline as wages increase. The statistically significant negative coefficient for robbery, combined with the pattern of insignificant relationships for crimes that do not result in economic gain, suggests that the correlation between minimum wage impact and arrest rates is not merely spurious.

Limiting attention to the statistically significant coefficient estimates, elasticities of the arrest rate with respect to the minimum wage impact were calculated. Coefficient estimates from the second specification in Table 1.2 were used, and the elasticities were evaluated at the mean levels of f_i and the 1992 mean arrest rates shown in Table 1.1. This methodology

[9] Aggravated assault is "an unlawful attack by one person upon another for the purpose of inflicting severe or aggravated bodily injury" by the FBI. Assaults that occur in the course of a theft are classified by the FBI as robberies.

produced predicted arrest elasticities of −0.16 for burglary, −0.10 for larceny, −0.14 for motor vehicle theft, −0.18 for vandalism, and −0.18 for robbery.

1.5 The Effect of Wages on Youth Arrests

1.5.1 Estimation and Results

As discussed in the introduction, a simple regression of arrest rates on wages may produce biased estimates of the effect of wages on youth crime. To identify the effect of wages on arrest rates, the minimum wage legislation of 1989 is employed as a source of an exogenous increase in the wages of teenagers. The magnitude of the policy intervention is quantified by using the fraction of teenagers earning between the state minimum wage and $4.25 per hour in 1989; this variable is used as an instrument for changes in average log wages.

The models used to estimate the effect of wage changes on juvenile crime are

$$\Delta y_{ic} = \alpha_c + \beta_c \Delta w_i + u_{ic} \tag{1.9}$$

and

$$\Delta y_{ic} = \alpha_c + \beta_c \Delta w_i + \gamma_c \Delta a_{ic} + u_{ic}. \tag{1.10}$$

Here, as before, Δy_{ic} is the change in teenage arrest rate between 1989 and 1992 in state i for offense category c, and Δa_{ic} is the change in adult arrest rate between 1989 and 1992 in state i for offense category c. Denote by Δw_i the change in mean log hourly wage for 16–19 year olds between 1989 and 1992. The subscript c indexes the eight types of crime: burglary, larceny-theft, motor vehicle theft, vandalism, murder, forcible rape, robbery, and assault.

For each category of crime these models were estimated using two-stage least squares (2SLS); the fraction of teenagers earning between the state minimum wage and $4.25 per hour in 1989 ($f_i$ in previous notation) serves as an instrument for the change in average log wages. For comparison, equation (1.10) was also estimated by ordinary least squares. Estimates computed by ordinary least squares appear in columns labeled *OLS* in Table 1.3, while the two-stage least squares estimates are in columns labeled *2SLS*.

Evidence documenting the first-stage relationship between change in the mean log wage of teenagers and the minimum wage legislation reveals that the legislation had a strong impact on teenage wages.[10] The strength of this relationship is confirmed by a regression of the change in mean log wage of 16–19 year olds between 1989 and 1992 on the fraction of teenagers earning between the state minimum wage and $4.25 per hour in 1989. The estimated coefficient from the first stage is 0.2202 with a standard error of 0.0289; the R^2 from this regression is 0.4988. Inclusion of controls for the change in adult arrest rates between 1989 and 1992 does not alter the first-stage relationship significantly: the estimated coefficient from the first stage ranges between 0.1917 with a standard error of 0.0338 and 0.2110 with a standard error of 0.0290.

Visual evidence of this relationship can be seen in Figure 1-6. The states were grouped by fraction of affected teenagers—the three groups correspond to states in which the fraction of affected teenagers is less than 20 percent, between 20 percent and 40 percent, and over 40 percent—and kernel density estimates of the wage distribution for teenagers in 1989 and 1992 were calculated. These estimates reveal that states characterized by a large fraction of affected teenagers saw a more pronounced shift in the wage distribution of teenagers.

Results for property crimes are shown in panel A of Table 1.3. The OLS results show that wages and arrest rates for property crimes are negatively correlated. The 2SLS results mirror the negative relationship between changes in arrest rates and change in mean log wages for burglary, larceny, motor vehicle theft, and vandalism seen in the reduced-form estimates (Table 1.2). The coefficients on mean log wages in the models for all types of property crime are negative and significant after controlling for changes in the arrest rate of 25–44 year olds. A comparison the OLS estimates and the 2SLS estimates of β_c for larceny reveals an interesting trend: the OLS coefficient estimates for property crimes are uniformly larger than those predicted by the instrumental variables approach.

The bias in the OLS estimates may be caused by positive relationship between economic growth and property crime rates.[11] An article recently published in The Times (1999) calls

[10]Discussion of the impact of the minimum wage on the earnings of teenagers can be found in Card (1991) or Card and Krueger (1995).

[11]Schwartz and Exter (1990) document that states experiencing a high rate of economic growth and pop-

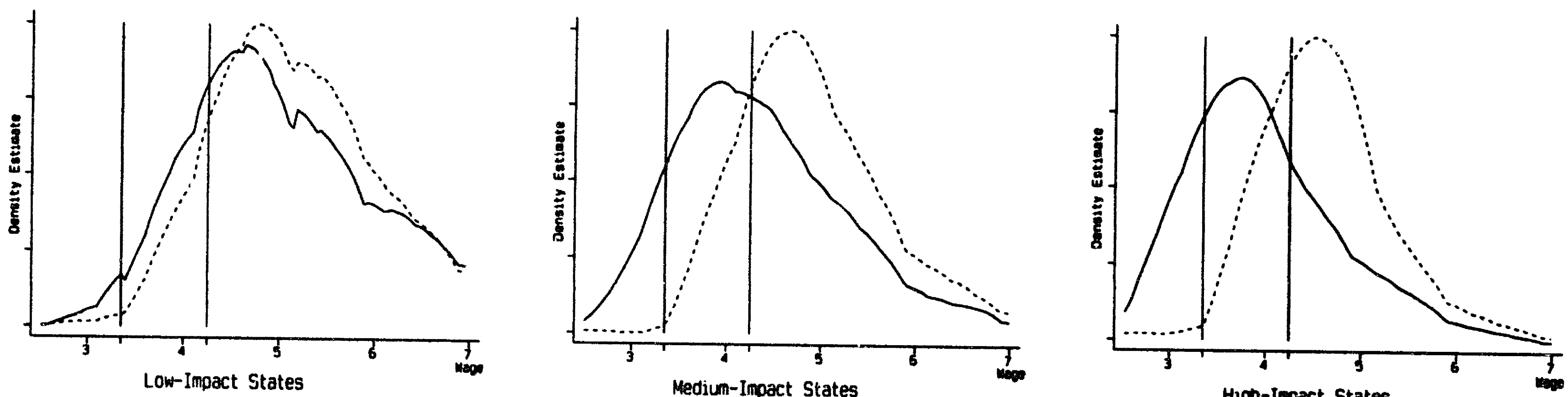

Figure 1-6: These three figures show kernel density estimates of the hourly wage distribution of 16–19 year old workers as reported in the outgoing rotation groups files of the 1989 and 1992 CPSs. *Low-Impact States*, *Medium-Impact States*, and *High Impact States* are those in which the fraction of affected teenagers in 1989 was less than 20 percent, between 20 percent and 40 percent, and over 40 percent. The density estimate for 1989 is shown as a solid line, and the density estimate for 1992 is shown as a dotted line. The units of the horizontal axis are dollars, with vertical lines drawn at the 1989 level of the federal minimum wage [$3.35 per hour] and at the 1992 level of the federal minimum wage [$4.25 per hour].

Table 1.3: Two-Stage Least Squares Estimates for Changes in State Arrest Rates, 1989–1992.

A. Property Crimes

	Burglary			Larceny			Auto Theft			Vandalism		
	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Mean of Dependent Variable in 1992		6.694			19.38			4.312			6.148	
Change in Mean Log Wage	-6.592	-8.737	-13.05	-5.800	-0.1506	-21.86	-3.808	1.563	-7.678	-6.495	-8.215	-12.90
	(2.574)	(5.184)	(3.827)	(5.450)	(13.46)	(8.446)	(2.164)	(3.506)	(3.506)	(2.619)	(4.036)	(3.993)
Change in Adult Arrest Rate	3.247	—	3.265	2.592	—	2.701	5.369	—	6.101	2.778	—	3.180
	(0.4599)		(0.4898)	(0.2557)		(0.2812)	(0.9977)		(1.151)	(0.8486)		(0.9187)

B. Violent Crimes

	Murder			Rape			Robbery			Assault		
	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Mean of Dependent Variable in 1992		0.3785			0.3785			2.999			4.925	
Change in Mean Log Wage	-0.0744	0.0629	-0.0938	0.1589	0.03754	-0.0343	-2.962	-5.553	-6.521	-0.9280	-1.168	-2.573
	(0.1957)	(0.3077)	(0.2691)	(0.1884)	(0.3474)	(0.2734)	(1.165)	(1.927)	(1.806)	(1.298)	(4.062)	(1.838)
Change in Adult Arrest Rate	1.488	—	1.486	1.305	—	1.344	1.965	—	2.105	1.319	—	1.307
	(0.4305)		(0.4310)	(0.0236)		(0.0571)	(0.5458)		(0.5997)	(0.0957)		(0.0978)

Notes: Standard errors are shown under parameter estimates. All specifications include a constant. The dependent variables are changes in arrest rates for burglary, larceny-theft, motor vehicle theft, vandalism, murder, forcible rape, robbery, and assault per 1000 for 16–19 year olds from 1989 to 1992. *Change in Mean Log Wage* is the change in average log wages for 16–19 year olds between 1989 and 1992. *Change in Adult Arrest Rate* is the change in adult arrest rates for the corresponding offense type between 1989 and 1992. Columns labeled OLS were estimated using weighted least squares with weight equal to the number of teenagers in the state in 1989. Columns labeled 2SLS were estimated using two-stage least squares: the fraction of teenagers earning between the state minimum wage and $4.25 per hour in 1989 was used to instrument for change in mean log wages. Weights equal to the 1989 state teenage population were used in computations.

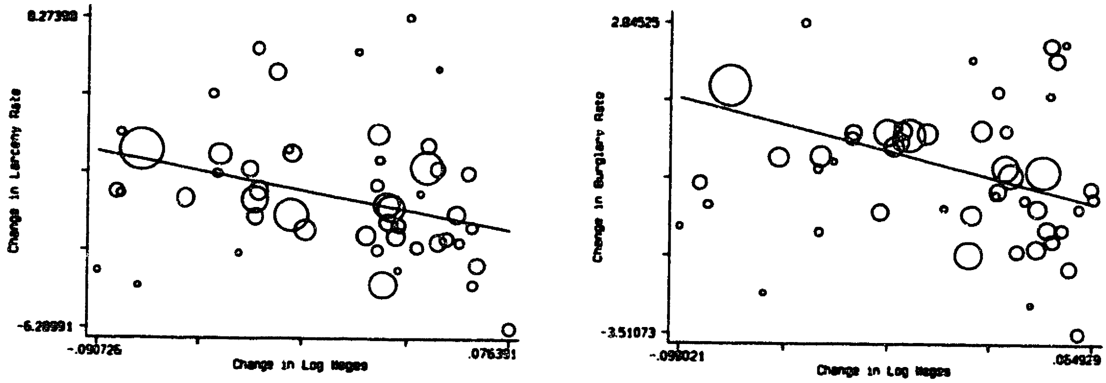

Figure 1-7: These figures show the relationship between the change in the mean log wage of teenagers and the change in arrest rates for two types of property crime: larceny and burglary. The fraction of teenagers affected by the minimum wage in each state is used as an instrument for the change in mean log wage. Larger circles correspond to states with higher populations. Also shown is the two-stage least squares estimate of the regression line. The slope estimates and standard errors are −13.05 and 3.827 for burglary and −21.86 and 8.446 for larceny. Complete regression results are presented in panel A of Table 1.3.

attention to this link: "A prosperous economy and a new generation of young criminals could signal a surge in theft and burglaries ... new theories [predict] that long-term economic growth encourages an eventual rise in property crime." Suppose that states which exhibited the largest wage growth of teenagers between 1989 and 1992 were states in which the economy was growing most quickly. The declines in property crime due to an increase in the return to market work may be diluted by the increased returns to property crime. A booming economy means that targets for theft and robbery are more lucrative.

Estimation results for violent crimes are shown in panel B of Table 1.3. The impact of wages on robbery arrest rates is both negative and significant, while the estimates of β_c for murder, forcible rape, and assault vary in sign and are insignificant.

The instrumental variables estimates of the relationship between wages and arrests were used to construct wage elasticities of arrest rates. The elasticities were calculated using parameter estimates taken from the specifications which control for changes in the adult arrest rate, and were evaluated at the 1992 mean arrest rate. The estimated elasticities are

ulation growth are also states that have the highest crime rates. Norström (1988) finds positive relationship between theft and economic conditions. He concludes that the relationship is a result of the increased availability of desirable objects to steal.

−1.9 for burglary, −1.1 for larceny, −1.8 for motor vehicle theft, −2.1 for vandalism, and −2.2 for robbery.

1.5.2 Cost-Benefit Analysis

Estimates of the relationship between wages and arrests can be used to estimate the social benefit of reducing youth crime by raising the wages of teenagers. For the purposes of these calculations, assume that the government funds a wage subsidy of 10 percent for teenagers in 1992.

Not every criminal is arrested for each crime that he commits, so a count of the reduction in number of arrests will understate the impact of a wage increase on the level of criminal activity.[12] For each category of crime, the ratio of victimizations to total arrests is calculated. The number of victimizations is taken from the National Crime Victimization Survey in 1992; total arrests are as reported in the 1992 Uniform Crime Reports.

Figures published by the Department of Justice (1992) give the ratio of victimizations to arrests in the United States for 1992 as 12.73 for burglary, 9.45 for larceny, 11.44 for motor vehicle theft, and 7.97 for robbery. Statistics on the number of incidents of vandalism in 1992 were not reported by the Department of Justice. A correction to the underestimate of the reduction in crime as measured by arrest rates is obtained by multiplying the decline in arrest rate by the ratio of victimizations to arrests.

Estimates of the social costs of crime were taken from Levitt (1995); these estimates are based partly on the costs of crime produced by Cohen (1988) and Miller, Cohen, and Rossman (1993). The authors provide estimates of both direct monetary loss from crime—losses associated with lost productivity, medical bills, and property loss—and "quality of life" reductions. Crimes which provide a source of income to the offender appear to show the greatest response to wages. As such, attention here is limited to burglary, larceny-theft, motor vehicle theft, and robbery. The relevant estimates of monetary loss are $1,200 for burglary, $200 for larceny, $4,000 for motor vehicle theft, and $2,900 for robbery. Dollar

[12]Of course, not every person arrested is guilty of committing a crime. Of the 1,330,455 juveniles arrested in 1992, in 30.1 percent of the cases the matter was handled within the police department and the suspect was released; 62.5 percent of cases were referred to the juvenile court jurisdiction; and 4.7 percent of cases were transferred to criminal or adult court.

Table 1.4: Estimated Impact of a Teenage Wage Subsidy.

	Decrease in Arrest Rate per 1000	Decrease in Crime Rate per 1000	Monetary Loss per Crime	Expected Social Gain per Subsidized Teenager
Burglary	0.5220	6.904	$1,600	$11.04
Larceny	0.8742	8.263	$200	$1.65
Auto Theft	0.3071	3.512	$4,000	$14.05
Robbery	0.2608	2.083	$17,800	$37.08

Notes: Figures in the first column were calculated as the estimated response in arrest rates per 1000 due to a wage increase of 4 log-points. Estimates were taken from the instrumental variables estimates presented in Table 1.3. The entries of column two were computed by multiplying the reduction in arrests per 1000 by the ratio of crimes known to police to arrests for the corresponding offense category in 1992. Cost estimates appearing in column three were taken from work by Levitt (1995). Gross social gain is the product of the entries in the second and third columns.

estimates of loss due to pain and suffering are $400 for burglary and $14,900 for robbery. Quality of life losses are negligible for larceny and motor vehicle theft.

Estimates of the impact of the policy on the labor market were calculated in the following manner. Let s be the wage subsidy rate, η_{LL} the elasticity of demand for teenage labor, ε the elasticity of labor supply, w^* the equilibrium wage rate, and L^* the equilibrium employment level. Static models of the labor market predict

$$\frac{\partial \log w^*}{\partial s} = \frac{\eta_{LL}}{\varepsilon - (1+s)\eta_{LL}} \tag{1.11}$$

and

$$\frac{\partial \log L^*}{\partial s} = \eta_{LL}\left[1 + \frac{\eta_{LL}}{\varepsilon(1+s)^{-1} - \eta_{LL}}\right]. \tag{1.12}$$

Labor supply and labor demand elasticities are taken to be 0.3 (Killingsworth, 1983) and −0.2 (Hamermesh, 1993). Equation (1.11) predicts a 10 percent wage subsidy will increase equilibrium wages by approximately 4 percent. Equation (1.12) predicts the wage subsidy will increase equilibrium employment by approximately 1.2 percent.

Estimates from Table 1.3 predict that a 4 percent increase in wages will reduce the teenage arrest rate for larceny by 0.5220 per 1000, burglary by 0.8742 per 1000, and motor vehicle theft by 0.3071 per 1000, and robbery by 0.2608 per 1000. These estimates appear in the first column of Table 1.4. Scaling these numbers by the ratio of victimizations to arrests gives the estimates of the reduction in teenage crime rates per 1000 that appear in the second column of Table 1.4. Multiplying the decline in crime rates per 1000 from column two by the cost per crime in column four yields the expected social benefits per subsidized teenager due to the decrease in crime; summing over the rows in column four gives the gross social benefit per subsidized teenager.

We now turn to calculating the costs of the wage subsidy. The average log wage for teenagers in 1992 was 1.562. The wage subsidy is estimated to increase the equilibrium log-wage to 1.602, of which the government pays 10 percent. In dollar terms, this amounts to the government subsidizing wages by an average of $0.47 per hour. Median usual weekly hours for 16–19 year olds in 1992 as reported in the outgoing rotation group files of the CPS was 20 hours per week. Equilibrium hours will increase by approximately 1.2 percent as a result of the wage subsidy. The 1992 March CPS shows that 16–19 year-olds with positive weeks worked in the previous year worked an average of 22 weeks. Using these numbers, the yearly wage cost of subsidizing a single teenager is calculated to be approximately $209.28. Feldstein (1995) estimates the deadweight loss of raising revenue through income taxes to be as much as 30 percent of the revenue raised. If we treat the $209.28 in wage costs as a transfer to teenagers and employers, the social cost of the wage subsidy is the deadweight loss due to the increase in income tax.

Adding the expected social gain for the four offenses yields a social benefit of $63.82 per year for each subsidized teenager. This social gain is offset by the deadweight loss of funding the wage subsidy through an increase in taxes. Using estimates from before, this deadweight loss is calculated to be approximately $62.78. These calculations suggest that the marginal social benefits of reduced crime are roughly equal to the social costs of subsidizing the wages of teenagers.

The estimates presented here are likely a lower bound on the social benefits of reduced crime resulting from an increase in wages. The cost-benefit analysis presented above measures the effect of a wage subsidy on the average teenager. Targeting a wage subsidy to at-risk groups of teenagers may provide larger social benefits for two reasons. On the benefit side, criminal participation among at-risk teenagers is higher, and so the marginal reduction in crime resulting from a wage increase may be greater than the average decrease measured in the population of teenagers. On the cost side, if at-risk teenagers work in economically depressed areas, the dollar amount of a proportional wage subsidy is lower because market wages are lower.

Grogger (1992,1995) documents transitory decreases in employment rates and earnings of arrestees. If arrests or incarceration reduce the pool of employable labor force participants, this shift in the supply of labor may decrease equilibrium employment and production. Also, if lower participation in crime as a teenager decreases recidivism later in life, higher wages may shift the lifetime profile of criminal involvement. Accounting for the social benefits of long-term reductions in crime will raise the present value of the wage subsidization policy.

An interesting comparison can be made of the social benefits of reduced crime resulting from two diverse policies: a wage subsidy for teenagers, and an increase in the number of police officers. Salary costs of an additional police officer are approximately $40,000, with roughly an equal amount incurred in non-salary overhead costs (Levitt, 1995). Levitt finds that an additional police officer reduces the number of reported violent crimes between 3.2 and 7.0 per year, and an additional police officer reduces the number of reported property crimes between 1.6 and 12.4 per year. Using these figures, the yearly social benefit of an additional police officer is roughly $200,000. Virtually all of this social benefit is generated by the reduction in violent crime. There are two reasons for this: violent crime has a more elastic response to an increase in police staffing; and the average social cost of a violent crime is estimated to be approximately $33,000, a figure much larger than the estimated social cost of a property crime of roughly $1,100.

The effects of a wage increase on crime provide a contrast to Levitt's findings that more police officers reduce violent crime, but that police have smaller effects on property crime rates. The strongest effects of a wage increase are measured as decreases in property crime arrest rates, while a wage increase appears to have little effect on violent crime arrest rates. As a crime prevention measure, it is difficult to justify bearing the burden of a wage subsidy because the social benefit of such a program results from a reduction in property crime, which has relatively low costs to society. The analysis does suggest, however, that some combination of an increase in policing and a wage subsidy may be a useful way to target reductions in both violent and property crimes.

1.6 Robustness Checks

1.6.1 State-Specific Trends in Arrest Rates

The estimation strategy presented in the previous sections controls for fixed differences between states by measuring the effect of change in wages on the change in arrest rates. The key assumption of the previous analysis is that the impact of the federal minimum wage change is correlated with the change in teenage wages, but affects youth crime only insofar as it changes the wages of teenagers. The federal minimum change provides an exogenous shift in the wage distribution of teenagers, with lower wage states (those with a higher fraction of 16–19 year olds earning between the state minimum wage and $4.25 per hour in 1989) experiencing a larger movement in its wage distribution. Problems may arise if states with different initial wage distributions experienced differential trends in crime rates before 1989.

Those states experiencing stronger growth in teenage wages between 1989 and 1992 are also those states which initially had a lower wage distribution. Suppose that some states experienced increases in crime during the late 1980s, causing a depression in economic conditions and wages. Teenage wages in these states will be more affected by the federal minimum wage increase, but pre-existing trends in crime could overwhelm the effect of a minimum wage increase. In this case, the instrumental variables estimates would be biased towards finding no effect of wages on crime.

Alternatively, suppose that there are societal factors that contribute both to high crime and low wages. Further suppose that these factors change in a mean-reverting manner: Low-wage, high-crime states tend to get better over time. Low-wage states will have a greater fraction of teenagers affected by the minimum wage legislation, but this measure of minimum wage impact will also be correlated with trends in crime rates. The reduced form could measure a spurious relationship between minimum wage impact and crime that was not caused by higher wages. A negative relationship between wages and crime could exist

Table 1.5: Reduced-Form Estimates for Changes in State Arrest Rates, 1986–1989.

A. Property Crimes

	Burglary		Larceny		Auto Theft		Vandalism	
Mean of Dependent Variable in 1989	6.738		18.46		4.382		4.562	
Fraction of Affected Teenagers	0.0722 (1.318)	1.051 (1.131)	-0.8925 (2.159)	0.5333 (2.022)	-3.216 (0.9074)	-0.5940 (0.7720)	0.0841 (0.8006)	-0.2680 (0.6771)
Change in Adult Arrest Rate	—	2.818 (0.8513)	—	2.515 (0.6416)	—	7.375 (1.388)	—	2.849 (1.321)

B. Violent Crimes

	Murder		Rape		Robbery		Assault	
Mean of Dependent Variable in 1989	0.2704		0.3260		2.321		3.781	
Fraction of Affected Teenagers	-0.4936 (0.4170)	-0.3051 (0.2630)	0.0696 (0.0856)	0.0669 (0.0778)	0.5378 (0.3995)	0.7668 (0.4415)	-1.368 (0.5236)	-1.012 (0.6047)
Change in Adult Arrest Rate	—	6.816 (3.506)	—	1.338 (0.2913)	—	1.416 (1.037)	—	0.9758 (0.2289)

Notes: Standard errors are shown in parentheses under parameter estimates. All specifications include a constant. The dependent variables are changes in arrest rates for burglary, larceny-theft, motor vehicle theft, vandalism, murder, forcible rape, robbery, and assault per 1000 for 16–19 year olds from 1986 to 1989. *Fraction of Affected Teenagers* is the fraction of teenagers earning between the state minimum wage and \$4.25 per hour in 1989. *Change in Adult Arrest Rate* is the change in adult arrest rates for the corresponding offense type between 1986 and 1989.

simply because high-crime, low-wage states get better over time.

Examination of Figure 1-3 suggests that arrest rates followed similar time trends in low-impact, medium-impact, and high-impact states prior to the minimum wage change. To verify this observation in a statistical framework, I regress change in arrest rates from 1986 to 1989 on the fraction of affected teenagers. As before, the change in arrest rates between 1986 and 1989 for four types of property crime–burglary, larceny-theft, motor vehicle theft, and vandalism—and four types of violent crime—murder, forcible rape, robbery, and assault—were constructed. If pre-treatment trends in arrest rates across initial wage distributions do not vary, we expect to see statistically insignificant coefficients on the fraction of affected teenagers. Results of this exercise are presented in Table 1.5.

Inspection of these tables suggests that the correlation between the fraction of affected teenagers in 1989 and change in arrest rates between 1986 and 1989 is weak. Of the sixteen coefficients of *Fraction of Affected Teenagers*, half are positive and half are negative. After controlling for changes in adult arrest rates, none are statistically significant. This analysis of the 1986 to 1989 period shows no evidence of state-specific trends in arrest rates.

1.6.2 Employment Effects of Minimum Wage Increases

One mechanism linking minimum wages to crime is a possible increase in unemployment rates, making it more difficult for criminals to find legal employment. This mechanism may complicate the interpretation of minimum wage effects on crime: the minimum wage increase raises the return to legal work, but it also reduces the probability that teenagers can find work. Empirical evidence examining the disemployment effects of a minimum wage increase has been mixed, but there is an influential literature—Card (1992) and Card and Krueger (1995), for example—that suggests minimum wage increases have little effect on the unemployment rates of teenagers.

To address these issues, I control for changes in the employment rates of teenagers and re-examine the reduced-form effect of a minimum wage increase on arrest rates. These results provide a clearer picture of the direct effect of higher wages on a decrease in teenage crime. However, many of the factors that bias naive evaluations of the relationship between wages and crime also apply to the relationship between employment and crime. For example, arrests and incarceration lower the future employment prospects of criminals; high rates of crime discourage new business formation and encourage existing businesses to change location; and property crime has been shown to rise with economic growth. As a result, directly controlling for changes in the teenage employment to population ratio between 1989 and 1992 introduces a possibly endogenous regressor to the model.

To correct for the possible endogeneity of unemployment rates, I calculate changes in demand for teenage labor using the technique outlined in Katz and Murphy (1992). This strategy assumes that the industry-level shifts in labor demand lead to changes in the employment rates of teenagers, but that the demand shifts are uncorrelated with other determinants of teenage crime. Under these assumptions, the shifts in labor demand provide a source of exogenous variation in the employment rates of teenagers. In what follows, I explain how these demand shifts were calculated, and then examine the first-stage relationship between the employment to population ratios of teenagers and constructed demand shifts.

Finally, the demand shifts are used to estimate the reduced-form relationship between the fraction of affected teenagers and changes in arrest rates.

Constructing Demand Shifts

Data are taken from the March Annual Demographic Supplement to the Current Population Survey. Series for weekly wages, hours worked, weeks worked, and industry by state and age were constructed using the 1989 and 1992 March CPS. The labor force was divided by age into two categories: workers between the ages of 16 and 19, and those between the ages of 20 and 65. Industries were categorized into roughly 50 two-digit industry codes.

Total annual hours of labor demanded can be calculated for each industry-age cell by computing the product of weeks worked and usual weekly hours. This product is weighted by the individual CPS weights for the March supplement and summed over all individuals within the industry-age cells. Deflating this number by the sum of hours over all industry-age cells expresses the labor supplied by a given cell as a fraction of the total hours demanded of all workers. Denote the vector of the average of 1989 and 1992 employment shares by H.

A wage series was constructed for industry-age cell in 1989 and 1992 in the following manner. Average weekly wage was computed within industry-age cells using the individual CPS weights for March supplement. Denote the vector of average weekly wages in 1989 and 1992 by W. A wage index for each year can be constructed by deflating the elements of W by $H'W$. This series provides a measure of the average relative wage of each industry-age cell.

Labor supply of an aggregate group may be calculated by weighting relative wages for each industry-age cell in the group by hours worked in that group. Summing over all cells in the aggregate provides a measure of labor supplied by the aggregate group. Denote by E_{sk} this measure of labor supply for industry s and factor k. To simplify notation, we omit subscripts for state and year. With this convention, s is an index of the 50 two digit industry codes from the CPS and k is an index of the two age categories.

A state-level labor demand shift for teenagers between 1989 and 1992 can be constructed using our data on labor supply. For teenagers, this demand shift was calculated by computing the change in employment for each industry between 1989 and 1992 expressed as a fraction of total industry employment in 1989. The percentage increase in employment for industry s is weighted by the fraction of total teenage labor employed by that sector. Summing over all sectors yields a measure of the change in demand for teenage labor. Formally, the demand shift, ΔX_k, was computed as

$$\Delta X_k = \sum_s \left(\frac{E_{sk}}{E_k}\right)\left(\frac{\Delta E_s}{E_s}\right) \tag{1.13}$$

where E_k is the total employment of factor k, E_s is the total employment in sector s, and ΔE_s is the change in employment in sector s between 1989 and 1992.

Demand Shifts and Teenage Employment

With a measure of the 1989 to 1992 change in demand for teenage labor in hand, the first-stage relationship between demand shifts and teenage employment can be calculated. Teenage employment to population ratios were calculated in each state for 1989 and 1992 using labour force status codes from the monthly files of the outgoing rotation group files of the CPS. The change in teenage employment to population ratios between 1989 and 1992 has a mean of −0.067 with a standard deviation of 0.046. The largest drop in teenage employment was −0.206 in the District of Columbia, while Montana saw the largest gain, 0.060. The constructed demand shifts for teenage labor have a mean of −0.001 with a standard deviation of 0.085. The minimum measured demand shift was −0.195 in Colorado. The largest increase in demand for teenage workers was 0.1793 in North Dakota.

The first-stage regression estimates the relationship between change in teenage employment to population ratios and change in demand for teenage labor. Denote by Δe the change in teenage employment to population ratios. Computed at the state level, the first stage estimates are

$$\Delta e = \underset{(0.006)}{-0.068} + \underset{(0.087)}{0.183}\ \Delta X_k.$$

Standard errors appear in parentheses beneath the estimated parameters. As expected, states experiencing a greater increase in demand for teenage labor saw greater increases in the teenage employment rates between 1989 and 1992. This first-stage analysis suggests that the constructed demand shifts provide a suitable instrument for changes in teenage employment between 1989 and 1992.

Reduced-Form Analysis

Using the change in demand for teenage labor as an instrument for the change in the employment to population ratio of teenagers, the reduced-form relationship between minimum wage impact and arrest rates can be re-examined. As explained earlier, because of the possible disemployment effects of a minimum wage increase, controlling for changes in the employment opportunities available to teenagers presents a cleaner view of the direct effects of a wage increase on crime.

Tables 1.7 and 1.6 display the reduced-form estimates of the effect of fraction of affected teenagers (which serves as an instrument for changes in teenage wages) and change in demand for teenage labor (which serves as an instrument for changes in teenage employment to population ratios) on the change in teenage arrest rates between 1989 and 1992. For each of the eight offense categories, columns one and two duplicate the reduced form estimates of Table 1.2. Columns three and four omit the *Fraction of Affected Teenagers* variable and estimate the reduced form relationship between the change in demand for teenage labor and the change in arrest rates. Columns five and six use both instruments to estimate the effect of the minimum wage law and labor demand on arrest rates.

An examination of columns three and four of Tables 1.7 and 1.6 shows the effect of demand for teenage labor on arrest rates is always insignificant. Columns five and six reveal that the addition of controls for changes in employment opportunities of teenagers has little effect on the reduced form coefficients of *Fraction of Affected Teenagers.* Controlling for changes in demand for teenage labor, estimates of the effect of the minimum wage legislation on arrest rates tend to become slightly more negative. For the five offense categories in which the minimum wage impact had a significant negative effect on the change in arrest rate, addition of the instrument for change in teenage employment does not change the relationship.

This reduced form evidence suggests that employment effects did not play an important role in changing arrest rates between 1989 and 1992. This is consistent with the general

Table 1.6: Reduced-Form Estimates for Changes in State Property Crime Arrest Rates, 1989-1992.

	Burglary						Larceny					
Fraction of Affected Teenagers	-1.773 (1.048)	-2.714 (0.6929)	—	—	-1.710 (1.068)	-2.876 (0.6996)	-0.0306 (2.730)	-4.415 (1.482)	—	—	0.1275 (2.781)	-4.783 (1.506)
Change in Demand for Teenage Labor	—	—	-2.040 (3.176)	1.133 (2.370)	-1.335 (3.157)	2.673 (2.083)	—	—	-3.299 (8.057)	1.921 (4.642)	-3.351 (8.222)	4.350 (4.318)
Change in Adult Arrest Rate	—	3.512 (0.4323)	—	3.278 (0.5002)	—	3.644 (0.4415)	—	2.750 (0.2428)	—	2.567 (0.2572)	—	2.789 (0.2458)
	Motor Vehicle Theft						Vandalism					
Fraction of Affected Teenagers	0.3172 (0.7107)	-1.472 (0.6363)	—	—	0.3214 (0.7152)	-1.517 (0.6505)	-1.667 (0.7957)	-2.681 (0.7337)	—	—	-1.750 (0.8074)	-2.720 (0.7436)
Change in Demand for Teenage Labor	—	—	0.0432 (2.105)	0.0720 (1.725)	-0.0893 (2.144)	0.7048 (1.671)	—	—	1.037 (2.453)	0.2519 (2.332)	1.758 (2.387)	1.040 (2.086)
Change in Adult Arrest Rate	—	5.821 (1.024)	—	4.649 (0.9393)	—	5.858 (1.037)	—	3.396 (0.8301)	—	2.358 (0.8925)	—	3.360 (0.8399)

Notes: Standard errors are shown under parameter estimates. All specifications include a constant. The dependent variables are changes in arrest rates for burglary, larceny-theft, motor vehicle theft, and vandalism per 1000 for 16–19 year olds from 1989 to 1992. *Fraction of Affected Teenagers* is the fraction of teenagers earning between the state minimum wage and $4.25 per hour in 1989. *Change in Demand for Teenage Labor* is a weighted average of changes in industry demand for teenage labor between 1989 and 1992. *Change in Adult Arrest Rate* is the change in adult arrest rates for the corresponding offense type between 1989 and 1992. Weights equal to the 1989 state teenage population were used in computations.

Table 1.7: Reduced-Form Estimates for Changes in State Violent Crime Arrest Rates, 1989-1992.

	Murder						Forcible Rape					
Fraction of Affected Teenagers	0.0128 (0.0622)	-0.0198 (0.0568)	—	—	0.0135 (0.0634)	-0.0283 (0.0580)	0.0762 (0.0712)	-0.0071 (0.0551)	—	—	0.0800 (0.0726)	-0.0026 (0.0560)
Change in Demand for Teenage Labor	—	—	-0.0103 (0.1839)	0.1237 (0.1690)	-0.0159 (0.1876)	0.1391 (0.1734)	—	—	-0.0490 (0.2130)	-0.1000 (0.1576)	-0.0821 (0.2146)	-0.0991 (0.1607)
Change in Adult Arrest Rate	—	1.521 (0.4360)	—	1.568 (0.4392)	—	1.612 (0.4521)	—	1.344 (0.2174)		1.343 (0.2102)	—	1.346 (0.2188)
	Robbery						Assault					
Fraction of Affected Teenagers	-1.126 (0.3512)	-1.330 (0.3019)	—	—	-1.133 (0.3583)	-1.362 (0.3075)	-0.2371 (0.8284)	-0.5277 (0.3647)	—	—	-0.0964 (0.8325)	-0.4979 (0.3717)
Change in Demand for Teenage Labor	—	—	-0.3350 (1.143)	-0.0238 (1.053)	0.1321 (1.059)	0.6004 (0.9023)	—	—	-3.021 (2.412)	-0.8165 (1.106)	-2.982 (2.461)	-0.5943 (1.109)
Change in Adult Arrest Rate	—	2.180 (0.4938)	—	1.847 (0.5826)	—	2.219 (0.5002)	—	1.333 (0.0939)	—	1.315 (0.0963)	—	1.325 (0.0958)

Notes: Standard errors are shown under parameter estimates. All specifications include a constant. The dependent variables are changes in arrest rates for murder, forcible rape, robbery, and assault per 1000 for 16–19 year olds from 1989 to 1992. *Fraction of Affected Teenagers* is the fraction of teenagers earning between the state minimum wage and $4.25 per hour in 1989. *Change in Demand for Teenage Labor* is a weighted average of changes in industry demand for teenage labor between 1989 and 1992. *Change in Adult Arrest Rate* is the change in adult arrest rates for the corresponding offense type between 1989 and 1992. Weights equal to the 1989 state teenage population were used in computations.

findings of previous work that higher levels that higher levels of unemployment are associated with higher levels of crime, but that the relationship is not as strong as one might suspect.[13]

1.7 Conclusions

This chapter measures the responsiveness of youth crime to market wages. The 1990 and 1991 federal minimum wage changes are used as a source of exogenous variation for the change in teenage wages. The analysis is conducted at the state level, using the varying impact of the federal legislation to create a framework in which the differing impact of the law can be exploited in instrumental variables estimation.

Estimates suggest that youth property crime, as measured by arrest rates, is highly responsive to wages. In contrast, Wages do not appear to have a strong impact on arrest rates for violent crimes such as murder, forcible rape, and assault. The negative relationship between wages and property crime combined with the pattern of insignificant relationships between wages and violent crime suggests that the correlation between change in wages and arrest rates is not merely spurious. Comparision of ordinary least squares and two-stage least squares reveals that ordinary least squares consistently underestimates the effect of a wage increase on a decrease in property crime.

Wage elasticities of arrest rates calculated using the two-stage least squares estimates are -1.9 for burglary, -1.1 for larceny, -1.8 for motor vehicle theft, -2.1 for vandalism, and -2.2 for robbery. These estimates suggest that rising teenage wages may be responsible for as much as 30 percent of the decline in arrest rates during the 1990s.

Estimates comparing the yearly wage costs arising from a 10 percent youth wage supplement of $63.82 per subsidized youth, to the social benefit of $62.78 per subsidized youth resulting from the reduction in crime, show roughly a zero net benefit of a wage increase. These figures, however, may be conservative estimates of the effects of such a policy. This work also shows that a decline in youth crime may be a previously unnoticed social benefit of a mandated minimum wage.

[13]Freeman (1999) concludes "... unemployment is related to crime, but if your prior was that the relation was overwhelming, you were wrong."

Chapter 2

The Effect of Criminal Victimization on Employment and Income

2.1 Introduction

Despite recent declines, crime remains a major concern. Law enforcement agencies reported 12.5 million crimes in 1998, corresponding to roughly 4.6 crimes reported per 100 residents. Responses collected in the National Crime Victimization Survey show that police reports significantly understate the level of crime: respondents reported 31 million crimes in 1998, a rate of 11.5 victimizations per 100 residents.

Society devotes significant resources towards preventing crime. In 1997, the criminal justice system had a budget on the order of $100 billion, almost half being spent on police, a third on corrections, and the remainder on the judicial system (Freeman, 1999). Law enforcement agencies in the United States employed 0.64 million sworn officers and 0.25 million civilian employees in 1998. Including civilian employees, the overall law enforcement employment rate was 3.4 employees per 1,000 residents. In addition to these public expenditures, significant private resources were allocated to crime prevention activities: Cunningham et al. (1991) report that private expenditures exceeded governmental expenditures by 73 percent.

The combination of both high levels of criminal activity and high levels of expenditures on crime prevention leads to the question of whether society allocates the optimal level of resources to crime prevention. Analysis of this question requires estimates of the social costs of crime that include both direct monetary losses (lost property or medical bills, for example) and the cost of victims' pain and suffering, as well as indirect costs. Estimates of the first two components of the social cost of crime have been made by Perkins et al. (1996), who use the NCVS to tabulate victims' estimates of the property loss and medical bills associated with crime, and by Cohen (1988) and Miller et al. (1993), who use jury awards to victims of crime to measure the costs of victims' pain and suffering. Less is known about the magnitude of the indirect costs of crime.

In this chapter, I focus on one aspect of the indirect cost of crime: decreases in the earnings of victims of crime.[1] Psychologists have documented a strong relationship between posttraumatic stress disorder (PTSD) and criminal victimization; they have also noted that individuals with PTSD suffer from lower employment rates as a result of the disorder.[2] In addition, behavior undertaken by the victim to guard against repeat victimization may lead to changes in working patterns and, subsequently, a reduction in earnings. Victims, for example, may choose lower-paying jobs in safer neighborhoods, or may withdraw from the labor market entirely if they feel that work exposes them to sufficiently high levels of risk.[3]

The effect of victimization on the employment status of victims is estimated using a longitudinal version of the National Crime Victimization Survey (NCVS) that contains data on the employment outcomes and victimization history of a representative sample of U.S. households. Multiple observations for each individual allow estimation of models that control for both observed and unobserved differences between victims and non-victims of crime. The results suggest that violent crime victimization is associated with a transitory

[1] Another indirect cost of crime has been studied by Dugan (1999). She finds that victims of recent crimes are more likely to move from their current residence.

[2] See Brewin et al. (1999), Schutzwohl and Maercher (1997), Norris and Kaniasty (1991), and Michaels et al. (1998) for evidence on the incidence of PTSD among victims of crime and the consequences of PTSD.

[3] Avoidance behavior may also have other costly consequences for victims. Dugan (1999) finds that recently victimized households are more likely to move.

decrease in employment rates of between 2 and 3 percent, but that the decline lasts no more than 18 months after the victimization. Results for property crimes show little effect of victimization on employment rates.

Focusing solely on employment as a measure of the labor market consequences of crime may undercount employment-related costs of crime. For example, in order to reduce the risk of repeat victimizations, victims may choose to take lower-paying jobs in safer neighborhoods or in locations closer to home. Similarly, victims may reduce the number of hours that they work at their existing job. Neither of these costs are counted by estimates that focus on employment as the outcome of victimization. To address this issue, I limit the sample to heads of households and estimate the effect of victimization on household income. Victims of violent crime suffer a short-lived decrease in household income of between 2 and 3 percent, but there is no income loss associated with property crime victimization.

Estimates of the effect of violent crime on employment and household income can be used to calculate the average earnings loss of crime victims. Using these estimates, violent crime is calculated to cost victims an average of roughly $700 in lost earnings. These estimates of lost earnings are roughly equal to the direct costs of injuries and property loss of victimization (Perkins et al., 1996), suggesting that indirect costs may be an important component of the social cost of crime. On the other hand, estimates of lost earnings are significantly less than the pain and suffering costs estimated by Cohen (1988) and Miller et al. (1993). For violent crime, estimates of pain and suffering costs are more than ten times the combined cost of lost or damaged property, medical bills, and lost earnings. The substantial difference between these estimates may be attributed, in part, to selection bias: violent crimes that lead to civil lawsuits may have higher costs than the average violent crime. Whatever the source, the discrepancy suggests that previous estimates of the pain and suffering costs of crime may overstate the social cost of crime.

The chapter is organized as follows. First, background information on previous work estimating the costs and describing the consequences of crime is presented. The third section outlines a simple framework that decomposes the costs associated with crime. The fourth section describes the longitudinal version of the NCVS that is used in this chapter. Estimates of the effect of victimization on employment are presented in the fifth section.

These estimates show that violent crime victims suffer a transitory drop in employment rates of roughly 2 percent. There appears to be little effect of property crime victimization on employment rates. Using data on household income, the sixth section estimates earnings losses associated with victimization. Violent crime leads to a short-lived decline in household income of between 2 and 3 percent, while property crime is shown to have little effect on household income. The seventh section uses estimates of the effect of violent crime on employment and household income to calculate the average earnings loss suffered by victims of violent crime. These estimates are compared to previous estimates of the monetary and pain and suffering costs of victimization; they are also compared to estimates of the cost of non-fatal job injuries. The conclusion follows.

2.2 The Cost of Crime

A cost-benefit analysis of policies targeted at reducing crime requires an estimate of the social value of a marginal reduction in crime. This figure must include both estimates of the monetary losses associated with victimization (for example, property loss, medical costs, and emergency response services) and, the perhaps larger, non-monetary losses of victimization due to pain, suffering, and a decline in the victim's quality of life. Estimates of the marginal cost of crime prove to be difficult to make; even the average cost of crime, especially average non-monetary loss associated with crime, is difficult to estimate.

The NCVS can be used to measure the direct monetary losses of crime. The survey asks victims to estimate losses from theft or damages to property, medical expenses resulting from victimization, and pay lost as a result of injuries. Perkins et al. (1996) use this source to estimate the average monetary loss associated with crime. They find that the average cost (in 1999 dollars) of a burglary was $1,001, the average cost of a motor vehicle theft was $4,788, the average cost of a robbery was $666, and so on. Using estimates for a range of crimes, they estimate the direct monetary loss to victims of crime to be $638 per crime, which translates to a modest monetary loss of $21.1 billion for all crimes committed in the United States in 1992.

Some attempts have been made to measure the non-monetary costs of crime that result from the loss in quality of life that victims suffer. Cohen (1988) uses jury awards to victims of crime as a measure of the pain, suffering, and fear endured by victims of crime. The author estimates that the average cost of a burglary is $2,310, the average cost of an assault is $17,080, and the average cost of a robbery is $17,883. Incorporating measures of pain and suffering, DiIulio and Piehl (1991) estimate the cost of the average crime to be roughly $2,300.

Interestingly, for most violent crimes, Cohen's estimates show that the cost to victims in direct monetary loss is a small fraction of the total cost of the crime to the victim.[4] The author estimates that the total cost of crime to victims of FBI index crimes is $131.5 billion, compared to the monetary loss of $21.1 billion obtained by Perkins et al. (1996).[5] Miller et al. (1993) revise the estimates of Cohen (1988) to include the costs of psychological problems associated with victimization. The authors conclude that the annual cost of crime due to lost productivity and a reduction in the quality of life is $249.2 billion.

Previous estimates of the monetary losses associated with crime have included in calculations lost earnings as a result of physical injury or time spent dealing with the criminal justice system. For example, Perkins et al. (1996) use responses to the following NCVS questions: "Did you lose time from work because of the injuries you suffered in this incident?" and "Did you lose any (other) time from work because of this incident for such things as cooperating with a police investigation, testifying in court, or repairing or replacing damaged or stolen property?" The authors find that crime costs the victim an average of 3.4 days of work.

However, these estimates may understate the cost of crime. Time lost from work because of injuries suffered during the crime may not include unemployment due to the psychological trauma of victimization. For example, the psychological trauma of victimization may mean that the victim is unable or unwilling to work. Cohen (1988) estimates that approximately 40 percent of rape victims suffer traumatic neurosis and that an additional 10 percent suffer from more severe psychological injuries. Robbery and attempted robbery victims suffer a 2

[4]For example, Cohen's estimates of the monetary loss of robbery is $1,581 while the total social cost is $17,883.

[5]The FBI index crimes are murder, rape, robbery, assault, burglary, larceny, motor vehicle theft, and arson.

percent rate of severe psychological distress.

Previous work by psychologists and psychiatrists has documented a strong relationship between victimization and posttraumatic stress disorder (PTSD). For example, Brewin et al. (1999) find that 20 percent of a group of violent assault victims suffered from PTSD six months after the assault. There is also evidence to suggest that the effects of victimization are not short-lived. Schutzwohl and Maercker (1997) interview a sample of violent crime victims two years after the incident and find that many of the victims still suffered from symptoms of PTSD. Although most of this research has focused on violent crime victims and psychological distress, Norris and Kaniasty (1991) use a representative sample of victims of violent crime, victims of property crime, and non-victims. They find that violent crime had a direct effect on psychological distress, but that property crime had only indirect effects on psychological distress, mediated by victims' concerns about their safety.

As a consequence of psychological distress, victims of crime may be unable to work. For example, Michaels et al. (1998) find that psychological disturbance is associated with low rates of employment five months after injury. Pathe and Mullen (1997) interviewed victims of stalking, finding that 37 percent suffered from PTSD and 53 percent of victims reported changing or ceasing employment as a result of the victimization. The high rate of change in employment may be due to either the psychological trauma of victimization or to a desire of victims to reduce the chances of repeat victimization.

Conventional estimates of the cost of crime may also understate the true cost because they omit the costs to victims of attempts to guard against repeat victimization. For instance, Pathe and Mullen (1997) find in the sample of stalking victims that 39 percent of the sample moved, perhaps as a consequence of the victim's desire to reduce the chances of future incidents. Similarly, Dugan (1999) has used the NCVS to show that victims of crime are more likely to move as a result of recent victimizations. Both results imply that victims may be willing to bear significant costs to avoid crime, and, as such, victims may change their working patterns in response to crime. Employment may expose them to a higher risk of victimization so that they are unwilling to work at certain times, in certain areas, or with certain people.

2.3 Theoretical Framework

It is useful at this point to analyze a framework that formalizes the costs of crime discussed previously: direct monetary cost, losses due to the pain and suffering of victims, lost production due to an inability of the victim to work, and production losses associated with efforts undertaken by victims to guard against repeat victimization.

Consider a two-period economy with a set of identical workers indexed by the unit interval, $[0, 1]$. Suppose that there are an equal number of identical firms, and that firms require a worker in both periods in order to produce an output of 1 at the end of the second period.

In the first period, all workers are involved in production. At the end of the first period, each worker is the victim of a crime with probability p. Victims of crime incur a monetary cost of victimization equal to c and also suffer trauma associated with the victimization. The severity of the crime is represented by a random variable, s, with support $(0, \infty)$. Let F denote the distribution function of s. Assume that some crimes are so traumatic that the victim is unable to work in the second period. Formally, for some constant $\underline{s}$, if $s > \overline{s}$ then the victim cannot participate in second-period production.

Workers with $s \leq \overline{s}$ decide whether to participate in production in the second period. If a worker participates in production in the second period, he earns a fixed wage, $w > 0$. Non-participants earn a wage of 0. In the second period, workers have the following beliefs about the probability of victimization. Non-victims have a subjective probability of victimization equal to 0. Victims who choose to participate in production during the second period have a subjective probability of repeat victimization equal to q. Victims who choose not to participate in production during the second period have a subjective probability of repeat victimization equal to 0.

A key assumption is that workers are more likely to be victims of crime, so that workers who engage in production during the second period face a higher subjective rate of victimization, $q > 0$.[6] For the sake of simplicity, assume that, despite the perceptions of the individuals, the true probability of victimization in the second period is zero. Also, assume

[6]Meithe and Stafford (1987) find that people with relatively low levels of nighttime activity and whose major activity occurred within the home had the lowest risk of violent crime victimization.

that victims believe the severity of a repeat victimization will be equal to the severity of the first crime.

Workers have a utility function, $u(w, s_1, s_2)$, that is a function of the wage that they receive, w; the severity of a first-period victimization, s_1; and the severity of a second-period victimization, s_2. It is assumed that u is increasing in wages, $u_w > 0$, and decreasing in the severity of crimes, $u_{s_1} < 0$ and $u_{s_2} < 0$.

Consider the employment decisions of non-victims: they will always participate in production in the second period since $u(w,0,0) > u(0,0,0)$. Now consider the employment decisions of victims of crime: they will participate if the expected utility from working exceeds the expected utility from non-participation:

$$(1-q)u(w-c,s,0) + qu(w-c,s,s) > u(-c,s,0).$$

Define $U(s) \equiv (1-q)u(w-c,s,0) + qu(w-c,s,s) - u(-c,s,0)$ and assume that $U(0) > 0$, $U' > 0$, and that for some $\underline{s}$, $U(\underline{s}) = 0$. For $s = \underline{s}$, the victim is indifferent between employment and unemployment in the second period:

$$(1-q)u(w,\underline{s},0) + qu(w,\underline{s},\underline{s}) = u(0,\underline{s},0),$$

and for $\underline{s} < s < \overline{s}$ the victim will choose not to work in the second period.

The cost of crime due to the lost utility of workers can be computed as p multiplied by

$$u(w,0,0) - \int_0^{\underline{s}} u(w-c,s,0)\, dF(s) - \int_{\underline{s}}^{\overline{s}} u(-c,s,0)\, dF(s) - \int_{\overline{s}}^{\infty} u(-c,s,0)\, dF(s). \quad (2.1)$$

The first term in equation (2.1) is total worker's utility in an economy without crime, the second term is the utility for victims of crime who choose to work in the second period, the third term is the utility of victims who choose not to work in the second period to reduce the risk of victimization, and the final term is the utility of victims who are unable to work in the second period.

The output loss from crime can be decomposed into the sum of direct monetary losses and indirect production losses:

$$pc + [1 - F(\bar{s})]\,p + [F(\bar{s}) - F(\underline{s})]\,p. \qquad (2.2)$$

The first term in equation (2.2) is the total monetary loss suffered by victims of crime (although much of this monetary loss may be a transfer of property from victims to criminals). The second term in equation (2.2) is the production loss due to victims who are unable to work in the second period. The third term in equation (2.2) is the production loss of workers who choose not to participate in production to reduce the probably of repeat victimization. This chapter attempts to identify the cost of crime corresponding to these last two terms, but does not attempt to separate between the victim incapacitation and avoidance effects.

2.4 Data

The data source is the National Crime Victimization Survey (NCVS), a program initiated by the Department of Justice in 1973 to provide detailed measures of crime.[7] The survey was designed to achieve four primary objectives: to develop detailed information about the victims and consequences of crime, to estimate the numbers and types of crimes not reported to police, to provide uniform estimates of selected types of crimes, and to permit comparisons of crime over time and across regions. Household members in a nationally representative sample of approximately 49,000 households (about 101,000 individuals) are interviewed by personnel from the U.S. Census Bureau. Sampling units are interviewed seven times at six-month intervals, and new units are rotated into the sample on an ongoing basis.

The NCVS collects information on crimes suffered by individuals, regardless of whether those crimes were reported to police. Data on the incidence of both violent and property crimes are recorded. At each interview, members of the household are asked to describe crimes committed against them in the previous six months. In addition to victimization data, the survey provides background characteristics of the respondents: age, race, sex,

[7]Recent versions of this survey are referred to as the National Crime Victimization Survey. Older versions are referred to as the National Crime Survey.

ethnicity, household income, employment status, marital status, education, and geographic characteristics of the sampling unit.

Although the NCVS was designed to estimate cross-sectional crime rates, the rotation group design of the survey allows individuals in the sample to be followed over time. This chapter uses a longitudinal version of the NCVS that tracks a set of sample units through seven six-month periods between July 1986 and December 1990. A limitation of the longitudinal version of the NCVS is that the sampling unit is a physical housing unit, and not a household. Households who move are lost from the sample; no attempt is made to track households to their new residence. If moving decisions are related to victimization, then attrition from the sample on the basis of victimization may bias the estimates of the effect of violent and property crimes on employment contained in this chapter.[8]

The full NCVS longitudinal file contains records for 97,268 individuals. Because I am interested in the labor market outcomes of the individuals in the sample, I restrict the sample to men and women who were between the ages of 18 and 65 for each period that they appear in the sample. This leaves the data set with records for 74,096 individuals. Eliminating individuals with missing data or who appear only for a single period in the sample produces the final data set with records for 44,872 individuals. The average tenure in the sample is 3.3 periods with a standard deviation of 2.4 periods.

Descriptive statistics appear in Table 2.1. The column labeled *Full Sample* uses the entire sample of 44,872 individuals. The column labeled *Non-Victims* uses a sample containing only those individuals who report neither a violent nor property crime victimization for the duration of their time in the sample. Columns labeled *Victims of Violent Crime* and *Victims of Property Crime* use, respectively, a sample containing only those individuals who report a violent crime victimization in at least one of the enumeration periods, and a sample containing only those individuals who report a violent crime victimization in at least one of the enumeration periods.

Comparison of the columns of Table 2.1 highlights important differences between victims and non-victims of crime. Victims of violent crime are less likely to be employed, have lower

[8] Dugan (1999) finds that past victimizations in the vicinity of a household's residence significantly increase the probability that the household relocates.

Table 2.1: Descriptive Statistics.

	Full Sample	Non-Victims	Victims of Violent Crime	Victims of Property Crime
Employed	0.7016	0.6923	0.6880	0.7238
	(0.4576)	(0.4615)	(0.4634)	(0.4471)
Household Income	26590	26260	21067	27772
	(19748)	(19550)	(17746)	(20245)
Age	35.09	35.42	31.12	34.63
	(12.35)	(12.63)	(10.61)	(11.71)
Female	0.5380	0.5410	0.4680	0.5366
	(0.4947)	(0.4983)	(0.4991)	(0.4987)
Black	0.0979	0.0949	0.1056	0.1039
	(0.2971)	(0.2931)	(0.3073)	(0.3051)
Hispanic	0.0703	0.0706	0.0671	0.0694
	(0.2556)	(0.2561)	(0.2502)	(0.2541)
Resident in MSA	0.7732	0.7512	0.8279	0.8196
	(0.4188)	(0.4323)	(0.3776)	(0.3846)
Violent Crime Victim	0.0429	—	—	0.0713
	(0.2023)			(0.2574)
Property Crime Victim	0.2990	—	0.4977	—
	(0.4578)		(0.5001)	
Robbery Victim	0.0071	—	0.1659	0.0111
	(0.0840)		(0.3721)	(0.1048)
Assault Victim	0.0367	—	0.8575	0.0625
	(0.1881)		(0.3496)	(0.2421)
Burglary Victim	0.1579	—	0.3089	0.5281
	(0.3646)		(0.4621)	(0.4992)
Larceny Victim	0.1007	—	0.1825	0.3367
	(0.3009)		(0.3864)	(0.4726)
Auto Theft Victim	0.1051	—	0.1742	0.3516
	(0.3067)		(0.3794)	(0.4775)
N	44872	30489	1923	13417
Enumeration Periods	3.3	3.1	3.2	3.9

Notes: Means with standard deviations in parentheses are shown. Columns labeled *Full Sample*, *Non-Victims*, *Victims of Violent Crime*, and *Victims of Property Crime* present summary statistics for, respectively, the full sample of individuals contained in the NCVS, those individuals reporting no violent or property crime victimizations while in the sample, those individuals reporting at least one violent crime victimization while in the sample, and those individuals reporting at least one property crime victimization while in the sample.

household incomes, are younger, are more likely to be male, are more likely to be black, and are more likely to reside in a city than non-victims of crime. Victims of property crimes are more likely to be employed, have higher household incomes, are more likely to be black, and are more likely to reside in a city than non-victims of crime. These differences suggest that a simple comparison of the post-victimization employment outcomes of victims and non-victims of crime may confound the effect of victimization on employment with differences in pre-victimization characteristics.

2.5 Results for Employment Status

This section estimates the average effect of violent and property crime on the employment status of victims. Although detailed information about crimes is recorded in the NCVS, I classify each crime as simply being either a violent crime or a property crime. An individual is recorded as being a crime victim if it is indicated to the interviewer that he or she was the victim of at least one crime in the previous six months. The subsequent analysis separately estimates the effect on employment of violent crime (rape, robbery, or assault) and property crime (burglary, larceny, or motor vehicle theft). Employment for a given period is measured during the final week of that period. Consequently, most victimizations in a period precede the measuring of employment outcomes in that period.[9]

Section 2.5.1 estimates the effect of victimization on employment status using a statistical model that allows victims to differ from non-victims along observable and unobservable dimensions. This section also estimates the time pattern of changes in employment rates for victims of crime, demonstrating the dynamics and duration of the effect of victimization on employment status. Section 2.5.2 uses a more flexible approach to estimating the effect of victimization on employment. Victims and non-victims are matched on the basis of their characteristics; comparing employment outcomes within the matched groups gives estimates of the employment effects of victimization.

[9]Employment status is measured by the respondent's answer to the question "Did you have a job or work at a business last week? (Do not include volunteer work or work around the house.)"

2.5.1 Employment Dynamics of Victims

This section exploits the longitudinal aspects of the data to estimate the employment effects of victimization. Employment is modeled as a function of an unchanging individual characteristic, a time-varying effect that is constant across all individuals, a set of time-varying observed individual characteristics, and the victimization history of the individual. A simple way to represent an individual's victimization history is with a series of dummy variables representing the time elapsed since a victimization or the time until a victimization will occur. The coefficients for these dummy variables capture pre- and post-victimization differences in the employment rates of victims.

Define the variable $V_{it}^k = 1$ if individual i is the victim of a crime in period $t - k$. The statistical model for the effect of victimization on employment can be written

$$y_{it} = X_{it}\beta + \sum_{k=-m}^{m} \delta_k V_{it}^k + \alpha_i + \gamma_t + \varepsilon_{it}. \tag{2.3}$$

In this model, employment status, y_{it}, depends on an unobserved and unchanging individual effect that is correlated with both employment and victimization, α_i; a set of time-varying characteristics of the individual, X_{it}; and a time effect that is constant across individuals, γ_t. The set of dummy variables, V_{it}^k, for $k = -m, -(m-1), \ldots, m-1, m$, jointly represent the event of a victimization. The parameters δ_k measure the differential employment rates of victims of crime both pre- and post-victimization (controlling for other differences between victims and non-victims of crime).

Using unbalanced panel data from the NCVS, this model is estimated separately for victims of property crime and victims of violent crime, and for values of m between 1 and 4. The set of covariates, X_{it}, includes individual characteristics such as household income and the square of household income, marital status, controls for age, and other demographic characteristics. Because earnings may be affected by crime, all specifications are estimated with and without controls for household income. However, inclusion of these controls does not have a significant impact on the estimates of the effect of victimization.

Estimates of equation (2.3) are shown in Table 2.2 for violent crimes. Figure 2-1 graphs the estimated coefficients $\hat{\delta}_{-3}, \ldots, \hat{\delta}_3$ from specification (6) in Table 2.2 and $\hat{\delta}_{-4}, \ldots, \hat{\delta}_4$ from

Table 2.2: Estimates of the Effect of Violent Crime Victimization on Employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean of Dependent Variable	0.7168	0.7168	0.7168	0.7168	0.7168	0.7168	0.7168	0.7168
δ_{-4}							0.0106	0.0100
							(0.0262)	(0.0261)
δ_{-3}					-0.0085	-0.0080	-0.0025	-0.0022
					(0.0204)	(0.0204)	(0.0227)	(0.0227)
δ_{-2}			0.0031	0.0033	0.0038	0.0041	0.0099	0.0100
			(0.0161)	(0.0161)	(0.0175)	(0.0175)	(0.0202)	(0.0202)
δ_{-1}	-0.0190	-0.0184	-0.0244	-0.0237	-0.0233	-0.0224	-0.0171	-0.0166
	(0.0125)	(0.0125)	(0.0134)	(0.0134)	(0.0149)	(0.0149)	(0.0180)	(0.0180)
δ_0	-0.0309	-0.0305	-0.0371	-0.0366	-0.0356	-0.0350	-0.0295	-0.0292
	(0.0103)	(0.0103)	(0.0113)	(0.0113)	(0.0130)	(0.0130)	(0.0164)	(0.0164)
δ_1	-0.0227	-0.0216	-0.0291	-0.0279	-0.0276	-0.0263	-0.0215	-0.0204
	(0.0111)	(0.0111)	(0.0121)	(0.0121)	(0.0137)	(0.0137)	(0.0169)	(0.0169)
δ_2			-0.0257	-0.0252	-0.0240	-0.0234	-0.0179	-0.0176
			(0.0136)	(0.0136)	(0.0150)	(0.0150)	(0.0179)	(0.0179)
δ_3					0.0105	0.0105	0.0165	0.0162
					(0.0168)	(0.0168)	(0.0194)	(0.0194)
δ_4							0.0123	0.0118
							(0.216)	(0.0216)
Household Income		0.0464		0.0464		0.0464		0.0464
		(0.0059)		(0.0059)		(0.0059)		(0.0059)
Square of Household Income		-0.0384		-0.0384		-0.0383		-0.0383
		(0.0069)		(0.0069)		(0.0069)		(0.0069)
Number of Individuals	29093	29093	29093	29093	29093	29093	29093	29093
Mean Periods in Sample	4.6	4.6	4.6	4.6	4.6	4.6	4.6	4.6

Notes: The dependent variable is a dummy variable for employment status. Standard errors appear in parentheses below estimates. δ_k is the coefficient on the variable V_{it}^k in equation (2.3); $V_{it}^k = 1$ if individual i is the victim of a violent crime in period $t - k$. *Household Income* is household income of respondent divided by 10,000, *Square of Household Income* is the square of household income divided by 100,000. *Number of Individuals* is the number of individuals in the panel, and *Mean Periods in Sample* is the average tenure in the sample. All regressions include individual and time effects, and controls for age and other demographic characteristics.

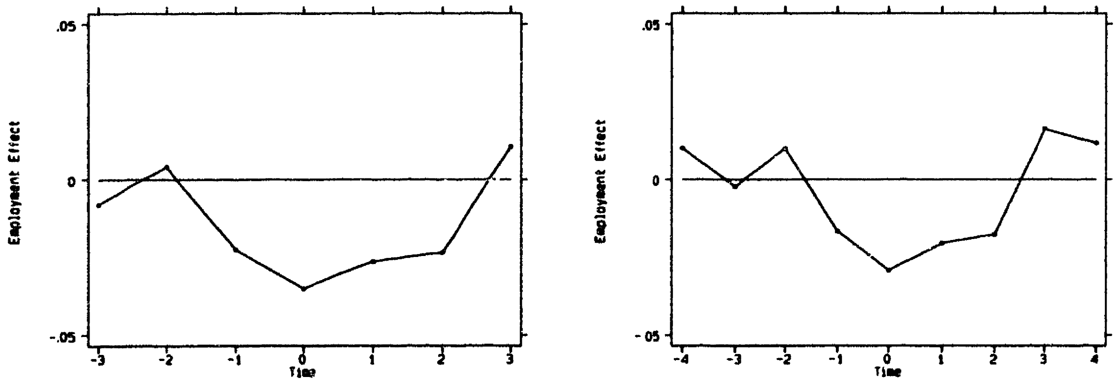

Figure 2-1: The Effect of Violent Crime Victimization on Employment.

specification (8) in Table 2.2. The results reveal that victims of violent crime suffer roughly a 3 percent decline in employment rates in the period of their victimization. These estimates also provide evidence on the persistence of the decrease in employment rates. The estimates in the column (8), for example, suggest that the largest decline in employment rates occurs in the period of the victimization, but that employment rates are also lower up to 18 months after victimization. However, this decrease in employment rate appears to be transitory, and differences in the employment rates of victims and non-victims appear to decrease in magnitude over time. After 18 months have elapsed since the time of victimization, victims return to their pre-victimization employment rates.

It is also interesting to note that in the period immediately preceding the victimization employment rates fall by more than 1 percent. This decline may indicate that the unemployed face a higher risk of victimization, or that there are shifts in the behavior of the victims of crime preceding the victimization that both decrease their employment rates and increase the likelihood of victimization. The pre-victimization decline in employment rates suggests that roughly half of the drop in employment rates associated with victimization may be due to preexisting differences in the behavior of victims, and not as a result of the victimization.

Table 2.3 presents estimates of equation (2.3) for property crime. Figure 2-2 graphs the estimated coefficients $\hat{\delta}_{-3},\ldots,\hat{\delta}_3$ from specification (6) in Table 2.3 and $\hat{\delta}_{-4},\ldots,\hat{\delta}_4$ from specification (8) in Table 2.3. Estimates of the contemporaneous effect of property crime

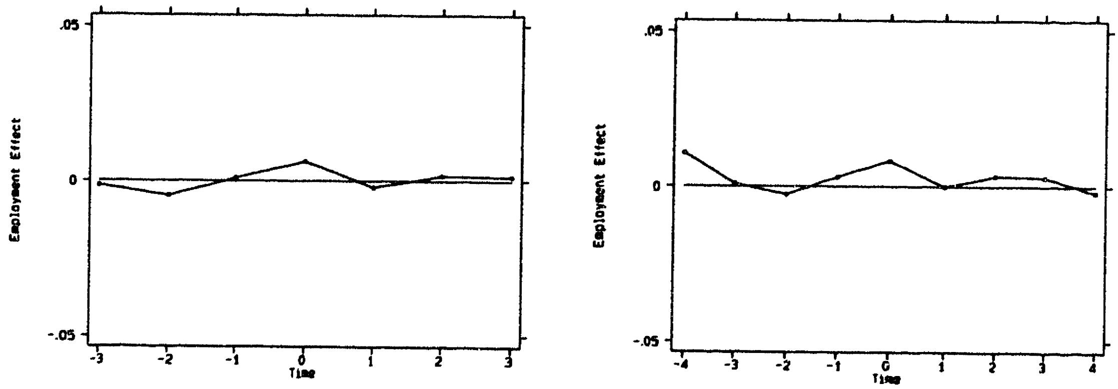

Figure 2-2: The Effect of Property Crime Victimization on Employment.

victimization show that the difference in employment rates of victims and non-victims is generally well under 1 percent in magnitude and is insignificant, suggesting that property crime victimization does not have a strong effect on the employment rates of victims. There also does not appear to be a pre-victimization increase or decrease in the employment rates of victims similar to the dynamics observed for victims of violent crimes.

2.5.2 Matching Estimators

The statistical model of the previous section controlled for observable and unobservable differences between victims and non-victims of crime by introducing individual characteristics and fixed effects in a regression framework. A more flexible alternative estimation strategy is to match victims and non-victims of crime on the basis of their characteristics in an attempt to limit pre-victimization differences. If the matched victims and non-victims are comparable, the difference in the employment rates of the two groups will yield an estimate of the effect of victimization on employment.

Write the average effect of victimization on employment status as

$$\delta_k \equiv \mathrm{E}(y_{i1}|V_i^k = 1) - \mathrm{E}(y_{i0}|V_i^k = 1).$$

In this notation, y_{i1} is the employment status of person i in period t if he is the victim of a crime in period $t-k$, y_{i0} is the employment status of person i in period t in the absence of victimization in period $t-k$, and V_i^k is an indicator for whether person i is the victim

Table 2.3: Estimates of the Effect of Property Crime Victimization on Employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean of Dependent Variable	0.7168	0.7168	0.7168	0.7168	0.7168	0.7168	0.7168	0.7168
δ_{-4}							0.0103	0.0107
							(0.0073)	(0.0073)
δ_{-3}					-0.0016	-0.0014	0.0007	0.0009
					(0.0057)	(0.0057)	(0.0064)	(0.0064)
δ_{-2}			-0.0050	-0.0047	-0.0050	-0.0047	-0.0028	-0.0026
			(0.0046)	(0.0046)	(0.0050)	(0.0050)	(0.0058)	(0.0058)
δ_{-1}	0.0013	0.0015	0.0007	0.0009	0.0008	0.0009	0.0029	0.0030
	(0.0038)	(0.0038)	(0.0040)	(0.0040)	(0.0045)	(0.0045)	(0.0054)	(0.0054)
δ_0	0.0068	0.0067	0.0064	0.0062	0.0064	0.0063	0.0084	0.0083
	(0.0033)	(0.0033)	(0.0036)	(0.0036)	(0.0041)	(0.0041)	(0.0051)	(0.0051)
δ_1	-0.0015	-0.0016	-0.0019	-0.0021	-0.0018	-0.0020	0.0001	-0.0001
	(0.0036)	(0.0036)	(0.0038)	(0.0038)	(0.0043)	(0.0043)	(0.0053)	(0.0053)
δ_2			0.0019	0.0016	0.0020	0.0017	0.0037	0.0034
			(0.0043)	(0.0042)	(0.0047)	(0.0046)	(0.0056)	(0.0056)
δ_3					0.0016	0.0014	0.0033	0.0030
					(0.0052)	(0.0052)	(0.0060)	(0.0060)
δ_4							-0.0016	-0.0019
							(0.0066)	(0.0066)
Household Income		0.0466		0.0465		0.0465		0.0466
		(0.0059)		(0.0059)		(0.0059)		(0.0059)
Square of Household Income		-0.0385		-0.0385		-0.0385		-0.0385
		(0.0069)		(0.0069)		(0.0069)		(0.0069)
Number of Individuals	29093	29093	29093	29093	29093	29093	29093	29093
Mean Periods in Sample	4.6	4.6	4.6	4.6	4.6	4.6	4.6	4.6

Notes: The dependent variable is a dummy variable for employment status. Standard errors appear in parentheses below estimates. δ_k is the coefficient on the variable V_{it}^k in equation (2.3); $V_{it}^k = 1$ if individual i is the victim of a property crime in period $t-k$. *Household Income* is household income of respondent divided by 10,000, *Square of Household Income* is the square of household income divided by 100,000. *Number of Individuals* is the number of individuals in the panel, and *Mean Periods in Sample* is the average tenure in the sample. All regressions include individual and time effects, and controls for age and other demographic characteristics.

of a crime at time $t-k$. To simplify notation, I omit the time index, t, and do not specify whether V_i^k is an indicator for violent or property crime victimization. The parameter of interest, δ_k, is the average difference in the employment rates that crime victims suffer as a consequence of their victimization.

The difficultly in estimating this effect is that the employment rate of victims in the absence of victimization, $\mathrm{E}(y_{i0}|V_i^k=1)$, can not be observed. A simple comparison of the difference in employment rates between victims and non-victims of crime can be decomposed as

$$\mathrm{E}(y_{i1}|V_i^k=1)-\mathrm{E}(y_{i0}|V_i^k=0)=\underbrace{\mathrm{E}(y_{i1}-y_{i0}|V_i^k=1)}_{\delta_k}+\underbrace{\mathrm{E}(y_{0i}|V_i^k=1)-\mathrm{E}(y_{0i}|V_i^k=0)}_{\text{bias}}. \tag{2.4}$$

The first term in equation (2.4) is equal to the average effect of victimization on employment status. The second term is a bias term attributable to the fact that the employment rates of non-victims are not necessary the same as the employment rates that victims of crime would have experienced in the absence of victimization. Suppose, for example, that victims of property crime are more likely to be employed because criminals target wealthy individuals. The bias term in equation (2.4) would then be positive, and the simple comparison would yield an estimated effect larger than δ_k. Similarly, suppose that victims of violent crime have lower employment rates because they live in urban centers with high unemployment and crime rates. In this case, the bias term in equation (2.4) is negative, and the simple comparison yields an estimated effect smaller than δ_k.

The bias term in equation (2.4) is zero if victimization is independent of y_{0i} and y_{1i}. This would be the case, for instance, if the targets of crime were randomly selected by criminals. However, if both employment and victimization are correlated with individual characteristics such as age, race, sex, income, and location, the bias term in equation (2.4) will be non-zero. The effect, δ_k, can be estimated if it is assumed that victimization if independent of employment status conditional on a set of observable characteristics such as geographic and demographic controls.

Under this assumption, the effect of victimization on victims of crime can be constructed as

$$\delta_k = \mathrm{E}(y_{1i} - y_{0i}|V_i^k = 1) = \mathrm{E}\left[\mathrm{E}\left(y_{1i}|X_i, V_i^k = 1\right) - \mathrm{E}\left(y_{0i}|X_i, V_i^k = 0\right)|V_i^k = 1\right]$$

where X_i is a set of characteristics of individual i. This method provides a way of identifying the effect of criminal victimization, but may be impractical if the set of covariates is large, or in situations where X_i is continuous. As a remedy to this problem, Rosenbaum and Rubin (1983, 1984, 1985) show that controlling for the full set of covariates can be replaced by controlling for a particular function of the covariates. In the context of this chapter, define the ***propensity score*** to be the probability of victimization conditional on the set of covariates, $p(X_i) = \mathrm{P}(V_i^k = 1|X_i)$. Using the propensity score, the effect of victimization on the victims of crime can be estimated as

$$\delta_k = \mathrm{E}\left[\mathrm{E}\left(y_{1i}|p(X_i), V_i^k = 1\right) - \mathrm{E}\left(y_{0i}|p(X_i), V_i^k = 0\right)|V_i^k = 1\right].$$

Implementation of this estimator proceeds in two steps. In the first step, the propensity score, the probability of victimization conditional on personal characteristics, is estimated. I parameterize the propensity score function using a probit specification, $p(X_i) = \Phi(X_i\gamma)$. The set of covariates, X_i, includes controls for previous victimizations, lags of employment status and household income, sex, race, ethnic background, marital status, age, and city size. Estimates of the propensity scores are the fitted values from the probit regression of a dummy variable for victimization, V_i^k, on the full set of covariates, X_i.

The second step takes these estimated propensity scores and uses them to estimate $\mathrm{E}(y_i|p(X_i), V_i^k = j)$ for $j = 0, 1$. I use a simple method of matching individuals on their estimated propensity score. The observations are sorted on propensity score and grouped into 10 strata so that the characteristics of victims and non-victims within strata are roughly equivalent. Within each stratum I calculate the difference in employment rates between the victims and non-victims, and weight this by the number of victims in the stratum. Averaging these weighted differences gives an estimate of the average effect of victimization.

An alternative method for estimating the effect of victimization on employment uses a regression-adjusted measure of the difference in employment rates between victims and non-victims within a stratum. Within each stratum, consider the regression

$$y_i = Z_i\beta + V_i^k\delta_k^s + \varepsilon_i$$

where y_i is employment status indicator, Z_i is a vector of covariates that includes controls for age, race, ethnic background, marital status, and city size, and V_i^k is an indicator of victimization status. The estimated effect of victimization within the stratum is $\hat{\delta}_k^s$, which is weighted by the number of victims in the stratum. Averaging the weighted $\hat{\delta}_k$'s provides a regression-adjusted measure of the effect of victimization on employment.

The estimated effect of victimization on employment rates are presented in Table 2.4. Columns labeled *Difference in Means* are an estimate of the unadjusted difference in employment rates between victims and non-victims, $E(y_{i1}|V_i^k = 1) - E(y_{i0}|V_i^k = 0)$. Columns labeled *Propensity Score* are estimates of δ_k computed using the propensity score matching procedure described above. Columns labeled *Regression Adjusted* are estimates of δ_k based on the regression-adjusted propensity score matching technique. Panels A, B, C, D, and E present estimates of, respectively, the effect of victimization at times $t+2$, $t+1$, t, $t-1$, and $t-2$ on employment status at time t.

Examination of these results shows that violent crime reduces the employment rates of victims by a little more than 2 percent up to 18 months after the crime. The estimates are similar to the estimates of Section 2.5.1, although the contemporaneous effect of victimization is somewhat smaller, while the effect of victimization on employment between 6 and 18 months after the victimization is somewhat larger. The regression-adjusted estimates of the effect of violent crime victimization are closer to the estimates in Section 2.5.1 (and smaller than the propensity-score based estimates). However, there are some differences between the regression-controlled estimates and the estimates of Section 2.5.1. The effect of victimization appears to be zero in the second period after victimization (between 12 and 18 months after the crime), and there does not appear to be a pre-victimization decline in the employment rates of victims.

Results for victims of property crimes confirm the findings of Table 2.3. While a simple comparison of the employment rates of victims of property crime with non-victims shows

Table 2.4: Estimates of the Effect of Victimization on Employment, Full Sample

	Violent Crimes			Property Crimes		
	Difference in Means	Propensity Score	Regression Adjusted	Difference in Means	Propensity Score	Regression Adjusted
A. Emp. at Time $t-2$						
Mean Emp. to Pop. Ratio	0.7163	0.7163	0.7163	0.7163	0.7163	0.7163
δ_{-2}	-0.0285	-0.0037	-0.0054	0.0310	-0.0011	0.0014
	(0.0269)	(0.0271)	(0.0203)	(0.0073)	(0.0074)	(0.0058)
N	52710	52710	52710	52710	52710	52710
B. Emp. at Time $t-1$						
Mean Emp. to Pop. Ratio	0.7175	0.7175	0.7175	0.7175	0.7175	0.7175
δ_{-1}	0.0241	-0.0159	-0.0090	0.0311	0.0092	0.0089
	(0.0213)	(0.0208)	(0.0193)	(0.0061)	(0.0059)	(0.0056)
N	74320	74320	74320	74320	74320	74320
C. Emp. at Time t						
Mean Emp. to Pop. Ratio	0.7122	0.7122	0.7122	0.7122	0.7122	0.7122
δ_0	-0.0137	-0.0228	-0.0192	0.0354	0.0092	0.0099
	(0.0161)	(0.0164)	(0.0129)	(0.0049)	(0.0048)	(0.0037)
N	103583	103583	103583	103582	103583	103582
D. Emp. at Time $t+1$						
Mean Emp. to Pop. Ratio	0.7174	0.7174	0.7174	0.7141	0.7141	0.7141
δ_1	0.0106	-0.0212	-0.0162	0.0335	0.0017	0.0022
	(0.0215)	(0.0212)	(0.0158)	(0.0065)	(0.0062)	(0.0049)
N	73879	73879	73879	68430	68430	68430
E. Emp. at Time $t+2$						
Mean Emp. to Pop. Ratio	0.7150	0.7150	0.7150	0.7086	0.7086	0.7086
δ_2	0.0506	-0.0218	-0.0073	0.0357	0.0045	0.0059
	(0.0272)	(0.0255)	(0.0220)	(0.0085)	(0.0079)	(0.0068)
N	52169	52169	52169	68430	68430	68430

Notes: Estimates of the effect of victimization at time t on employment rates. *Mean Emp. to Pop. Ratio* is the average employment to population ratio in the sample. Standard errors appear in parentheses below estimates. Columns labeled *Difference in Means* were estimated by computing the unadjusted difference in mean employment rates between victims and non-victims. Columns labeled *Propensity Score* were estimated by matching individuals by propensity scores. Propensity scores were estimated using victimization history, employment status, household income, race, age, demographic controls, and controls for city size. Columns labeled *Regression Adjusted* were estimated by first matching individuals by propensity score and then regression-controlling for individual characteristics.

that victims have consistently higher employment rates, both the propensity score and regression adjusted estimates show that almost all of this difference is attributable to differences in characteristics of victims and non-victims. For example, the simple comparison shows that victims of property crime are about 3.5 percent more likely to be employed than non-victims, but that this difference is reduced to less than 1 percent after controlling for other observed differences between victims and non-victims.

The previous analysis relied on the assumption that victimization is independent of employment status conditional on a set of observable characteristics, but there may be important unobservable differences between victims and non-victims. As an attempt to control for possible unobservable differences, I limit my sample to individuals who report at least one violent or property crime victimization during their tenure in the sample. This strategy exploits variation in the timing of the victimization to identify the employment impact, but limits cross-sectional differences between individuals in the sample. Using this sample, I re-estimate the effects of victimization on employment in Table 2.4. Results appear in Table 2.5.

Results using the victims sample are similar to those obtained using the full sample: violent crime victimization appears to decrease employment rates by 2 percent, and property crime victimization has little impact on the employment status of victims. The regression-adjusted estimates of the effect of victimization are similar to the propensity score estimates, suggesting that simply matching on propensity score may be more effective in this more homogeneous subsample. Unadjusted differences in employment rates between victims and non-victims now overstates the effect of violent crime victimization on crime. For example, victims of violent crime are about 4 percent less likely to be employed in the period following their victimization, but about half of this difference appears to be attributable to selection. It is interesting to note that, conditional on being the victim of a property crime, a simple comparison of the employment rates of victims and non-victims is close to the propensity score and regression adjusted estimates of the effect of victimization. Whereas this comparison in the full sample yielded a victimization effect of 3.5 percent, this estimate in the full sample is under 1 percent and not statistically significant.

Table 2.5: Estimates of the Effect of Victimization on Employment, Victims Sample

	Violent Crimes			Property Crimes		
	Difference in Means	Propensity Score	Regression Adjusted	Difference in Means	Propensity Score	Regression Adjusted
A. Emp. at Time $t-2$						
Mean Emp. to Pop. Ratio	0.7461	0.7461	0.7461	0.7403	0.7403	0.7403
δ_{-2}	-0.0677	-0.0030	-0.0054	0.0059	0.0001	0.0014
	(0.0280)	(0.0259)	(0.0203)	(0.0077)	(0.0077)	(0.0058)
N	1981	1981	1981	20260	20260	20260
B. Emp. at Time $t-1$						
Mean Emp. to Pop. Ratio	0.7452	0.7452	0.7452	0.7422	0.7422	0.7422
δ_{-1}	-0.0044	-0.0043	0.0126	0.0051	0.0020	-0.0002
	(0.0226)	(0.0228)	(0.0203)	(0.0064)	(0.0064)	(0.0060)
N	2865	2865	2865	28181	28181	28181
C. Emp. at Time t						
Mean Emp. to Pop. Ratio	0.7383	0.7383	0.7383	0.7434	0.7434	0.7434
δ_0	-0.0429	-0.0189	-0.0241	0.0080	0.0063	0.0064
	(0.0174)	(0.0184)	(0.0149)	(0.0053)	(0.0053)	(0.0041)
N	4123	4123	4123	38427	38427	38427
D. Emp. at Time $t+1$						
Mean Emp. to Pop. Ratio	0.7570	0.7570	0.7570	0.7426	0.7426	0.7426
δ_1	-0.0356	-0.0277	-0.0253	0.0032	-0.0036	-0.0021
	(0.0226)	(0.0239)	(0.0182)	(0.0069)	(0.0069)	(0.0055)
N	2424	2424	2424	22291	22291	22291
E. Emp. at Time $t+2$						
Mean Emp. to Pop. Ratio	0.7715	0.7715	0.7715	0.7405	0.7405	0.7405
δ_2	-0.0077	-0.0140	-0.0151	0.0017	-0.0084	-0.0039
	(0.0281)	(0.0292)	(0.0266)	(0.0089)	(0.0089)	(0.0077)
N	1440	1440	1440	12741	12741	12741

Notes: Estimates of the effect of victimization in period t on employment rates, restricting the sample to individuals who report at least one victimization. *Mean Emp. to Pop. Ratio* is the average employment to population ratio in the sample. Standard errors appear in parentheses below estimates. Columns labeled *Difference in Means* were estimated by computing the unadjusted difference in mean employment rates between victims and non-victims. Columns labeled *Propensity Score* were estimated by matching individuals by propensity scores. Propensity scores were estimated using victimization history, employment status, household income, race, age, demographic controls, and controls for city size. Columns labeled *Regression Adjusted* were estimated by first matching individuals by propensity score and then regression-controlling for individual characteristics.

2.6 Results for Household Income

The theoretical framework and empirical results of previous sections focused on the effect of violent and property crime on the employment status of victims. However, limiting attention to employment status may present a distorted view of the employment-related costs of crime. For example, in order to reduce the risk of repeat victimizations, victims may choose to take lower-paying jobs in safer neighborhoods or in locations closer to home. Similarly, victims may reduce the number of hours that they work at their existing job. Because these decreases in earnings are not counted by estimates of the cost of crime that use employment rates as the outcome, this section attempts to estimate the effect of violent and property crime on the income of victims.

The NCVS provides information on the annual household income of the respondents at each interview period. Household income is a categorical variable, recorded as belonging to one of 14 intervals. For each household, I code household income as being the midpoint of the reported interval; income is top-coded at $75,000. In the subsequent analysis, the outcome will be the logarithm of this measure of household income. Because variation in income is measured only at the household level, I limit my sample to heads of households.

2.6.1 Income Dynamics of Victims

The statistical model for the effect of victimization on household income is

$$y_{it} = X_{it}\beta + \sum_{k=-m}^{m} \delta_k V_{it}^k + \alpha_i + \gamma_t + \varepsilon_{it} \tag{2.5}$$

where the dependent variable, y_{it}, is the logarithm of household income, X_{it} is a set of time-varying individual characteristics, α_i is an unobserved individual effect, and γ_t is a time effect that is constant across all individuals. As in Section 2.5.1, the event of a victimization is represented as a series of dummy variables, V_{it}^k, for $k = -m, -(m-1), \ldots, (m-1), m$ where $V_{it}^k = 1$ if individual i in period $t-k$ was the victim of a crime.

Equation (2.5) is estimated separately for violent and property crimes and for m between 1 and 4 using the unbalanced panel of heads of households from the NCVS. Results are shown in Table 2.6, and Figure 2-3 graphs the coefficients from specifications (3) and (4) in

Table 2.6: Estimates of the Effect of Victimization on Income

	Violent Crime				Property Crime			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
δ_{-4}				0.0127				-0.0181
				(0.0212)				(0.0062)
δ_{-3}			-0.0080	0.0002			-0.0056	-0.0080
			(0.0164)	(0.0183)			(0.0048)	(0.0054)
δ_{-2}		-0.0084	-0.0099	-0.0017		-0.0055	-0.0054	-0.0077
		(0.0129)	(0.0140)	(0.0162)		(0.0039)	(0.0042)	(0.0049)
δ_{-1}	-0.0161	-0.0205	-0.0220	-0.0143	-0.0060	-0.0055	-0.0053	-0.0073
	(0.0100)	(0.0107)	(0.0120)	(0.0145)	(0.0032)	(0.0034)	(0.0038)	(0.0046)
δ_0	-0.0135	-0.0180	-0.0192	-0.0119	0.0010	0.0018	0.0021	0.0004
	(0.0082)	(0.0090)	(0.0104)	(0.0131)	(0.0028)	(0.0031)	(0.0035)	(0.0043)
δ_1	-0.0324	-0.0369	-0.0380	-0.0309	0.0020	0.0028	0.0033	0.0018
	(0.0088)	(0.0096)	(0.0109)	(0.0135)	(0.0030)	(0.0032)	(0.0036)	(0.0044)
δ_2		-0.0122	-0.0133	-0.0066		0.0073	0.0079	0.0067
		(0.0108)	(0.0119)	(0.0142)		(0.0036)	(0.0039)	(0.0047)
δ_3			0.0006	0.0070			0.0058	0.0048
			(0.0133)	(0.0153)			(0.0044)	(0.0050)
δ_4				0.0055				0.0088
				(0.0170)				(0.0056)
Number of Individuals	19004	19004	19004	19004	19004	19004	19004	19004
Mean Periods in Sample	4.6	4.6	4.6	4.6	4.6	4.6	4.6	4.6

Notes: The dependent variable is the logarithm of household income. Standard errors appear in parentheses below estimates. δ_k is the coefficient on the variable V_{it}^k in equation (2.5); $V_{it}^k = 1$ if individual i is the victim of a property crime in period $t-k$. *Number of Individuals* is the number of individuals in the panel, and *Mean Periods in Sample* is the average tenure in the sample. All regressions include individual and time effects, and controls for age and other demographic characteristics.

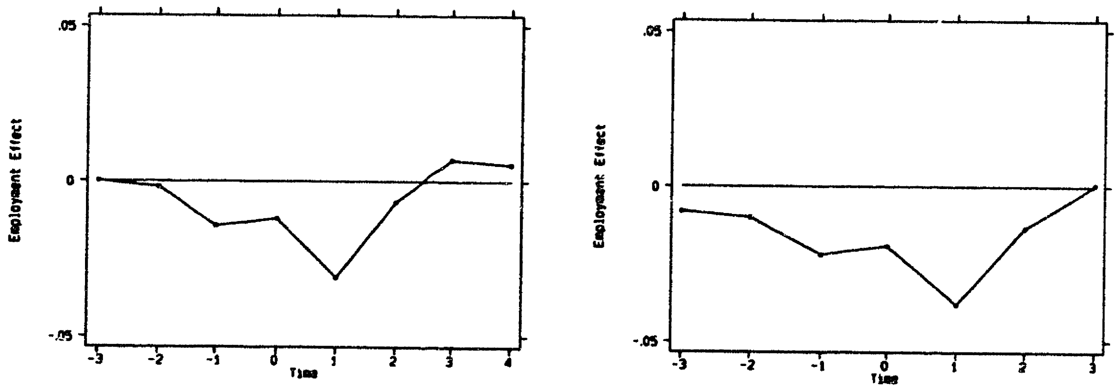

Figure 2-3: The Effect of Violent Crime Victimization on Household Income.

Table 2.6. For victims of violent crime, there are a number of interesting contrasts between the results for household income and the results using employment status as an outcome. Recall that victims of violent crime had lower employment rates in the six months before victimization and that in the six months after the victimization their employment rates declined further. The pattern of a pre-victimization decline in household income exists, but the magnitude of the decline is nearly equal to the decline in income in the six months following the crime. The largest drop in household income (on the order of 3 percent) appears to be between six months and a year after the victimization. Household income also appears to rebound to pre-victimization levels more quickly than employment status; the difference in household income between a year and 18 months after victimization is on the order of 1 percent and is smaller than the pre-victimization decline in household income.

Similar to the results for the effect of property crime victimization on employment status in Section 2.5.1, results for victims of property crimes reveal little effect of victimization on household income. The estimates of δ_k from equation (2.5) are generally well under 1 percent in magnitude and not statistically different from zero. Figure 2-4 graphs the coefficients from specifications (7) and (8) in Table 2.6.

2.6.2 Matching Estimates

Matching estimates are presented in Table 2.7. Analogous to Section 2.5.2, the identifying assumption is that victimization is independent of income conditional on a set of

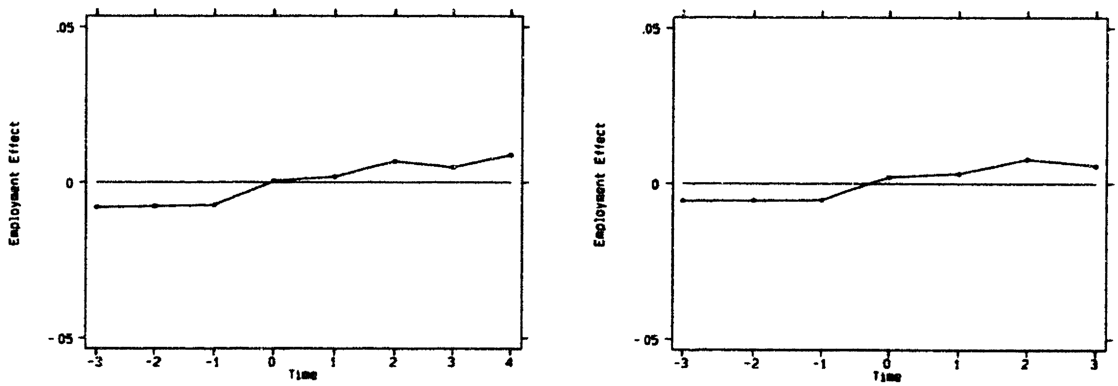

Figure 2-4: The Effect of Property Crime Victimization on Household Income.

observable characteristics. The probability of victimization is estimated as a function of victimization history, employment status, household income, city size, age, race, sex, and other demographic characteristics. Observations are sorted into 10 strata on the basis of this estimated probability of victimization so that pre-victimization characteristics within a strata are roughly equivalent. A comparison of the difference in the mean logarithm of household income between victims and non-victims within a strata yields an estimate of the effect of crime on household income.

Columns labeled *Difference in Means* in Table 2.7 are an unadjusted comparison of the difference in average log income between victims and non-victims of crime. These estimates reveal large differences in the household incomes of victims of violent crime: in the period of victimization, they earn almost 33 percent less than non-victims. Columns labeled *Propensity Score* present estimates of the effect of victimization on household income obtained using propensity score matching. Columns labeled *Regression Adjusted* first match individuals on the basis of their propensity score, and then compute the regression-adjusted difference in household income between victims and non-victims. Regression controls include lagged values of household income and employment status, city size, age, race, ethnic background, and other demographic controls. Comparison of these last two columns to the first column reveals that a simple comparison of the household incomes of victims and non-victims significantly overstates the effect of victimization: much of the difference in incomes is a result of the differences in pre-existing characteristics of victims and non-victims.

Table 2.7: Estimates of the Effect of Victimization on Household Income

	Violent Crimes			Property Crimes		
	Difference in Means	Propensity Score	Regression Adjusted	Difference in Means	Propensity Score	Regression Adjusted
C. Emp. at Time $t-2$						
δ_{-2}	-0.2591	0.0003	-0.0021	0.0512	-0.0088	0.0053
	(0.0563)	(0.0058)	(0.0356)	(0.0162)	(0.0164)	(0.0127)
N	34653	34653	34653	34653	34653	34653
D. Emp. at Time $t-1$						
δ_{-1}	-0.2521	-0.0079	0.0076	0.0262	-0.0081	-0.0107
	(0.0457)	(0.0046)	(0.031)	(0.0136)	(0.0139)	(0.0073)
N	48850	48850	48850	48850	48850	48850
C. Emp. at Time t						
δ_0	-0.3321	-0.0099	-0.0239	-0.0069	-0.0039	0.0046
	(0.0351)	(0.0343)	(0.0095)	(0.0114)	(0.0118)	(0.0032)
N	67973	67973	67973	67973	67973	67973
D. Emp. at Time $t+1$						
δ_1	-0.2046	-0.0298	-0.0343	0.0569	0.0103	0.0173
	(0.0415)	(0.0465)	(0.0129)	(0.0142)	(0.0142)	(0.0047)
N	48526	48526	48526	44897	44897	44897
E. Emp. at Time $t+2$						
δ_2	-0.0903	-0.0281	-0.0224	0.0734	0.0169	0.0217
	(0.0558)	(0.0587)	(0.0164)	(0.0178)	(0.0174)	(0.0074)
N	34257	34257	34257	29636	29636	29636

Notes: Estimates of the effect of victimization at time t on household income. The dependent variable is the logarithm of household income. Standard errors appear in parentheses below estimates. Columns labeled *Difference in Means* were estimated by computing the unadjusted difference in mean employment rates between victims and non-victims. Columns labeled *Propensity Score* were estimated by matching individuals by propensity scores. Propensity scores were estimated using victimization history, employment status, household income, race, age, demographic controls, and controls for city size. Columns labeled *Regression Adjusted* were estimated by first matching individuals by propensity score and then regression-controlling for individual characteristics.

Estimates for violent crimes show that the household incomes of victims falls by roughly 2 to 3 percent in the 18 months following victimization. In contrast to the estimates of Table 2.6, the matching estimates do not show a significant difference in the pre-victimization income of victims and non-victims. Compared to the estimates of Table 2.6, matching estimates also suggest that victimization has a significant effect on household income between a year and 18 months after the crime. Estimates predict that victims of violent crime have household incomes that are roughly 2 percent lower than non-victims over this time period. However, in accordance with the previous estimates, the largest decline in income—a drop of roughly 3 percent—occurs between six months and a year after victimization.

Matching estimates for the effect of property crime on household income are similar to the results presented in Section 2.6.1. Victimization is not associated with significant differences in household income in the months following the crime, nor with differences in household income in the months preceding the crime.

2.7 Estimates of Earnings Losses

The estimates of the effect of violent-crime victimization on employment and household earnings can be used to revise previous estimates of the social costs of crime. This section presents calculations that use estimates from Section 2.5 and Section 2.6 to compute the average earnings loss suffered by victims of violent crime. These figures are compared to previous estimates of the social cost of crime in order to measure the relative importance of the employment-related costs of crime. Finally, to put the employment-related costs of victimization in context, I compare these costs to the cost of lost work and pain and suffering resulting from job injuries.

The results of Section 2.5 show that victims of violent crime suffer a 2 to 3 percent decrease in employment rates in the 18 months following the victimization. A rough estimate of the average employment-related loss of victimization can be calculated by multiplying the average decrease in victims' employment rates by the average earnings of victims, then summing these costs over all periods in which the victim is affected. Specification (8) of Table 2.2 estimates that the average decrease in employment rates caused by violent-crime victimization is approximately -0.0292 in the six months after the incident, -0.0204 between six months and one year after the incident, and -0.0176 between one year and 18 months after the incident. To simplify calculations, the effects in all other periods are assumed to be zero, and the effect of victimization on employment is assumed to be constant over each six-month period.

The next step is to calculate the average earnings of victims of violent crime. Because the NCVS provides information only on the household income of respondents, I use data from the March supplement of the 1989 Current Population Survey (CPS) to measure earnings. Restricting the sample to workers between the ages of 18 and 65 gives an estimate of average earnings for this group of $16,167. Table 2.1, however, shows that victims of violent crime have lower household income than a representative sample of households; the figure of $16,167 may be an overestimate of the average earnings of victims of violent crime. Nevertheless, because victimization data and earnings are not available from the same data source, I take the earnings figures from the March CPS as the estimate of the annual earnings of victims of violent crime. Inflating $16,167 using the Consumer Price Index (CPI) yields an estimate of annual earnings of $22,000 in 1999 dollars. Multiplying the average decrease in victims' employment rates by the average earnings (for the sake of simplicity, income is not discounted over the time periods) produces a figure of roughly $740 as the expected cost of violent-crime victimization due to the reduction in the employment rates of victims.

Results from Section 2.6 provide an alternate way of calculating the dollar value of the employment-related costs of victimization. Estimates show that victims of violent crimes suffer a short-lived decline in household income of roughly 3 percent. An estimate of the expected employment-related loss of victimization can be calculated by multiplying the average decrease in victims' household income by the average household income of victims, then summing these costs over all periods in which the victim is affected. Estimates from column (4) in Table 2.6 show that the decrease in household income caused by violent-crime victimization is approximately -0.0119 in the six months after the incident, -0.0309 between six months and one year after the incident, and -0.0066 between one year and 18 months after the incident. As before, to simplify the calculations, the effects in all other periods are assumed to be zero, and the effect of victimization on household income is assumed to be constant over each six-month period.

As shown in Table 2.1, the average household income of victims of violent crime is $21,000. Inflating this figure by the CPI gives an estimate of the household income of victims of violent crime of $28,700 in 1999 dollars. Using the estimates of the effect of victimization on household income, the decrease in the logarithm of household income has a dollar value of approximately $710. This estimate of the earnings losses of victims of violent crime is close to the figure of $740 dollars that was computed using estimates of the effect of violent crime victimization on employment.

To put these findings in context, estimates of the earnings losses of victims can be compared with the cost of lost or damaged property and medical bills associated with violent crime. Estimates of the monetary cost of victimization from Perkins et al. (1996) are similar in magnitude to the estimates of earnings losses calculated above, suggesting that indirect costs may be an important component of the social cost of crime. However, the indirect costs of crime seem relatively modest when compared to the estimates of Cohen (1988) and Miller et al. (1993). For violent crimes such as robbery, estimates of the costs of pain and suffering are roughly $17,000; this figure is more than ten times the combined cost of lost or damaged property, medical bills, and lost earnings resulting from the crime. The substantial difference between these estimates suggests that previous estimates of the pain and sufferings costs of victimization may overstate the true cost of crime. This difference may be attributed, in part, to the methodology employed by previous studies. For example, Cohen (1988) uses jury awards to victims of violent crime as a measure of the cost of crime, but violent crimes that lead to civil lawsuits may not be representative of all violent crimes. In particular, such crimes may have a higher cost than the average violent crime.

An interesting comparison can also be made between the costs of violent crime and job injuries. Viscusi (1979, 1981) has used variation in the risk of job injuries and compensating wage differentials to measure the implicit cost of job-related hazards. Using data from the Survey of Working Conditions, Viscusi (1979) finds that the average cost of nonfatal job injuries and illnesses is between $24,200 and $42,400. Using data from the Panel Study of Income Dynamics, Viscusi (1981) finds slightly higher numbers; the implicit cost of nonfatal job injuries and illnesses is between $54,500 and $60,500. Viscusi and Moore (1987) use similar methodology to estimate the implicit value of non-monetary cost of job injuries. The authors find that the pain and suffering and non-work disability costs of job injuries are between $24,900 and $38,000. Interestingly, the pain and suffering costs of job injuries account for roughly half of the total cost of job injuries.

Surprisingly, the non-monetary costs of workplace injuries ($25–38 thousand) are slightly higher than estimates of the pain and suffering costs of violent crimes such as robbery and assault ($17–18 thousand). Combining the estimates of medical and property losses of violent crimes obtained by Perkins et al. (1996) with the employment-related losses calculate above, the monetary losses of violent crime are approximately $1,500. This figure is well below the estimates of Viscusi (1979, 1981) that place the cost of medical care and lost work of nonfatal job injuries of between $20,000 and $30,000.

2.8 Conclusions

This chapter measures the effect of violent and property crimes on the employment status and household income of victims. Longitudinal data from the National Crime Victimization Survey is used to estimate models that control for both observed and unobserved differences between victims and non-victims of crime. Estimates obtained using panel data models and matching techniques yield similar results: violent crime appears to be associated with a transitory decrease in the employment rates and household income of victims, while property crime has little effect on the employment rates and household income of victims.

I estimate that a violent crime victimization in the previous six months reduces employment rates by roughly 2 to 3 percent. The reduction in unemployment rates appears to be transitory: crimes that occurred more than eighteen months ago do not decrease the employment rates of victims. Victims of violent crime also appear have lower employment rates in the period immediately prior to victimization. This difference may reflect the fact that unemployment induces lifestyle changes that put individuals at a higher risk of victimization. Property crimes do not appear to reduce the employment rates of victims.

Victimization may also be associated with a reduction in earnings that is not the result of an individual leaving the labor force. Estimating the effect of crime on household income reveals that violent crimes lead to a drop in household income of 2 to 3 percent. However, the duration of this effect appears somewhat shorter than the effect of violent crime victimization on employment status.

The estimates of the effect of violent crime on employment and household income are used to calculate the average employment-related loss of victimization. Violent crime costs victims an average of roughly $700 in lost earnings. This amount is approximately equal to estimates of the costs of injuries and property loss of violent crime victimization, suggesting that earnings losses may be important in calculating the social costs of crime. Estimates of the earnings loss suffered by victims of violent crime can also compared to previous estimates of the pain and suffering costs of crime (Cohen, 1988, and Miller et al., 1993); pain and suffering costs of crime are significantly larger than the combined cost of lost or damaged property, medical bills, and lost earnings. The substantial difference between these estimates suggests that previous estimates of the pain and suffering costs of crime may overstate the true cost.

Chapter 3

Efficient Bootstrapping for GMM*

3.1 Introduction

There are many important applications of generalized method of moments (GMM) estimators for cross-section and panel data. For example, there are a wide variety of GMM estimators for dynamic panel models.[1] Also, instrumental variables estimators, which are also GMM estimators, are important in the estimation of treatment effects. It is well known that the usual asymptotic theory can be a poor approximation to the distribution of the estimators, particularly when there are many overidentifying restrictions or when the parameters of interest are not well identified. The bootstrap provides one approach to improvements in the approximation; this chapter describes a relatively efficient bootstrap method for GMM in cross-section and panel data. We show that our method improves on the standard asymptotic approximation under certain regularity conditions. We also illustrate that the improvement can be large, particularly in dynamic panel data models, using Monte Carlo simulations and an empirical example.

Hall and Horowitz (1996) have previously proposed a bootstrap for GMM. Their approach is based on centering the moment conditions in GMM, while ours is based on bootstrapping the original moment conditions with an efficient estimator of the distribution.[2]

*This chapter is joint work with Bryan W. Brown and Whitney K. Newey.

[1]For example, see Anderson and Hsaio (1982), Blundell (1992), Holtz-Eakin, Newey, and Rosen (1988), and Nickell (1981)

[2]Our original work, including an argument for bootstrap improvements, was carried out simultaneously

Our approach has a computational advantage in that the bootstrap does not require modifying the form of the estimator. Also, our approach is asymptotically efficient relative to theirs. On the other hand, their approach has wider applicability than ours because they cover dependent data and we do not.

The bootstrap uses an estimate of the distribution of the data to form an estimate of the distribution of a statistic. Under certain conditions, the improved approximation to the distribution of the statistic can be expressed in a form similar to an Edgeworth expansion. These expansions are based on large sample approximations that are of higher order in the sample size than the usual asymptotic approximation. There are also other approaches to improvements in approximation, including those of Bekker (1994) and Staiger and Stock (1997). Although to date there have been few comparisons of these different approaches, the bootstrap appears to work well in some examples where the parameters are well identified and there are many overidentifying restrictions.

The organization of the chapter is as follows. Section 3.2 presents background on the GMM estimation and inference problem, and discusses the problems with the usual empirical distribution bootstrap. Section 3.3 presents our approach to bootstrapping from moment restricted distributions. Section 3.4 presents regularity conditions and formal results to show that our bootstrap improves on the standard asymptotic approximation. Section 3.5 presents heuristic arguments comparing the efficiency of our bootstrap to that of Hall and Horowitz (1996). Results of some Monte Carlo experiments are reported in Section 3.6, and Section 3.7 gives an empirical example. The conclusion follows.

3.2 Inference for GMM

Asymptotic Inference for GMM

We first describe GMM estimation and then turn to the inference problem for GMM. Let $Z_1, \ldots, Z_n$ denote the data, which is assumed throughout the chapter to independent and identically distributed. Also, let β denote a $q \times 1$ vector of parameters, Z a single data

and independent of Hall and Horowitz (1996) as in Brown and Newey (1992). The current version of this work relies heavily on their theoretical results, however.

observation, and $g(z,\beta)$ a $r \times 1$ vector of functions where $r \geq q$. Suppose that the true parameter β_0 satisfies the population moment conditions

$$\mathrm{E}[g(Z_i,\beta_0)] = 0. \tag{3.1}$$

A GMM estimator of β_0 is obtained by choosing β so that the sample moments of the function $g(z,\beta)$ are close to zero. Let $\hat{g}_n(\beta) = \sum_{i=1}^n g(z_i,\beta)/n$ be the vector of sample moments and Ω a positive semidefinite weighting matrix. The quadratic form $\hat{g}_n(\beta)'\Omega\hat{g}_n(\beta)$ is a measure of how close the sample moments are to zero, and a GMM estimator can be obtained as

$$\tilde{\beta} = \underset{\beta \in B}{\operatorname{argmin}}\, \hat{g}_n(\beta)'\Omega\hat{g}_n(\beta).$$

A widely used GMM estimator is one where the weighting matrix Ω is optimal, in the sense that it minimizes the asymptotic variance of the GMM estimator (Hansen, 1982). An optimal weighting matrix is one with a probability limit that is the inverse of the asymptotic variance of $\sqrt{n}\hat{g}_n(\beta_0)$. Because the data are i.i.d., this asymptotic variance will be $\mathrm{E}[g(Z,\beta_0)g(Z,\beta_0)']$, which is assumed to be non-singular. An estimator of the inverse of the asymptotic variance can be obtained as $\Omega(\tilde{\beta}) = \left[\sum_{i=1}^n g(z_i,\tilde{\beta})g(z_i,\tilde{\beta})'/n\right]^{-1}$. The estimator corresponding to this weighting matrix will be

$$\hat{\beta} = \underset{\beta \in B}{\operatorname{argmin}}\, \hat{g}_n(\beta)\Omega(\tilde{\beta})\hat{g}_n(\beta). \tag{3.2}$$

In this chapter we will focus on inference for such an optimal GMM estimator because confidence intervals and tests of overidentification are simplest for it.

Asymptotic confidence intervals for the components of β are based on the asymptotic normality of $\hat{\beta}$ and consistent estimation of the asymptotic variance. A consistent estimator, $\hat{V}$, of the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta_0)$ can be formed as $\hat{V} = \left(\hat{G}'\Omega(\tilde{\beta})\hat{G}\right)^{-1}$, using the Jacobian of the moments, $\hat{G} = \partial\hat{g}_n(\hat{\beta})/\partial\beta$, and $\Omega(\tilde{\beta})$. Thus, a consistent estimate of the standard error for the jth component of $\hat{\beta}$ is $\mathrm{SE}_j = \sqrt{\hat{V}_{jj}}$. Asymptotic normality of $\sqrt{n}(\hat{\beta}-\beta_0)$ and consistency of SE_j imply that the t-ratio, $t_j = \sqrt{n}(\hat{\beta}_j - \beta_{0j})/\mathrm{SE}_j$, has a limiting distribution that is standard normal. For the $1-\alpha$ quantile, z_α, of the distribution of the absolute value of a standard normal random variable, a $1-\alpha$ confidence interval can be constructed as

$$\mathcal{I} = \left[\hat{\beta}_j - z_\alpha \frac{\mathrm{SE}_j}{\sqrt{n}}, \hat{\beta}_j + z_\alpha \frac{\mathrm{SE}_j}{\sqrt{n}}\right].$$

Standard asymptotic theory implies that $\mathrm{P}(\beta_0 \in \mathcal{I}) \to 1-\alpha$ as $n \to \infty$.

The overidentification test statistic based on the efficient GMM estimator is formed as $J = n\hat{g}_n(\hat{\beta})'\Omega(\tilde{\beta})\hat{g}_n(\hat{\beta})$, the minimized value of the objective function in equation (3.2) . Under the null hypothesis that the model is correctly specified with equation (3.1) holding, J will converge in distribution to a chi-squared random variable with $r-q$ degrees of freedom. Then for the $1-\alpha$ quantile, J_α, of a chi-squared distribution with $r-q$ degrees of freedom, $\mathrm{P}(J > J_\alpha) = \alpha$ as $n \to \infty$.

Bootstrap Inference for GMM

Bootstrap versions of these confidence intervals and test statistics can be obtained by replacing the asymptotic critical values, z_α and J_α, by critical values that are estimated from the data. The bootstrap is based on some estimator, $\hat{F}_n$, of the distribution of $\mathbf{Z}_n = (z_1, \ldots, z_n)$. Under the assumption of i.i.d. data, $\hat{F}_n$ can be taken to be the product of the distribution of a single observation. Bootstrap (estimated) critical values for t-ratios and overidentifying test statistics can then be obtained from the distribution of these statistics under $\hat{F}_n$, that is, from the distribution of these statistics under the assumption that $\hat{F}_n$ were true. Because the distributions of the t-ratio and overidentifying test statistic are so complicated, the bootstrap has to be done by Monte Carlo simulation.

Let $\mathbf{Z}_n^* = (z_1^*, \ldots, z_n^*)$ denote a sample of n observations with distribution $\hat{F}_n$; there will be B such samples, where B denotes the number of bootstrap replications. Let β_j^*, SE_j^*, and J^* be computed exactly as described above, only with the sample $\mathbf{Z}_n^*$ replacing the sample $\mathbf{Z}_n$. In particular, the sample moments $\hat{g}_n(\beta)$ would be replaced by the Monte Carlo moments $\hat{g}_n^*(\beta) = \sum_{i=1}^n g(z_i^*, \beta)/n$ and the weighting matrices by their counterparts from the Monte Carlo data. Let $t_j^* = \sqrt{n}(\beta_j^* - \hat{\beta}_j)/\mathrm{SE}_j^*$ be the corresponding t-ratio (where $\hat{\beta}$ takes the place of β_0) and J^* be the value of the overidentifying test statistic obtained using the sample $\mathbf{Z}_n^*$.

The distribution of t_j^* and J^* across the Monte Carlo replications then can be used to estimate the corresponding critical values. The critical value for a symmetric $1-\alpha$ confidence interval for β_{0j} is the $1-\alpha$ quantile, $z^*_{\alpha j}$, of the Monte Carlo distribution of $|t^*_j|$.[3] A confidence interval based on these bootstrapped critical values can be constructed as

$$\mathcal{I}^* = \left[\hat{\beta}_j - z^*_{\alpha j}\frac{\text{SE}_j}{\sqrt{n}}, \hat{\beta}_j + z^*_{\alpha j}\frac{\text{SE}_j}{\sqrt{n}}\right]. \tag{3.3}$$

Also, an estimated α level critical value for the overidentifying test statistic is the $1-\alpha$ quantile, J^*_α, of the Monte Carlo distribution of J^*. The critical region of the overidentifying test is then $C^* = \{J \geq J^*_\alpha\}$. The corresponding inference procedures are analogous to the asymptotic ones where the critical values from the asymptotic distribution have been replaced by ones from the bootstrap distribution.

The bootstrap is critically dependent on the estimate of the distribution of the data, $\hat{F}_n$. Because the data are i.i.d., we can restrict attention to distributions that are the product of n copies of a distribution estimator of a single observation, $\hat{F}$. The bootstrap is often based on the empirical distribution, which is a discrete distribution that places probability weight $1/n$ on each observed data point. The empirical distribution is an appropriate choice in many cases where inference procedures are not based on restricted distributions. However, the inference procedures for GMM depend on the validity of the moment conditions in equation (3.1) that restrict the distribution of the data if $r > q$.

Specifically, consistency of $\hat{V}$ for the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta_0)$ depends on the validity of equation (3.1). If equation (3.1) is violated then $\hat{\beta}$ is still asymptotically normal, but the asymptotic variance formula changes, as discussed by Massoumi and Philips (1982) for instrumental variables estimators. Consequently, bootstrap confidence intervals in equation (3.3) will not provide an improvement when they are based on the empirical distribution. They will, however, be asymptotically valid for reasons discussed below.

The problem is more serious with the overidentification statistic, in that bootstrapping from the empirical distribution will give the wrong size, even asymptotically. This failure results from the need to impose the null hypothesis of a test in bootstrapping to obtain

[3] An alternative approach to bootstrap confidence intervals is to use quantiles of the empirical distribution of $\beta^*_j - \hat{\beta}_j$, as discussed by Hahn (1996). This will by asymptotically valid, but does not yield an improvement for reasons that will be discussed below. For this reason, we focus on bootstrap confidence intervals based on the t-ratio.

asymptotic validity. The null hypothesis of equation (3.1) is not imposed by the empirical distribution because the moments, $\hat{g}_n(\hat{\beta}) = \sum_{i=1}^{n} g(z_i, \hat{\beta})/n$, are generally not zero when $r > q$. Consequently, the empirical distribution corresponds to an alternative hypothesis where the moments are non-zero. This problem does not disappear as the sample size grows because $\hat{g}_n(\hat{\beta})$ will only shrink towards zero at rate $1/\sqrt{n}$. The problem can be quite severe because of the correlation between the magnitude of $\hat{g}_n(\hat{\beta})$ and the size of the bootstrap critical value. The overidentification test should tend to reject when $\hat{g}_n(\hat{\beta})$ is far from zero, but it is exactly those cases where bootstrapping from the empirical distribution should yield large critical values because they correspond to cases where the moments are far from their null hypothesis value of zero. For example, in Monte Carlo experiments where we have tried using the critical values generated by the empirical distribution, the tests never reject: the size of the test based on critical values generated by the empirical distribution is zero.

3.3 Bootstrapping from Moment Restricted Distributions

To obtain asymptotically correct critical values requires centering the bootstrap distribution in some way. There are several ways to center this distribution. Hall and Horowitz (1996) center the moments at the realized value of the test statistic. We consider another approach that involves centering the distribution that generates the bootstrap so that it has zero moments. As discussed below, this approach seems to have some theoretically attractive features relative to other approaches.

Our approach to bootstrapping for GMM is to use a distribution function estimator, $\hat{F}$, that imposes the moment conditions

$$\int g(z, \hat{\beta})\, d\hat{F} = 0.$$

We focus on distributions that are asymptotically efficient estimators of the distribution of a single observation in the semiparametric model where the distribution is unrestricted but for the moment restriction $\mathrm{E}[g(Z, \beta_0)] = 0$. There are many ways to construct such a distribution estimator; we begin by considering one interesting approach.

The empirical likelihood approach chooses $\hat{F}$ to be discrete with probability $\hat{p}_i$ for observation i, where $\hat{p}_i$ for $i = 1, \ldots, n$ solves

$$\max_{p_1,\ldots,p_n} \sum_{i=1}^{n} \log(p_i) \quad \text{subject to} \quad p_i > 0, \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i g(z_i, \hat{\beta}) = 0. \tag{3.4}$$

The term *empirical likelihood* comes from Owen (1990) who suggest construction of confidence intervals from variation in the maximized objective function as $\hat{\beta}$ changes. Our purpose is different. We consider using the probabilities $\hat{p}_i$ obtained from this maximization (with $\hat{\beta}$ equal to an optimal GMM estimator) as the basis for the bootstrap. We will construct confidence intervals and critical values by resampling from a discrete distribution with probability $\hat{p}_i$ that $Z = z_i$.

Fortunately, there is a way to compute $\hat{p}_i$ without solving the n-dimensional maximization problem in equation (3.4). Let λ denote an $r \times 1$ vector, where r is the dimension of $g(z, \beta)$, and consider the solution to

$$\max_{\lambda} \sum_{i=1}^{n} \log\left(1 + \lambda' g(z_i, \hat{\beta})\right) \text{ subject to } 1 + \lambda' g(z_i, \hat{\beta}) > 0. \tag{3.5}$$

The probabilities, $\hat{p}_i$, from the maximization problem in equation (3.5) can be computed as

$$\hat{p}_i = \frac{1}{n(1 + \hat{\lambda}' g(z_i, \hat{\beta}))}. \tag{3.6}$$

This is a simpler calculation method because the maximization problem of equation (3.5) is a lower-dimensional, concave problem.

The empirical likelihood estimator is one member of a class of distribution estimators. Let $T(v)$ denote a differentiable concave function of a scalar argument, v, with $\nabla T(v) = dT(v)/dv$ and with domain that is an open interval containing zero. Consider a distribution estimator that is discrete with probability that $Z = z_i$ given by

$$\hat{p}_i = \frac{\nabla T(\hat{\lambda}' g(z_i, \hat{\beta}))}{\sum_{j=1}^{n} \nabla T(\hat{\lambda}' g(z_j, \hat{\beta}))} \quad \text{where } \hat{\lambda} \text{ maximizes } \sum_{i=1}^{n} T(\lambda' g(z_i, \hat{\beta})). \tag{3.7}$$

The empirical likelihood estimator is an example of such a distribution estimator with

$T(v) = \log(1+v)$. Other examples include $T(v) = -(1+v)^2$ and $T(v) = -\exp(v)$. Any such estimator will be straightforward to compute as a concave programming problem where the gradient and Hessian matrix are easily calculated. In each case, the probabilities satisfy $\sum_{i=1}^n \hat{p}_i = 1$ by construction and the moment conditions, $\sum_{i=1}^n \hat{p}_i g(z_i, \hat{\beta}) = 0$, follow from the first order conditions for $\hat{\lambda}$. Furthermore, under regularity conditions discussed below, the probabilities $\hat{p}_i$ will all be positive with probability approaching one in large samples.

One of these estimators has a closed form solution. The first order condition for equation (3.7) with $T(v) = -(1+v)^2$ is given by $0 = \sum_{i=1}^n g(z_i, \hat{\beta})(1 + g(z_i, \hat{\beta})'\hat{\lambda})$. Solving for $\hat{\lambda} = -\hat{V}^{-1}\bar{g}$, where $\hat{V} = \sum_{i=1}^n g(z_i, \hat{\beta})g(z_i, \hat{\beta})'/n$ and $\bar{g} = \sum_{i=1}^n g(z_i, \hat{\beta})/n$, and plugging in to equation (3.7) gives

$$\hat{p}_i = \frac{1 - \bar{g}'\hat{V}^{-1}g(z_i, \hat{\beta})}{n(1 - \bar{g}'\hat{V}^{-1}\bar{g})}.$$

This estimator has been considered by Back and Brown (1993).

In the case of discrete data, where the moment functions $g(z, \beta)$ correspond to a full vector of cell probabilities, this GMM bootstrap reduces to the parametric bootstrap. Specifically, suppose that the support of Z consists of $r+1$ points, $\{z^1, \ldots, z^{r+1}\}$, and that the model specifies probabilities $p_j(\beta)$ for each possible outcome ($j = 1, \ldots, r+1$). Also, suppose that $g(z, \beta) = \mathrm{I}(Z = z^j) - p_j(\beta)$; the moment conditions $\mathrm{E}[g(Z, \beta_0)] = 0$ correspond to the specification of the probability of the jth outcome to be $p_j(\beta_0)$ for each outcome but the $r+1$. The $r+1$ probability is determined by the adding-up restriction that the probabilities sum to one.

In this case, the optimal GMM estimator, $\hat{\beta}$, will be asymptotically equivalent to the minimum chi-square estimator, which is known to be efficient. Furthermore, for $n_j = \sum_{i=1}^n \mathrm{I}(z_i = z^j)$ and $\nabla T^j = \nabla T(\hat{\lambda}' g(z^j, \hat{\beta}))$, the first order conditions for the maximization in equation (3.7) are $0 = n_j \nabla T^j - p_j(\hat{\beta}) \sum_{i=1}^n \nabla T(\hat{\lambda}' g(z_i, \hat{\beta}))$. Solving for $\hat{p}_i$ in equation (3.7) then gives

$$\hat{p}_i = \frac{p_j(\hat{\beta})}{n_j} \quad \text{for } z_i = z^j,\ j = 1, \ldots, r.$$

Thus, the probability of each possible outcome z^j is $p_j(\hat{\beta})$, so that bootstrapping from this distribution corresponds to the parametric bootstrap for a discrete distribution.

3.4 Theoretical Results

Regularity Conditions

Assumption 1 *β_0 is the unique solution in the compact set B to $E[g(Z,\beta)] = 0$. For every $\beta \in B$, $g(z,\beta)$ is continuous at β with probability one and* $\mathrm{E}[\sup_{\beta\in B} \|g(Z,\beta)\|] < \infty$. *$\hat{\Omega} \xrightarrow{p} \Omega$ where Ω is positive definite.*

Assumption 2 *β_0 is in the interior of B. $g(z,\beta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of β_0 with probability approaching one. $G'\Omega G$ is nonsingular for $G =$* $\mathrm{E}[\partial g(z,\beta_0)/\partial\beta]$. *For some $p > 2$,* $\mathrm{E}[\sup_{\beta\in\mathcal{N}} \|g(Z,\beta)\|^p] < \infty$*;* $\mathrm{E}[\sup_{\beta\in\mathcal{N}} \|\partial g(Z,\beta)/\partial\beta\|] < \infty$*; and $E[g(Z,\beta_0)g(Z,\beta_0)']$ is nonsingular.*

Assumption 3 *On an open interval containing zero the function $T(v)$ is twice continuously differentiable with $\nabla T(0) > 0$ and $\nabla^2 T(0) < 0$.*

Assumption 4 $\mathrm{E}[g(Z,\beta)g(z,\beta)']$ *has smallest eigenvalue that is bounded away from zero on a neighborhood $\mathcal{N}$ of β_0. There is a function $C_g(z)$ with $\|g(z,\beta_1)-g(z,\beta_2)\| \le C_g(z)\|\beta_1-\beta_2\|$ for $\beta_1,\beta_2 \in B$. $g(z,\beta)$ is four times differentiable on $\mathcal{N}$ and for $\bar{g}(z,\beta)$ a vector of the unique components of the derivatives of $g(z,\beta)$ through order 4 with respect to β, there is a function $C_{\bar{g}}$ such that $\|\bar{g}(z,\beta_1) - \bar{g}(z,\beta_2)\| \le C_{\bar{g}}(z)\|\beta_1 - \beta_2\|$ for $\beta_1,\beta_2 \in \mathcal{N}$. Let C denote C_g or $C_{\bar{g}}$. Then* $\mathrm{P}[C(Z) > r] = \mathrm{O}(r^{-33})$ *as $r \to \infty$.*

Assumption 5 *For $f(z)$ a vector of the unique components of $g(z,\beta_0)$ and $g(z,\beta_0)g(z,\beta_0)'$ and their derivatives with respect to β through order 4, $f(z)$ is a Lipschitz continuous function of z and as $r \to \infty$,* $\mathrm{P}(\|f(z)\| > r) = \mathrm{O}(r^{-33})$.

Assumption 6 *Z can be partitioned $(Z^{(c)}, Z^{(d)})$ where $Z^{(c)} \in \mathbb{R}^c$ for some $c > 0$; the distribution of $Z^{(c)}$ and $\partial g(z,\beta)/\partial\beta$ are absolutely continuous with respect to Lebesgue measure, and the distribution of $Z^{(d)}$ is discrete. There need not be any discrete components of Z, but there must be at least one continuous component.*

Assumption 7 $\limsup_{\|t\|\to\infty} |\chi(t)| < 1$ *where $\chi(t)$ is the characteristic function of Z,* $\mathrm{E}[\exp(it'Z)]$.

Theorems

This section contains a formal statement of the results of the chapter. The following additional notation is used in the statements of these results. Let P denote the probability measure induced by the data-generation process and let P^* denote the probability measure induced by bootstrap sampling conditional on the estimation data. Expectation under the probability measure P is denoted by E; E^* corresponds to expectation under the probability measure P^*. Our first result is that the distribution estimators discussed in Section 3.3 attain the semiparametric efficiency bound of Brown and Newey (1994). The efficiency property of $\hat{F}$ is potentially important because it should help ensure that the bootstrap has good large sample properties. In a sense that will be discussed in Section 3.5, an efficient bootstrap method should result from an efficient estimator of the distribution.

Theorem 1 *Let Assumptions 1–3 hold. Then with probability approaching one* $\hat{F}(z) = \sum_{i=1}^{n} \hat{p}_i \mathrm{I}(z \leq z_i)$ *exists and attains the semiparametric efficiency bound for estimation of the distribution of a single observation in the model* $\mathrm{E}[g(Z, \beta_0)] = 0$.

The next result gives conditions under which the bootstrap critical values for J^* and T_j^* yield an improvement over the asymptotic critical values. Let $z_{\alpha j}^*$ and J_α^*, respectively, denote the α-level bootstrap critical values of the t-statistic and the test of overidentifying restrictions.

Theorem 2 *Let Assumptions 1–7 hold. Under* $H_0 : \beta_j = \beta_{0j}$

$$\mathrm{P}(|t_j| > z_{\alpha j}^*) = \alpha + o(n^{-1}).$$

If $r > q$ *and the overidentifying restrictions hold,*

$$\mathrm{P}(J > J_\alpha^*) = \alpha + o(n^{-1}).$$

To prove this result, we show that J^* and T_j^* can be written as smooth functions of bootstrap sample moments up to an asymptotically negligible remainder term in Proposition 1. Using this result, Proposition 2 demonstrates that the distributions of J^* and T_j^* have an Edgeworth expansion through $O(n^{-1})$. Application of Theorem 1 of Hall and Horowitz (1996) to independent data proves the analogous result for J and T_j. Theorem 2 follows from a combination of these two results.

3.5 The Bootstrap Improvement

To see that the GMM bootstrap we have proposed leads to better approximations it is useful to consider a theoretical justification. This theory is similar to that of Beran (1988) and is based on an Edgeworth expansion. Let F denote a possible distribution for Z_i and let $S = S_n(Z_1, \ldots, Z_n, F)$ be a random variable whose distribution we are interested in approximating when F is the true distribution of a single observation (examples of S are the overidentifying test statistic and a t-ratio). The overidentifying test statistic will not depend on F, but the t-ratio will because the t-ratio depends on the true parameter, which, in turn, depends on the true distribution. We will focus on the distribution of S when F is true given by

$$H_n(F) = \mathrm{P}(S_n(Z_1, \ldots, Z_n) \leq s|F)$$

where s is an argument that is suppressed for notational convenience.

The true distribution of S will be $H_n(F_0)$ where F_0 is the true distribution of Z_i, and the asymptotic distribution of S will be the limit of $H_n(F_0)$ as $n \to \infty$. A bootstrap estimator of the distribution of S is $H_n(\hat{F})$ where $\hat{F}$ is some estimator of the distribution of a single observation. Typically, $H_n(\hat{F})$ will be difficult to calculate and so is approximated by Monte Carlo simulation methods. A theoretical comparison of the asymptotic and bootstrap approximation errors can then be based on comparing the size of $[\lim_{n\to\infty} H_n(F_0)] - H_n(F_0)$ and $H_n(\hat{F}) - H_n(F_0)$. When the bootstrap distribution is closer, the replacement of asymptotic critical values with bootstrap critical values (as discussed in Section 3.2) should lead to an improved approximation. This improvement will occur because these critical values are essentially quantiles of the bootstrap distribution, which should be closer to the true quantiles when the distributions are closer.

A formal expansion in n of the distribution is useful for understanding when the bootstrap will improve on the asymptotic approximation. Suppose that

$$H_n(F) = H_\infty + n^{-\alpha}R(F) + o(n^{-\alpha}) \tag{3.8}$$

uniformly in F as $n \to \infty$ for $\alpha > 0$. This corresponds to to an Edgeworth expansion where H_∞ is the asymptotic or limiting distribution and $n^{-\alpha}R(F)$ is the next order term (in many cases α will be 1/2 or greater). An important feature of S that is implicit in this expansion is that the limiting distribution does not depend on F, that is, S is *asymptotically pivotal*. Assume that $R(\hat{F})$ is consistent for $R(F_0)$, $R(\hat{F}) \xrightarrow{p} R(F_0)$. Then the approximation errors of asymptotic and bootstrap distributions are given by

$$\begin{aligned} H_\infty - H_n(F_0) &= -n^{-\alpha}R(F_0) + o(n^{-\alpha}) = O(n^{-\alpha}) && \text{(Asymptotic)} \\ H_n(\hat{F}) - H_n(F_0) &= n^{-\alpha}[R(\hat{F}) - R(F_0)] + o(n^{-\alpha}) = o_p(n^{-\alpha}) && \text{(Bootstrap).} \end{aligned} \tag{3.9}$$

The remainder term in the large sample approximation is of order $n^{-\alpha}$, while the remainder in the bootstrap approximation is of smaller order. Thus, the bootstrap is an improvement in the sense that in large enough samples the remainder term will be smaller with high probability.

A feature that is essential for the bootstrap to yield an improvement is that S be asymptotically pivotal: the limiting distribution of S, H_∞, can not depend on F. When S satisfies this condition, the leading term in the difference $H_n(\hat{F}) - H_n(F_0)$ drops out. If H_∞ depended on F then the bootstrap approximation error would be of the same order as $H_\infty(\hat{F}) - H_\infty(F_0)$, which would typically be of no smaller order than the asymptotic approximation. This is why it is important to bootstrap the t-ratio using a moment restricted distribution. If the distribution does not satisfy the moment restrictions, then the formula for the asymptotic variance of a GMM estimator is not $(G\Omega G)^{-1}$, it is more complicated (for an example, see Maasoumi and Phillips, 1982). Hence, the t-ratio based on this formula would not have a standard normal distribution; it would have an asymptotic variance that depended on F, and H_∞ would depend on F. Essentially, by imposing the moment restrictions, $\hat{F}$ is restricted to a class of distributions for which the t-ratio and overidentifying test statistic are asymptotically pivotal, so that the bootstrap yields an improvement.

Efficiency considerations also motivate the choice of a moment-restricted distribution in bootstrapping for GMM. As previously shown, the moment-restricted distribution estimators we have proposed are the asymptotically most efficient estimators available if all that is known is the moment condition, $\mathrm{E}[g(Z, \beta_0)] = 0$. Also, since the bootstrap approximation error is $H_n(\hat{F}) - H_n(F_0)$, it is reasonable to expect that an asymptotically more efficient estimator $\hat{F}$ should lead to a smaller approximation error in large samples. This argument can be made more precise when the expansion of equation (3.8) has some further structure. Suppose that we include higher-order terms in the expansion of the distribution of S

$$H_n(F) = H_\infty + n^{-\alpha} R_1(F) + n^{-\gamma} R_2(F) + o(n^{-\gamma})$$

where $\gamma \geq \alpha + 1/2$. Then

$$\begin{aligned} n^{\alpha+1/2}[H_n(\hat{F}) - H_n(F_0)] &= \sqrt{n}[R_1(\hat{F}) - R_1(F_0)] + n^{-\gamma+\alpha+1/2}[R_2(\hat{F}) - R_2(F_0)] + o_p(1) \\ &= \sqrt{n}[R_1(\hat{F}) - R_1(F_0)] + o_p(1). \end{aligned}$$

So long as the remainder term, $R_1(F)$, and the distribution estimator, $\hat{F}$, satisfy certain regularity conditions, this equation implies that $n^{\alpha+1/2}[H_n(\hat{F}) - H_n(F_0)]$ is asymptotically normal, and that its asymptotic variance depends directly on that of $\hat{F}$. This result is analogous to the well-known delta method result that the asymptotic variance of a nonlinear function of a vector of estimators is proportional to the asymptotic variance of the estimators. Thus, the more efficient is $\hat{F}$, the closer $H_n(\hat{F})$ will be to $H_n(F_0)$ in large samples, in the sense that $n^{\alpha+1/2}[H_n(\hat{F}) - H_n(F_0)]$ will have smaller asymptotic variance.

These efficiency considerations suggest that the GMM bootstrap of Section 3.3 should have superior large sample properties to the procedure proposed by Hall and Horowitz (1996). Their approach is to bootstrap from the empirical distribution and to center the moments in the bootstrap sample around the empirical estimate. For example, for a bootstrap sample $z_1^b, \ldots, z_n^b$, one would compute an initial weighting matrix Ω^b and, using the centered moments, $\tilde{g}_n^b(\beta) = \sum_{i=1}^n g(z_i^b, \beta)/n - \sum_{i=1}^n g(z_i, \hat{\beta})/n$, compute the initial estimator $\tilde{\beta}^b$ by minimizing

$$\tilde{g}_n^b(\beta)' \Omega^b \tilde{g}_n^b(\beta).$$

The optimal weighting matrix, $\Omega^b(\tilde{\beta})$, would be computed as the inverse of the sample variance of $\tilde{g}_n^b(\tilde{\beta})$, and the overidentifying test statistic would be computed as the minimized value of

$$n\tilde{g}_n^b(\beta)'\Omega^b(\tilde{\beta})\tilde{g}_n^b(\beta).$$

The overidentifying test statistic is a random variable of the form $S(z_1^b, \ldots, z_n^b, \tilde{F})$; $\tilde{F}$ is the empirical distribution, which affects S through $\sum_{i=1}^n g(z_i, \hat{\beta})/n = \int g(z, \hat{\beta})\, d\tilde{F}$. As discussed by Hall and Horowitz (1996), the overidentifying test statistic will be asymptotically pivotal over all distributions because the moments are centered at their expectation. However, it is easy to see that the moment-restricted bootstrap is exactly this calculation with the empirical distribution, $\tilde{F}$, replaced by the restricted distribution, $\hat{F}$. This occurs because the moments for the restricted distribution are zero ($\int g(z, \hat{\beta})\, d\hat{F} = 0$), so that the centered moments are the same as the uncentered moments. Therefore, the moment-restricted bootstrap is identical to the Hall and Horowitz (1996) bootstrap except that the empirical distribution has been replaced with the more efficient estimator that imposes the moment restrictions. The efficiency considerations discussed above suggest that the moment restricted bootstrap should have superior large sample properties.

A simple example may help to clarify the efficiency issue. Suppose that Z is a scalar and the hypothesis of interest is $\mathrm{E}[Z] = 0$. A standard test statistic is $T = (\sum_{i=1}^n z_i)^2/(\sum_{i=1}^n z_i^2)$. This statistic corresponds to the GMM test where there is a single moment condition, the parameter β is not present, and $g(z, \beta) = z$. The standard approach to bootstrap estimation of the distribution of T centers the observations at the sample mean by simulating the distribution of $[\sum_{i=1}^n (z_i^b - \bar{z})]^2/[\sum_{i=1}^n (z_i^b - \bar{z})^2]$, where $\bar{z} = \sum_{i=1}^n z_i/n$. This standard bootstrap method also is a special case of that suggested by Hall and Horowitz (1996), and is that same as bootstrapping the distribution of T from the empirical distribution of $z_1 - \bar{z}, \ldots, z_n - \bar{z}$.

This distribution estimator is not efficient except in the case where Z has a normal distribution. Consequently, using one of the efficient estimators from Section 3.3 should, under the null hypothesis, produce a more efficient bootstrap for some distributions. Specifically, if Z is continuously distributed and has moments of sufficient order then there is an expansion

$$\mathrm{P}(T \geq k|F) = \mathrm{P}(\chi_1^2 \geq k) + \\ n^{-1}k\left[\frac{k^2-3}{6}\kappa(F) - \frac{k^4+2k^2-3}{9}\gamma(F)^2 - \frac{k^2+4}{2}\right]\phi(k) + \mathrm{O}(n^{-2})$$

where $\kappa(F)$ and $\gamma(F)$ denote the skewness and kurtosis of the distribution F, and ϕ is the standard normal density (see Hall, 1992, page 73). Multiplying by $n^{3/2}$ and differencing then gives

$$n^{3/2}\left[\mathrm{P}(T \geq k|\hat{F}) - \mathrm{P}(T \geq k|F_0)\right] = \\ \sqrt{n}k\left[\frac{k^2-3}{6}\{\kappa(\hat{F}) - \kappa(F_0)\} - \frac{k^4+2k^2-3}{9}\{\gamma(\hat{F})^2 - \gamma(F_0)^2\}\right]\phi(k) + o_p(1).$$

Thus, the efficiency of the bootstrap will be determined by the efficiency of the corresponding estimators of moments of the distribution (a linear combination of skewness and kurtosis in this example). In this example, the efficiency gain obtained by using the moment restricted bootstrap will depend on the distribution.

3.6 Monte Carlo Experiments

A Monte Carlo experiment was performed to obtain information on the performance of the moment-restricted bootstrap. Consider an autoregressive panel data model with an individual effect of the form

$$y_{it} = \rho_0 y_{i,t-1} + \alpha_i + \varepsilon_{it} \quad \text{and} \quad y_{i0} = \frac{\alpha_i}{1-\rho_0} + \nu_i \tag{3.10}$$

for $t = 1, \ldots, 4$ and $i = 1, \ldots, n$. We assume that ε_{it} and α_i have standard normal distributions, ν_i has a normal distribution with mean zero and variance $1/(1-\rho_0^2)$, and $\varepsilon_{i1}, \ldots, \varepsilon_{i4}, \alpha_i, \nu_i$ are mutually independent. In this model, y_{it} is stationary over time because of the specification of the equation describing y_{i0}.

The fixed effects estimator that includes a distinct constant for each i is inconsistent by a wide margin (Nickel, 1981). One approach to consistent estimation is to time difference

Table 3.1: Dynamic Panel Model Monte Carlo Results

	$n = 50$		$n = 100$	
Nominal	Asymptotic	Bootstrap	Asymptotic	Bootstrap
0.90	0.80	0.88	0.85	0.90
0.10	0.12	0.11	0.126	0.114
0.05	0.066	0.058	0.065	0.058
0.01	0.012	0.013	0.014	0.015

the observations and use lagged levels as instrumental variables, similar to Anderson and Hsaio (1982) and Holtz-Eakin, Newey, and Rosen (1988). Equation (3.10) implies that

$$\Delta y_{it} = \rho_0 \Delta y_{i,t-1} + \Delta \varepsilon_{it}$$

and $\mathrm{E}(y_{i,t-j}\Delta\varepsilon_{it}) = 0$ for $j \geq 2$. In the example where there are four time periods, these conditions lead to a three dimensional moment vector where $\beta = \rho$ and

$$g(z_i, \beta) = [y_{i1}(\Delta y_{i3} - \beta\Delta y_{i2}), y_{i1}(\Delta y_{i4} - \beta\Delta y_{i3}), y_{i2}(\Delta y_{i4} - \beta\Delta y_{i3})]'.$$

The initial weighting matrix used in estimation was a block diagonal matrix, with the first diagonal block consisting of the inverse sample second moment of y_{i1} and the second block consisting of the inverse sample second moment of (y_{i1}, y_{i2}). The sample size n and the autoregressive coefficient ρ are parameters of the design. We use a value of $\rho = 0.5$ and two sample sizes, $n = 50$ and $n = 100$. Table 3.1 reports the results of this experiment: the table lists the actual sizes of asymptotic and bootstrap tests of nominal sizes 0.10, 0.05, and 0.01 and the coverage probabilities for nominal 90 percent confidence intervals.

An experiment was also carried out using an additional moment condition that was suggested by Ahn and Schmidt (1994): uncorrelatedness of α_i and ε_{it} for each time period implies that $\mathrm{E}[(y_{i4} - \rho_0 y_{i3})(\Delta y_{i3} - \rho_0 \Delta y_{i2})] = 0$. This moment condition is nonlinear in ρ, but it is straightforward to use an initial estimator to form a linearized version. Linearizing this moment condition around an initial estimator $\tilde{\rho}$ gives

$$y_{i4}\Delta y_{i3} - \tilde{\rho}^2 y_{i3}\Delta y_{i2} - (y_{i3}\Delta y_{i3} + y_{i4}\Delta y_{i2} - 2\tilde{\rho} y_{i3}\Delta y_{i2})\,\rho.$$

Table 3.2: Dynamic Panel Model Monte Carlo Results

	$n = 50$		$n = 100$	
Nominal	Asymptotic	Bootstrap	Asymptotic	Bootstrap
0.90	0.70	0.89	0.79	0.90
0.10	0.23	0.099	0.11	0.083
0.05	0.13	0.045	0.057	0.040
0.01	0.037	0.009	0.014	0.007

We add this fourth moment condition to the three from equation (3.6) and re-evaluate the performance of asymptotic and bootstrapped critical values. Table 3.2 reports the results of this experiment. With an extra degree of overidentification, the bootstrap gives an even bigger improvement than in the previous case.

3.7 Empirical Example

An empirical example may help shed light on how the bootstrap improvement can affect results. The example involves estimation of a Q model of investment using a panel of firms taken from Blundell et al. (1998). The statistical model takes the form:

$$(I/K)_{it} = \beta Q_{it} + \gamma_t + \alpha_i + \nu_{it} \tag{3.11}$$

for firms $i = 1, \ldots, n$ and time periods $t = 1, \ldots, T$. Here Q_{it} is an empirical measure of the ratio of the shadow value of capital and the unit price of investment goods, and $(I/K)_{it}$ is the investment rate. We use a subset of the data from Blundell et al., choosing a balanced panel of 532 firms over 10 years with different initial years ranging from 1971 to 1977. As such, the coefficients γ_t in equation (3.11) are specified by calendar year, rather than by time elapsed from the initial year in the sample.

To estimate the investment coefficient, β, consider instrumental variables estimators that use lagged values of Q as instruments to estimate the differenced equation. These instruments allow for contemporaneous correlation of Q_{it} and ν_{it}, and for correlation with the fixed effect. This specification leaves different numbers of instruments for different time periods because more lags are available as instruments for the later time periods. After differencing and using a year of data to generate lagged variables we were left with eight years of data. We also dropped the last year of data for each firm because of a singularity problem with the bootstrap and because the γ_t coefficient for the last year seemed to be quite different than the others, suggesting that it might be an outlier. This leaves seven years of data and a total of $1 + 2 + 3 + 4 + 5 + 6 + 7 = 28$ instruments.

In describing the results we focus on the estimator of β and on the overidentifying test statistic. The estimate of β was 0.00758 with a standard error of 0.00376. The point estimate is similar to that of Blundell et al. (1992), although it is smaller and much less precisely estimated. According to the asymptotic standard errors, it is significant at the 0.05 significance level. The overidentifying test statistic is 32.14 and has 27 degrees of freedom; it is below the asymptotic critical value of 40.11.

The bootstrap confidence intervals for the coefficient are wider than the asymptotic confidence intervals. The bootstrap p-value for the t-statistic was 0.103, compared to the asymptotic value which was less than 0.05. Thus, the coefficient is less statistically significant when the bootstrap approximation is used in place of the asymptotic critical values. Also, the bootstrap distribution of the overidentifying test statistic has more weight in the tail than the asymptotic distribution; the bootstrap 0.05 critical value is 42.79, compared to the asymptotic critical value of 40.11.

3.8 Conclusions

Critical values based on asymptotic theory often provide a poor approximation to the distribution of GMM estimators. In particular, asymptotic theory provides a poor approximation to the distribution of GMM estimators when there are many overidentifying restrictions or the parameters of interest are not well identified. As a way of dealing with this difficulty, this chapter describes a relatively efficient bootstrap based on a moment-restricted distribution.

The results of this chapter show that our moment-restricted distribution estimator attains the semiparametric efficiency bound for estimation of the distribution of a single observation in the model $\mathrm{E}[g(Z, \beta_0)] = 0$. We also show that the moment-restricted bootstrap provides asymptotic refinements to the critical values of symmetrical t-tests and the test of overidentifying restrictions under certain conditions. In both Monte Carlo simulations and an empirical example using a dynamic panel data model, we show that the bootstrap improvement over asymptotic theory can be large.

3.9 Appendix

Lemma 1 *Let Assumptions 1–5 hold. Consider a distribution estimator that is discrete with probability that* $z = z_i$ *given by*

$$\hat{p}_i = \frac{\nabla T\left(\hat{\lambda}'\hat{g}_i\right)}{\sum_{j=1}^{n} \nabla T\left(\hat{\lambda}'\hat{g}_j\right)}$$

where $\hat{\lambda}$ *solves* $\max_{\lambda} \sum_{j=1}^{n} T\left(\lambda'\hat{g}_j\right)$. *Then with* P *probability* $1 - o(n^{-1})$, $\hat{p}_i > 0$ *for* $i = 1, \ldots, n$. *Furthermore, with* P *probability* $1 - o(n^{-1})$, *there exists a* $C < \infty$ *such that* $\max_{i \leq n} |\hat{p}_i| \leq Cn^{-1}$.

Proof First we show that there exists an α with $0 < \alpha < 1/2$ such that

$$\lim_{n \to \infty} n \, \mathrm{P}\left[\|\bar{g}\| > n^{-\alpha}\right] = 0.$$

With the function C_g as defined in the assumptions, consider $\|\bar{g}\|$

$$\begin{aligned}
\|\bar{g}\| &= n^{-1}\left\|\sum_{i=1}^{n} g(Z_i, \hat{\beta})\right\| \\
&\leq n^{-1}\sum_{i=1}^{n}\left\|g(Z_i, \hat{\beta}) - g(Z_i, \beta_0)\right\| + \left\|n^{-1}\sum_{i=1}^{n} g(Z_i, \beta_0)\right\| \\
&\leq n^{-1}\sum_{i=1}^{n} \|\hat{\beta} - \beta_0\| C_g(Z_i) + \left\|n^{-1}\sum_{i=1}^{n} g(Z_i, \beta_0)\right\|.
\end{aligned}$$

Therefore

$$n\,\mathrm{P}[\|\bar{g}\| > n^{-\alpha}] \leq$$
$$n\,\mathrm{P}\left[n^{-1}\sum_{i=1}^{n}\|\hat{\beta}-\beta_0\|C_g(Z_i) > (1/2)n^{-\alpha}\right] + n\,\mathrm{P}\left[\left\|n^{-1}\sum_{i=1}^{n} g(Z_i,\beta_0)\right\| > (1/2)n^{-\alpha}\right]$$

where $\lim_{n\to\infty} n\,\mathrm{P}[\|n^{-1}\sum_{i=1}^{n} g(Z_i,\beta_0)\| > (1/2)n^{-\alpha}] = 0$ for $\alpha = (2+\epsilon)/5$ with $0 < \epsilon < 1/64$ by Lemma 1 of Hall and Horowitz (1996). Next we note that

$$n\,\mathrm{P}\left[\|\hat{\beta}-\beta_0\|n^{-1}\left|\sum_{i=1}^{n}[C_g(Z_i) - \mathrm{E}\,C_g(Z) + \mathrm{E}\,C_g(Z)]\right| > (1/2)n^{-\alpha}\right] \leq$$
$$n\,\mathrm{P}\left[\|\hat{\beta}-\beta_0\|\,\mathrm{E}\,C_g(Z) > (1/4)n^{-\alpha}\right] +$$
$$n\,\mathrm{P}\left[\|\hat{\beta}-\beta_0\|n^{-1}\left|\sum_{i=1}^{n}[C_g(Z_i) - \mathrm{E}\,C_g(Z)]\right| > (1/4)n^{-\alpha}\right]$$

where $\lim_{n\to\infty} n\,\mathrm{P}[\|\hat{\beta}-\beta_0\|\,\mathrm{E}\,C_g(Z) > (1/4)n^{-\alpha}] = 0$ by Lemma 3 and Lemma 4 of Hall and Horowitz (1996) for $0 < \alpha \leq 2/5$. Also

$$n\,\mathrm{P}\left[\|\hat{\beta}-\beta_0\|\left|n^{-1}\sum_{i=1}^{n}[C_g(Z_i) - \mathrm{E}\,C_g(Z)]\right| > (1/4)n^{-\alpha}\right] \leq$$
$$n\,\mathrm{P}\left[\|\hat{\beta}-\beta_0\| > 1/4\right] + n\,\mathrm{P}\left[n^{-1}\left|\sum_{i=1}^{n}[C_g(Z_i) - \mathrm{E}\,C_g(Z)]\right| > n^{-\alpha}\right] \quad \text{(A1)}$$

where the first summand on the right hand side of equation (A1) is shown to have a limit of zero in the proofs of Lemma 3 and Lemma 4 of Hall and Horowitz (1996) and the second summand on the right hand side of equation (A1) may be shown to have a limit of zero by Lemma 1 of Hall and Horowitz (1996) for $\alpha = (2+\epsilon)/5$ where $0 < \epsilon < 1/64$. Choosing $\alpha = 2/5$, we conclude that $\lim_{n\to\infty} n\,\mathrm{P}[\|\bar{g}\| > n^{-\alpha}] = 0$.

By Lemma 2 of Hall and Horowitz (1996), for any $r > 0$

$$\lim_{n\to\infty} n\,\mathrm{P}\left[\sup_{\beta\in B} n^{-1}\left\|\sum_{i=1}^{n}[g(Z_i,\beta)g(Z_i,\beta)' - \mathrm{E}\,g(Z,\beta)g(Z,\beta)']\right\| > r\right] = 0.$$

Defining $\hat{V} = n^{-1}\sum_{i=1}^{n} g(z_i,\hat{\beta})g(z_i,\hat{\beta})'$ and $V = \mathrm{E}\,g(Z,\beta_0)g(Z,\beta_0)'$ we see that $\mathrm{E}\,\hat{V} = V$.

Application of Lemma 2 yields

$$\mathrm{P}\left[\|\hat{V}-V\| \leq r\right] = 1 - \mathrm{o}(n^{-1}).$$

By the nonsingularity of V, there is a $\delta > 0$ such that with probability $1-\mathrm{o}(n^{-1})$ the smallest eigenvalue of $\hat{V}$ is larger than δ. Let $\bar{v}$ be a compact interval containing zero in its interior such that ∇T and $\nabla^2 T$ are bounded away from zero on $\bar{v}$. Define $\hat{Q}(\lambda) = n^{-1}\sum_{i=1}^{n} T(\lambda'\hat{g}_i)$ and $\hat{\Lambda} = \{\lambda : \lambda'\hat{g}_i \in \bar{v}, i = 1, \ldots, n\}$. Choose $\bar{\delta}$ such that $\nabla^2 T(v) \leq \bar{\delta} < 0$ for all $v \in \bar{v}$ so that

$$\hat{H}(\lambda) \equiv \frac{\partial^2 \hat{Q}(\lambda)}{\partial\lambda\partial\lambda'} = n^{-1}\sum_{i=1}^{n} \nabla^2 T(\lambda'\hat{g}_i)\hat{g}_i\hat{g}_i' \leq \bar{\delta}\hat{V}$$

for all $\lambda \in \hat{\Lambda}$. Therefore there is a $\tilde{\delta} < 0$ a such that with probability $1-\mathrm{o}(n^{-1})$ the largest eigenvalue of $H(\lambda)$ is less than $\tilde{\delta}$ for all $\lambda \in \hat{\Lambda}$. Let $\tilde{\lambda} = \max_{\lambda\in\hat{\Lambda}} \hat{Q}(\lambda)$. Then by a mean value expansion

$$\begin{aligned}
\hat{Q}(0) \leq \hat{Q}(\tilde{\lambda}) &= \hat{Q}(0) + \frac{\partial \hat{Q}(0)}{\partial\lambda}'\tilde{\lambda} + (1/2)\tilde{\lambda}'\hat{H}(\bar{\lambda})\tilde{\lambda} \\
&\leq \hat{Q}(0) + \left\|\frac{\partial \hat{Q}(0)}{\partial\lambda}\right\| \|\tilde{\lambda}\| + (1/2)\|\tilde{\lambda}\|^2\bar{\delta}
\end{aligned}$$

where $\bar{\lambda}$ lies on a line between $\tilde{\lambda}$ and 0 and so is in $\hat{\Lambda}$ by the convexity of $\hat{\Lambda}$. Subtracting and dividing through by $\|\tilde{\lambda}\|$ for $\|\tilde{\lambda}\| \neq 0$ gives

$$\|\tilde{\lambda}\| \leq (-2/\bar{\delta})\left\|\frac{\partial \hat{Q}(0)}{\partial\lambda}\right\| = (-2/\bar{\delta})|\nabla T(0)|\|\bar{g}\|.$$

It follows that there exists an $\bar{\alpha}$ with $0 < \bar{\alpha} < 1/2$ such that

$$\mathrm{P}\left[\|\tilde{\lambda}\| \leq n^{-\bar{\alpha}}\right] = 1 - \mathrm{o}(n^{-1}).$$

Furthermore, for any constant r and for $\delta_i = \max_{\beta\in B} \|g(z_i,\beta)\|$,

$$\begin{aligned}
\mathrm{P}\left[\max_{i\le n}|\tilde{\lambda}'\hat{g}_i| \ge r\right] &\le \mathrm{P}\left[\|\tilde{\lambda}\| \max_{i\le n}\|\hat{g}_i\| \ge r\right] \\
&\le \mathrm{P}\left[\|\tilde{\lambda}\| \ge n^{-\bar{\alpha}}\right] + \mathrm{P}\left[\max_{i\le n}\delta_i \ge rn^{\bar{\alpha}}\right] \\
&= \mathrm{o}(n^{-1}) + \mathrm{E}\left[\left(\max_{i\le n}\delta_i\right)^p\right]/r^p n^{\bar{\alpha}p} \\
&= \mathrm{o}(n^{-1}) + n^{1-\bar{\alpha}p}\,\mathrm{E}\left[(\delta_i)^p\right]/r^p \\
&= \mathrm{o}(n^{-1})
\end{aligned}$$

for p such that $\bar{\alpha}p > 2$. Therefore, with probability $1-\mathrm{o}(n^{-1})$, $\tilde{\lambda}$ is in the interior of $\hat{\Lambda}$ and hence satisfies the first-order condition $\partial\hat{Q}(\tilde{\lambda})/\partial\lambda = 0$. Since $H(\tilde{\lambda})$ is negative definite with probability $1-\mathrm{o}(n^{-1})$, $\tilde{\lambda}$ is a unique local maximum and hence a unique global maximum, so that $\tilde{\lambda} = \hat{\lambda}$ with probability $1-\mathrm{o}(n^{-1})$. Summarizing, we have now shown that, with probability $1-\mathrm{o}(n^{-1})$, $\hat{\lambda}$ exists, is unique, and satisfies

$$\sum_{i=1}^{n} \nabla T(\hat{\lambda}'\hat{g}_i)\hat{g}_i = 0.$$

It also follows that, for any $r>0$, $\mathrm{P}[\max_{i\le n}|\hat{\lambda}'\hat{g}_i| > r] = \mathrm{o}(n^{-1})$ so that, with probability $1-\mathrm{o}(n^{-1})$, $\nabla T(\hat{\lambda}'\hat{g}_i)$ and $\nabla T(0)$ have the same sign for $i=1,\ldots,n$. Hence

$$\hat{p}_i = \frac{\nabla T(\hat{\lambda}\hat{g}_i)}{\sum_{i=1}^{n}\nabla T(\hat{\lambda}\hat{g}_j)} > 0$$

for $i=1,\ldots,n$ with probability $1-\mathrm{o}(n^{-1})$.

It follows from the above that

$$\mathrm{P}\left[\max_{i\le n}|\nabla T(\hat{\lambda}'\hat{g}_i) - \nabla T(0)| > r\right] = \mathrm{o}(n^{-1}) \qquad \text{(A2)}$$

for any $r>0$. Then with $\hat{p}_i = \nabla T(\hat{\lambda}'\hat{g}_i)/\sum_{j=1}^{n}\nabla T(\hat{\lambda}'\hat{g}_j)$, we have that

$$\max_{i\le n}|\hat{p}_i| \le \frac{n^{-1}\left(\max_{i\le n}|\nabla T(\hat{\lambda}'\hat{g}_i) - \nabla T(0)| + |\nabla T(0)|\right)}{n^{-1}|\sum_{j=1}^{n}\nabla T(\hat{\lambda}'\hat{g}_j)|}.$$

Combining this with equation (A2) and the fact that

$$\operatorname*{plim}_{n\to\infty} n^{-1}\sum_{j=1}^{n}\nabla T(\hat{\lambda}'\hat{g}_j) = \nabla T(0),$$

we have that, with probability $1 - o(n^{-1})$, there exists a $C < \infty$ with

$$\max_{i\le n} |\hat{p}_i| \le Cn^{-1}. \blacksquare$$

Lemma 2 *Let Assumptions 1–5 hold. Let a be a function such that* $\mathrm{E}\,|a(Z)| < \infty$*. Then for* $\gamma > 1$ *and any* $r > 0$

$$\lim_{n\to\infty} n\,\mathrm{P}\left[\mathrm{E}^*\,|a(Z)| > rn^{\gamma}\right] = 0.$$

Proof Define the event S to be the set of all samples χ such that $\hat{p}_i > 0$ for $i = 1, \dots, n$ and $\max_{i\le n} |\hat{p}_i| \le Cn^{-1}$ where $C < \infty$. By Lemma 1, $\mathrm{P}(S) = o(n^{-1})$. Conditional on the event S occurring, the expectation in $n\,\mathrm{P}\left[\mathrm{E}^*\,|a(Z)| > rn^{\gamma}\right]$ may be bounded as follows:

$$\mathrm{E}^*\,|a(Z)| = \sum_{i=1}^{n}\hat{p}_i|a(Z_i)| \le Cn^{-1}\sum_{i=1}^{n}|a(Z_i)|.$$

Expanding $\mathrm{P}[\mathrm{E}^*\,|a(Z)| > rn^{\gamma}]$ and making use of the above inequality in conjunction with Markov's inequality yields:

$$\begin{aligned}
\mathrm{P}[\mathrm{E}^*\,|a(Z)| > rn^{\gamma}] &= \mathrm{I}_S(\chi)\,\mathrm{P}\left[\mathrm{E}^*\,|a(Z)| > rn^{\gamma}\right] + \mathrm{I}_{S^c}(\chi)\,\mathrm{P}\left[\mathrm{E}^*\,|a(Z)| > rn^{\gamma}\right] \\
&\le \mathrm{P}\left[Cn^{-1}\sum_{i=1}^{n}|a(Z_i)| > rn^{\gamma}\right] + o(n^{-1}) \\
&\le (C/r)n^{-\gamma}\,\mathrm{E}\,|a(Z)| + o(n^{-1})
\end{aligned}$$

which is $o(n^{-1})$ as $n \to \infty$ for $\gamma > 1$. $\blacksquare$

In subsequent proofs we will assume that $\hat{p}_i > 0$ for $i = 1, \dots, n$ and neglect explict treatment of the P probability $o(n^{-1})$ event in which the distribution function $\hat{F}$ does not exist. A more formal treatment of this issue would mimic the technique used above in the proof of Lemma 2.

Lemma 3 *Let Assumptions 1–5 hold. Let h be a function such that* $\mathrm{E}\, h(Z) = 0$, $\mathrm{P}(|h(Z)| > z) = \mathrm{O}(z^{-33})$ *as* $z \to \infty$, *and for each sample* $\{Z_i : i = 1, \dots, n\}$, $\sum_{i=1}^{n} \hat{p}_i h(Z_i) = 0$. *Let* Z^* *be a random sample from the distribution* $\hat{F}$. *Define*

$$R_i^* = h(Z_i^*)$$

and

$$\bar{R}^* = n^{-1} \sum_{i=1}^{n} R_i^*.$$

For each $\delta > 0$

$$\lim_{n\to\infty} n\,\mathrm{P}[n\,\mathrm{P}^*(|\bar{R}^*| > n^{-9/25}) > \delta] = 0.$$

Proof Let $\alpha = -9/25$ and define $\mathrm{P}' \equiv n\,\mathrm{P}^*(|\bar{R}^*| > n^{\alpha})$. Then for any $p > 0$ and some constant C_0 which depends only on p it follows from Markov's inequality and Burkholder's inequality that:

$$\mathrm{P}' = n\,\mathrm{P}^* \left(\left| \sum_{i=1}^{n} R_i^* \right| > n^{1+\alpha} \right) \leq n \frac{\mathrm{E}^* |\sum_{i=1}^{n} R_i^*|^p}{(n^{1+\alpha})^p} \leq n^{1-(1+\alpha)p} C_0\, \mathrm{E}^* \left| \sum_{i=1}^{n} (R_i^*)^2 \right|^{p/2}. \quad \text{(A3)}$$

It follows by Holder's inequality that for constants $a_1, a_2, \dots, a_n$ and $q > 1$,

$$\left(\sum_{i=1}^{n} |a_i| \right)^q \leq n^{q-1} \sum_{i=1}^{n} |a_i|^q.$$

Applying Holder's inequality to (A3) yields that

$$\mathrm{P}' \leq C_0 n^{1-(1+\alpha)p}\, \mathrm{E}^* \left[n^{p/2-1} \sum_{i=1}^{n} |R_i^*|^p \right] = C_0 n^{1-p(1/2+\alpha)}\, \mathrm{E}^* |R_1^*|^p. \quad \text{(A4)}$$

Using equation (A4) it follows that

$$\mathrm{P}\left[n\,\mathrm{P}^* \left(|\bar{R}^*| > n^{\alpha} \right) > \delta \right] \leq \mathrm{P}\left[C_0 n^{1-p(1/2+\alpha)}\, \mathrm{E}^* |R_1^*| > \delta \right]. \quad \text{(A5)}$$

The result follows upon application of Lemma 2 to the exprssion on the right hand side of equation (A5). ∎

Lemma 4 *Let Assumptions 1–5 hold. Let Z^* be a random sample from the distribution $\hat{F}$. For $\beta \in B$ define $H(z,\beta) = g(z,\beta)g(z,\beta)'$. For any $\delta > 0$ and $r > 0$,*

$$\lim_{n\to\infty} n\,\mathrm{P}\left[n\,\mathrm{P}^*\left(\sup_{\beta\in B} n^{-1}\left\|\sum_{i=1}^{n} g(Z_i^*,\beta) - \mathrm{E}^*\,g(Z,\beta)\right\| > r\right) > \delta\right] = 0 \qquad \text{(A6)}$$

and

$$\lim_{n\to\infty} n\,\mathrm{P}\left[n\,\mathrm{P}^*\left(\sup_{\beta\in B} n^{-1}\left\|\sum_{i=1}^{n}[H(Z_i^*,\beta) - \mathrm{E}^*\,H(Z,\beta)]\right\| > r\right) > \delta\right] = 0. \qquad \text{(A7)}$$

Analogous results hold for the first through third derivatives of G and H with respect to β.

Proof Only (A6) is proved. The proof of (A7) is similar. For $\beta \in B$ define $G(z,\beta) = g(z,\beta) - \mathrm{E}^*\,g(Z,\beta)$. Given $\epsilon > 0$ divide B into subsets $\beta_1, \ldots, \beta_{M(\epsilon)}$ such that $\|\beta_1 - \beta_2\| < \epsilon$ whenever $\beta_1, \beta_2 \in B_i$. Let β_i be a point in B_i for each $i = 1, \ldots, M(\epsilon)$. Then

$$\sup_{\beta\in B} n^{-1}\left\|\sum_{i=1}^{n} G(Z_i^*,\beta)\right\| = \max_j \sup_{\beta\in B_j} n^{-1}\left\|\sum_{i=1}^{n} G(Z_i^*,\beta)\right\|. \qquad \text{(A8)}$$

now define

$$\mathrm{P}' \equiv \mathrm{P}^*\left[\sup_{\beta\in B} n^{-1}\left\|\sum_{i=1}^{n} G(Z_i^*,\beta)\right\| > r\right].$$

Using equation (A8) we bound P' as follows

$$\begin{aligned}
\mathrm{P}' &= \mathrm{P}^*\left[\max_j \sup_{\beta\in B_j} n^{-1}\left\|\sum_{i=1}^{n} G(Z_i^*,\beta)\right\| > r\right] \\
&= \mathrm{P}^*\left[\bigcup_{j=1}^{M(\epsilon)}\left\{\sup_{\beta\in B_j} n^{-1}\left\|\sum_{i=1}^{n} G(Z_i^*,\beta)\right\| > r\right\}\right] \\
&\le \sum_{j=1}^{M(\epsilon)} \mathrm{P}^*\left[\sup_{\beta\in B_j} n^{-1}\left\|\sum_{i=1}^{n} G(Z_i^*,\beta)\right\| > r\right].
\end{aligned}$$

now consider the expression $\|\sum_{i=1}^n G(Z_i^*,\beta)\|$:

$$\left\|\sum_{i=1}^n G(Z_i^*,\beta)\right\| \leq \sum_{i=1}^n \|G(Z_i^*,\beta) - G(Z_i^*,\beta_j)\| + \left\|\sum_{i=1}^n G(Z_i^*,\beta_j)\right\|$$
$$\leq \sum_{i=1}^n \|g(Z_i^*,\beta) - g(Z_i^*,\beta_j)\| + \sum_{i=1}^n \|\mathrm{E}^*[g(Z,\beta) - g(Z,\beta_j)]\|$$
$$+ \left\|\sum_{i=1}^n G(Z_i^*,\beta_j)\right\|. \tag{A9}$$

Recall that our regularity conditions assume the existence of a function $C_g(z)$ such that $\|g(z,\beta_1) - g(z,\beta_2)\| \leq C_g(z)\|\beta_1 - \beta_2\|$ for any $\beta_1, \beta_2 \in B$. Applying this to equation (A9) we see that

$$\left\|\sum_{i=1}^n G(Z_i^*,\beta)\right\| \leq \sum_{i=1}^n \|\beta - \beta_j\| C_g(Z_i^*) + \sum_{i=1}^n \|\beta - \beta_j\| \,\mathrm{E}^*\, C_g(Z) + \left\|\sum_{i=1}^n G(Z_i^*,\beta_j)\right\|$$
$$\leq \epsilon \sum_{i=1}^n [C_g(Z_i^*) + \mathrm{E}^*\, C_g(Z)] + \left\|\sum_{i=1}^n G(Z_i^*,\beta_j)\right\|$$
$$\leq \epsilon \sum_{i=1}^n |C_g(Z_i^*) - \mathrm{E}^*\, C_g(Z)| + 2\epsilon n\, \mathrm{E}^*\, C_g(Z) + \left\|\sum_{i=1}^n G(Z_i^*,\beta_j)\right\|$$

where we note that this last expression does not depend on β. Using the above inequality we conclude that

$$\mathrm{P}^*\left[\sup_{\beta\in B_j} n^{-1}\left\|\sum_{i=1}^n G(Z_i^*,\beta)\right\| > r\right] \leq \mathrm{P}^*\left[(\epsilon/n)\sum_{i=1}^n |C_g(Z_i) - \mathrm{E}^*\, C_g(Z)| > r/3\right]$$
$$+ \mathrm{P}^*\left[2\epsilon\, \mathrm{E}^*\, C_g(Z) > r/3\right] + \mathrm{P}^*\left[n^{-1}\left\|\sum_{i=1}^n G(Z_i^*,\beta)\right\| > r/3\right].$$

With probability $1 - \mathrm{o}(n^{-1})$ there exists some $C < \infty$ such that

$$2\epsilon\, \mathrm{E}^*\, C_g(Z) = 2\epsilon \sum_{i=1}^n \hat{p}_i C_g(Z_i) \leq 2\epsilon C n^{-1} \sum_{i=1}^n C_g(Z_i)$$

so that we may choose ϵ small enought that $2\epsilon\, \mathrm{E}^*\, C_g(Z) < r/3$ with probability $1 - \mathrm{o}(n^{-1})$.

So

$$\begin{aligned}
\mathrm{P}[n\mathrm{P}' > \delta] &\le \mathrm{P}\left[n\sum_{j=1}^{M(\epsilon)} \mathrm{P}^*\left(\sup_{\beta\in B_j} n^{-1}\left\|\sum_{i=1}^{n} G(Z_i^*,\beta)\right\| > r\right) > \delta\right] \\
&\le \mathrm{P}\left[n\sum_{j=1}^{M(\epsilon)} \mathrm{P}^*\left(n^{-1}\left\|\sum_{i=1}^{n} G(Z_i^*,\beta_j)\right\| > r/3\right) > \delta/2\right] + \\
&\quad \mathrm{P}\left[n\sum_{j=1}^{M(\epsilon)} \mathrm{P}^*\left(n^{-1}\sum_{i=1}^{n} |C_g(Z_i^*) - \mathrm{E}^* C_g(Z)| > r/(3\epsilon)\right) > \delta/2\right].
\end{aligned}$$

This last expression, in turn, is no larger than

$$\begin{aligned}
&\sum_{j=1}^{M(\epsilon)} \mathrm{P}\left[n\mathrm{P}^*\left(n^{-1}\left\|\sum_{i=1}^{n} G(Z_i^*,\beta_j)\right\| > r/3\right) > \delta/(2M(\epsilon))\right] \\
&\qquad + \sum_{j=1}^{M(\epsilon)} \mathrm{P}\left[n\mathrm{P}^*\left(n^{-1}\sum_{i=1}^{n} |C_g(Z_i^*) - \mathrm{E}^* C_g(Z)| > r/(3\epsilon)\right) > \delta/(2M(\epsilon))\right].
\end{aligned}$$

Applying Lemma 3 to the two terms of this sum yields the desired result. ■

Lemma 5 *Let Assumptions 1–5 hold. Let Z_i^* be sampled randomly according to the distribution function $\hat{F}$. Let β^* solve*

$$\min_{\beta\in B} J^*(\beta) = \left[n^{-1}\sum_{i=1}^{n} g(Z_i^*,\beta)\right]' \Omega \left[n^{-1}\sum_{i=1}^{n} g(Z_i^*,\beta)\right]. \tag{A10}$$

For any $\delta > 0$, $r > 0$ and all sufficiently small η

$$\lim_{n\to\infty} n\mathrm{P}[n\mathrm{P}^*(\|\beta^* - \hat{\beta}\| > rn^{-(1+\eta)/3}) > \delta] = 0. \tag{A11}$$

Proof We first show that

$$\lim_{n\to\infty} n\mathrm{P}[n\mathrm{P}^*(\|\beta^* - \hat{\beta}\| > r) > \delta] = 0.$$

To do this, define $\Delta^*(\beta) \equiv n^{-1}\sum_{i=1}^n [g(Z_i^*,\beta) - \mathrm{E}^* g(Z,\beta)]$. Rearrangement of

$$[\Delta^*(\beta)]'\Omega[\Delta^*(\beta)] = \left(n^{-1}\sum_{i=1}^n [g(Z_i^*,\beta) - \mathrm{E}^* g(Z,\beta)]\right)'\Omega\left(n^{-1}\sum_{i=1}^n [g(Z_i^*,\beta) - \mathrm{E}^* g(Z,\beta)]\right)$$

yields that

$$J^*(\beta) = [\mathrm{E}^* g(Z,\beta)]'\Omega[\mathrm{E}^* g(Z,\beta)] + 2[\mathrm{E}^* g(Z,\beta)]'\Omega[\Delta^*(\beta)] + [\Delta^*(\beta)]'\Omega[\Delta^*(\beta)].$$

where $J^*(\beta)$ is as defined in the statement of the lemma. Given any $\epsilon > 0$ it follows from Lemma 4 that

$$\mathrm{P}\left[\mathrm{P}^*\left(\sup_{\beta\in B}\|\Delta^*(\beta)\| > \epsilon\right) > \delta\right] = \mathrm{o}(n^{-1}).$$

Moreover, $\|\mathrm{E}^* g(Z,\beta)\|$ is bounded uniformly over $\beta \in B$ with probability $1 - \mathrm{o}(n^{-1})$. Therefore, given any $\delta_1, \delta_2 > 0$

$$\lim_{n\to\infty} n\,\mathrm{P}\left[n\,\mathrm{P}^*\left(\sup_{\beta\in B}|2[\mathrm{E}^* g(Z,\beta)]'\Omega[\Delta^*(\beta)] + [\Delta^*(\beta)]'\Omega[\Delta^*(\beta)]| > \delta_1\right) > \delta_2\right] = 0.$$

Define $M = \inf_\Gamma \mathrm{E}^* g(Z,\beta)'\Omega\,\mathrm{E}^* g(Z,\beta)$ where $\Gamma = \{\beta \in B : \|\beta - \hat\beta\| > r\}$. Set $\delta < M/2$. Then

$$\begin{aligned} J^*(\beta) - J^*(\hat\beta) = {} & [\mathrm{E}^* g(Z,\beta)]'\Omega[\mathrm{E}^* g(Z,\beta)] + 2[\mathrm{E}^* g(Z,\beta)]\Omega[\Delta^*(\beta)] + [\Delta^*(\beta)]'\Omega[\Delta^*(\beta)] - \\ & [\mathrm{E}^* g(Z,\hat\beta)]'\Omega[\mathrm{E}^* g(Z,\hat\beta)] - 2[\mathrm{E}^* g(Z,\hat\beta)]\Omega[\Delta(\hat\beta)] - [\Delta^*(\hat\beta)]'\Omega[\Delta^*(\hat\beta)] \\ & > M - 2\delta > 0 \end{aligned} \tag{A12}$$

uniformly over Γ with P^* probability $1 - \mathrm{o}(n^{-1})$ except possibly in a set of P probability $\mathrm{o}(n^{-1})$. By the definition of β^*, $J^*(\beta^*) \le J^*(\hat\beta)$ which is inconsistent with (A12). Therefore $n\,\mathrm{P}[n\,\mathrm{P}^*(\|\beta^* - \hat\beta\| > r) > \delta] = \mathrm{o}(1)$ as $n \to \infty$.

It now follows that $\partial J^*(\beta^*)/\partial(\beta) = 0$ with probability $1 - \mathrm{o}(n^{-1})$ except possibly if χ is in a set of P probability $\mathrm{o}(n^{-1})$. Let $\delta^* = \beta^* - \hat\beta$, δ_i^* denote the ith component of δ^*, $J^*_{\beta ij}(\beta) = \partial^3 J^*(\beta)/\partial\beta\partial\beta_i\beta_j$, and $J^*_{\beta ijk} = \partial^4 J^*(\beta)/\partial\beta\partial\beta_i\beta_j\beta_k$. Using the convention of summing over common subscripts, a Taylor series expansion of $\partial J^*(\beta^*)/\partial\beta$ about $\beta^* = \hat\beta$

yields

$$\partial J^*(\hat\beta)/\partial\beta + \left[\partial^2 J^*(\hat\beta)/\partial\beta\partial\beta'\right]\delta^* + (1/2)J^*_{\beta ij}(\hat\beta)\delta_i^*\delta_j^* + (1/6)J^*_{\beta ijk}(\hat\beta)\delta_i^*\delta_j^*\delta_k^* + \xi^* = 0 \tag{A13}$$

with P^* probability $1-o(n^{-1})$ except possibly if χ is in a set of P probability $o(n^{-1})$, where $\bar\beta$ is between β^* and $\hat\beta$, and

$$\xi^* = (1/6)\left[J^*_{\beta ijk}(\bar\beta) - J^*_{\beta ijk}(\hat\beta)\right]\delta_i^*\delta_j^*\delta_k^*.$$

Let $V^* = [\partial^2 J^*(\hat\beta)/\partial\beta\partial\beta']^{-1}$. V^* exists and $\|V^*\|$ is bounded with P^* probability $1-o(n^{-1})$ except possibly in a set of P probability $o(n^{-1})$ by Lemma 4. Therefore, with P^* probability $1-o(n^{-1})$ except possibly if χ is in a set of P probability $o(n^{-1})$

$$\left(\beta^* - \hat\beta\right) = -V^*\left[\partial J^*(\hat\beta)/\partial\beta + (1/2)J^*_{\beta ij}(\hat\beta)\delta_i^*\delta_j^* + (1/6)J^*_{\beta ijk}(\hat\beta)\delta_i^*\delta_j^*\delta_k^* + \xi^*\right]. \tag{A14}$$

Application of Lemma 2 of Hall and Horowitz (1996) and Lemma 4 shows that with P^* probability $1-o(n^{-1})$ except possibly if χ is in a set of P probability $o(n^{-1})$, $\|\partial J^*(\hat\beta)/\partial\beta\| < n^{-(1+\eta)/3}$ for any sufficiently small $\eta > 0$, $\|J^*_{\beta ij}(\hat\beta)\|$ and $\|J^*_{\beta ijk}(\hat\beta)\|$ are bounded, and $\|\xi^*\| < M\|\beta^* - \hat\beta\|^4$ for some $M < \infty$. Therefore, with P^* probability $1 - o(n^{-1})$ except possibly if χ is in a set of P probability $o(n^{-1})$, the norm of the right-hand side of (A14) is less than $rn^{-(1+\eta)/3}$ whenever $\|\beta^* - \hat\beta\| \le rn^{-(1+\eta)/3}$ and n is sufficiently large. Application of the Brouwer fixed point theorem to the right-hand side of (A14) establishes (A11). ∎

Lemma 6 *Let Assumptions 1-5 hold. Let Z be sampled randomly according to the distribution function $\hat F$. Let β^* solve*

$$\min_{\beta\in B} J^*(\beta,\tilde\beta^*) = \left[n^{-1}\sum_{i=1}^n g(Z_i^*,\beta)\right]'\Omega^*(\tilde\beta^*)\left[n^{-1}\sum_{i=1}^n g(Z_i^*,\beta)\right] \tag{A15}$$

where $\tilde\beta^*$ solves (A10). For any $\delta > 0$, $r > 0$ and all sufficiently small η

$$\lim_{n\to\infty} n\,\mathrm{P}[n\,\mathrm{P}^*(\|\beta^* - \hat\beta\| > rn^{-(1+\eta)/3}) > \delta] = 0.$$

Proof The proof uses arguments similar to the proof of Lemma 4 of Hall and Horowitz (1996) with results from that paper replaced by the analogous results for the moment-restricted bootstrap found in this appendix. ■

Proposition 1 *Let Assumptions 1-7 hold, and let $\hat{\beta}$ and β^*, respectively, solve either*

$$\min_{\beta \in B} J(\beta) \equiv \left[n^{-1} \sum_{i=1}^{n} g(Z_i, \beta) \right]' \Omega \left[n^{-1} \sum_{i=1}^{n} g(Z_i, \beta) \right] \tag{A16}$$

and (A10) or

$$\min_{\beta \in B} J(\beta, \tilde{\beta}) \equiv \left[n^{-1} \sum_{i=1}^{n} g(Z_i, \beta) \right]' \Omega(\tilde{\beta}) \left[n^{-1} \sum_{i=1}^{n} g(Z_i, \beta) \right] \tag{A17}$$

(where $\tilde{\beta}$ solves equation (A16)) and (A15). Let Δ^* denote either t_j^* or (if $\hat{\beta}$ and β^* solve (A17) and (A15)) $K^*(\beta^*)$, where

$$K^*(\beta) = \Omega^{-1/2}(\tilde{\beta}^*) \left[n^{-1} \sum_{i=1}^{n} g(Z_i, \beta) \right].$$

Let $f^*(Z_i^*, \beta)$ be a vector containing the unique components of $g(Z_i^*, \beta)$, $g(Z_i^*, \beta) g(Z_i^*, \beta)'$, and their derivative through order 4 with respect to the compenents of β. Define $S^* = n^{-1} \sum_{i=1}^{n} f^*(Z_i^*, \hat{\beta})$. For each definition of Δ^* there is a function G that is continuously differentiable in a neighborhood of S^* such that $\Delta^* = n^{1/2} G(S^*) + \mathrm{o}(n^{-1})$ with P^* probability $\mathrm{o}(n^{-1})$ except, possibly, if χ is in a set of P probability $\mathrm{o}(n^{-1})$.

Proof The proof uses arguments similar to the proof of Proposition 2 of Hall and Horowitz (1996) with results from that paper replaced by the analogous results for the moment-restricted bootstrap found in this appendix. ■

Theorem 1 *Let Assumptions 1–3 hold. Then with probability approaching one $\hat{F}(z) = \sum_{i=1}^{n} \hat{p}_i \mathrm{I}(z \leq z_i)$ exists and attains the semiparametric efficiency bound for estimation of the distribution of a single observation in the model $\mathrm{E}[g(Z, \beta_0)] = 0$.*

Proof Let $\hat{Q}(\lambda) = n^{-1} \sum_{i=1}^{n} T(\lambda' \hat{g}_i)$ and $\tilde{\lambda} = \operatorname{argmax} \hat{Q}(\lambda)$ subject to the restriction that

$\lambda \in \hat{\Lambda} = \{\lambda : |\lambda' \hat{g}_i| \leq \bar{v}, i = 1, \ldots, n\}$. It follows by Theorem 3.4 of newey and McFadden (1994) that

$$n^{1/2}(\hat{\beta} - \beta_0) = -\left(G'V^{-1}G\right)^{-1} G'V^{-1} n^{1/2} \hat{g}_n(\beta_0) + o_p(1) = O_p(1)$$

where $G = \mathrm{E}[\partial g(Z, \beta_0)/\partial \beta]$ and $\hat{g}_n(\beta) = n^{-1} \sum_{i=1}^{n} g(Z_i, \beta)$. Also, by a standard mean value expansion

$$n^{1/2}\bar{g} = \left[\mathrm{I} - G\left(G'V^{-1}G\right)^{-1} G'V^{-1}\right] n^{1/2} \hat{g}(\beta_0) + o_p(1).$$

Therefore, $\partial \hat{Q}(0)/\partial \lambda = \nabla T(0)\bar{g} = O_p(n^{-1/2})$. Also, as shown in section 4 of Newey and McFadden, $n^{-1} \sum_{i=1}^{n} \hat{g}_i \hat{g}_i' \xrightarrow{p} V$, so by non-singularity of V and $\nabla^2 T(V)$ bounded away from zero, with probability approaching one the smallest eigenvalue of

$$\frac{\partial^2 \hat{Q}(\lambda)}{\partial \lambda \partial \lambda'} = n^{-1} \sum_{i=1}^{n} \nabla^2 T(\lambda' \hat{g}_i) \hat{g}_i \hat{g}_i'$$

is bounded away from zero, uniformly in $\lambda \in \hat{\Lambda}$ with probability approaching one. Therefore, by a mean value expansion, with probability approaching one,

$$\hat{Q}(0) \leq \hat{Q}(\tilde{\lambda}) = \hat{Q}(0) + \left[\frac{\partial \hat{Q}(0)}{\partial \lambda}\right]' \tilde{\lambda} + \tilde{\lambda}' \left[\frac{\partial^2 \hat{Q}(\bar{\lambda})}{\partial \lambda \partial \lambda'}\right] \tilde{\lambda} \leq \hat{Q}(0) + \left\| \frac{\partial \hat{Q}(0)}{\partial \lambda} \right\| \|\tilde{\lambda}\| - C \|\tilde{\lambda}\|^2$$

where $\bar{\lambda}$ lies on a line joining $\tilde{\lambda}$ and 0 and hence is in $\hat{\Lambda}$ by convexity of $\hat{\Lambda}$. Subtracting and dividing through by $\|\tilde{\lambda}\|$ gives

$$\|\tilde{\lambda}\| \leq C^{-1} \left\| \frac{\partial \hat{Q}(0)}{\partial \lambda} \right\| = O_p(n^{-1/2}).$$

Furthermore, for $y_i = \max_{\beta \in \mathcal{N}} \|g(z_i, \beta)\|$, with probability approaching one,

$$\max_{i \leq n} \|\hat{g}_i\| \leq \max_{i \leq n} y_i \leq \left(\sum_{i=1}^{n} y_i^p\right)^{1/p} = O_p\left([n \, \mathrm{E}\, y_i^p]^{1/p}\right) = O_p(n^{1/p}) = o_p(n^{1/2})$$

so that

$$\max_{i \leq n} |\tilde{\lambda}' \hat{g}_i| \leq \|\tilde{\lambda}\| \max_{i \leq n} \|\hat{g}_i\| = O_p(n^{-1/2})\, o_p(n^{1/2}) = o_p(1).$$

It follows that $\hat{\lambda}$ is an element of the interior $\hat{\Lambda}$ with probability approaching one, and hence is a unique interior maximum of the globally concave function $\hat{Q}(\lambda)$, implying $\hat{\lambda} = \tilde{\lambda}$. Also, $\partial^2\hat{Q}(0)/\partial\lambda\partial\lambda' \xrightarrow{p} \nabla^2 T(0)V$ and for any $\bar{\lambda} \xrightarrow{p} 0$,

$$\left\| \frac{\partial^2\hat{Q}(\bar{\lambda})}{\partial\lambda\partial\lambda'} - \frac{\partial^2\hat{Q}(0)}{\partial\lambda\partial\lambda'} \right\| \leq \max_{i\leq n} \left|\nabla^2 T(\bar{\lambda}'\hat{g}_i) - \nabla^2 T(0)\right| n^{-1}\sum_{i=1}^{n} y_i^2 = o_p(1)\, O_p(1) = o_p(1),$$

so that $\partial^2\hat{Q}(\bar{\lambda})/\partial\lambda\partial\lambda' \xrightarrow{p} \nabla^2 T(0)V$. Since $\hat{\lambda}$ is a global maximum, the first order conditions $\partial\hat{Q}(\hat{\lambda})/\partial\lambda = 0$ are satisfied, and can be expanded around zero to obtain

$$n^{1/2}\hat{\lambda} = -\left[\frac{\partial^2\hat{Q}(\bar{\lambda})}{\partial\lambda\partial\lambda'}\right] n^{1/2}\frac{\partial\hat{Q}(0)}{\partial\lambda} = -\left[\frac{\nabla T(0)}{\nabla^2 T(0)}\right] V^{-1} n^{1/2}\bar{g} + o_p(1).$$

Then a mean-value expansion gives

$$\left| n^{-1}\sum_{i=1}^{n} \nabla T(\hat{\lambda}'\hat{g}_i) - \nabla T(0) \right| \leq$$
$$\|\hat{\lambda}\| \max_{i\leq n} \left|\nabla^2 T(\bar{\lambda}'\hat{g}_i) - \nabla^2 T(0)\right| n^{-1}\sum_{i=1}^{n} y_i + \left|\nabla^2 T(0)\right| \|\hat{\lambda}\|\|\bar{g}\| = o_p(n^{-1/2}).$$

Also, by a mean-value expansion, for $\mathrm{I}_i = \mathrm{I}(z \leq z_i)$ and $H = \mathrm{E}[\mathrm{I}_i\, g(Z_i, \beta_0)]$,

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} \nabla T(\hat{\lambda}'\hat{g}_i)\, \mathrm{I}_i &= \nabla T(0) n^{-1}\sum_{i=1}^{n} \mathrm{I}_i + \left[n^{-1}\sum_{i=1}^{n} \nabla^2 T(\bar{\lambda}'\hat{g}_i)\, \mathrm{I}_i\, \hat{g}_i\right]' \hat{\lambda} \\ &= \nabla T(0) n^{-1}\sum_{i=1}^{n} \mathrm{I}_i + \nabla^2 T(0) H'\hat{\lambda} + o_p(n^{-1/2}). \end{aligned}$$

It then follows that

$$\begin{aligned}
n^{1/2}[\hat{F}(z) - F(z)] &= n^{-1/2}\sum_{i=1}^{n}[\mathrm{I}_i - F(z)] + \left[\frac{\nabla^2 T(0)}{\nabla T(0)}\right] H' n^{1/2}\hat{\lambda} + \mathrm{o}_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}[\mathrm{I}_i - F(z)] - H'V^{-1}n^{1/2}\bar{g} + \mathrm{o}_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}\left\{\mathrm{I}_i - F(z) - H'V^{-1}[I - G(G'V^{-1}G)^{-1}G'V^{-1}]g(z_i,\beta_0)\right\} \\
&\quad + \mathrm{o}_p(1).
\end{aligned}$$

Thus, the asymptotic variance of this estimator is

$$F(z)[1 - F(z)] - H'(V^{-1} - V^{-1}G(G'V^{-1}G)^{-1}G'V^{-1})H,$$

which is equal to the semiparametric efficiency bound given in Brown and Newey (1994). ■

Proposition 2 *Let Assumptions 1–7 hold. Let ν^* be a vector of bootstrap moments of the form*

$$\mathrm{E}^*\, n^{\alpha(m)} \prod_{k=1}^{m} \Psi_{j_k}^* \tag{A18}$$

where $2 \le m \le 6$, $\alpha(m) = 0$ if m is even and $1/2$ if m is odd, and $\Psi_{j_k}^*$ is the j_k component of $\Psi^* = n^{1/2}[S^* - \mathrm{E}^* S^*]$. Then, except possibly, if χ is contained in a set of probability $\mathrm{o}(n^{-1})$

$$\sup_z \left| \mathrm{P}^*(t_j^* \le z) - \left[1 + \sum_{i=1}^{2} n^{-i/2}\pi_i(\delta,\nu^*)\right]\Phi(z)\right| = \mathrm{o}(n^{-1}).$$

where $\delta = d/dz$, π_i $(i = 1,2)$ is a polynomial in δ whose coefficients are continuous functions of ν^*, $\pi_1(\delta,\nu^*)\Phi(z)$ is an even function of z, and $\pi_2(\delta,\nu^*)\Phi(z)$ is an odd function of z. In addition

$$\sup_z \left| \mathrm{P}^*(J^* < z) - \int_{-\infty}^{z} d\left\{\left[1 + n^{-1}\pi_J(\xi,\nu^*)\right]\mathrm{P}(\chi_\lambda^2 \le \xi)\right\}\right| = \mathrm{o}(n^{-1})$$

where χ^2_λ is a random variable that has the chi-square distribution with $\lambda \equiv r - q$ degrees of freedom, and $\pi_J(\xi, \nu^*)$ is a polynomial function of χ whose coefficients are continuous functions of ν^*.

Proof The proof of this proposition should be similar to the proof of Theorem 2 of Hall and Horowitz (1996) or Theorem 5.1 of Hall (1992), and so is omitted. ■

Lemma 7 *Let Assumptions 1-5 hold. Let $f(Z_i, \beta)$ be a vector containing the unique components of $g(Z_i, \beta)$, $g(Z_i, \beta)g(Z_i, \beta)'$, and their derivatives through order four with respect to the components of β. Let ν denote a vector of the form $\lim_{n\to\infty} \mathrm{E}\, n^{\alpha(m)} \prod_{k=1}^{m} \Psi_{j_k}$ where Ψ_{j_k} denotes the j_k component of*

$$\Psi = n^{1/2} \left[n^{-1} \sum_{i=1}^{n} f(Z_i, \beta_0) - \mathrm{E}\, f(Z, \beta_0) \right]$$

and $\nu^*_{m\{j\}}$ denote the quantity in (A18). Then for each integer m such that $2 \le m \le 6$,

$$\operatorname*{plim}_{n\to\infty} \nu^*_{m\{j\}} = \nu_{m\{j\}}.$$

Proof The proof of this lemma is similar to the proof of Lemma 11 of Hall and Horowitz (1996) and so is omitted. ■

Theorem 2 *Let Assumptions 1-7 hold. Under $H_0 : \beta_j = \beta_{0j}$*

$$\mathrm{P}(|t_j| > z^*_{\alpha j}) = \alpha + o(n^{-1}).$$

If $r > q$ and the overidentifying restrictions hold,

$$\mathrm{P}(J > J^*_\alpha) = \alpha + o(n^{-1}).$$

Proof By Theorem 1 of Hall and Horowitz (1996),

$$\sup_z \left| \mathrm{P}(t_j \le z) - \left[1 + \sum_{i=1}^{2} n^{-i/2} \pi_i(\delta, \nu) \right] \Phi(z) \right| = o(n^{-1}),$$

and

$$\sup_z \left| \mathrm{P}(J < z) - \int_{-\infty}^{z} d\left\{ \left[1 + n^{-1}\pi_J(\xi,\nu)\right] \mathrm{P}(\chi^2_\lambda \le \xi) \right\} \right| = o(n^{-1}).$$

Using this result with Proposition 2 and Lemma 7, and noting that $\pi_1(\delta,\nu)\Phi(z)$ and $\pi_1(\delta,\nu^*)\Phi(z)$ are even functions of z, it follows that

$$\mathrm{P}(|t_j| > z) - \mathrm{P}^*(|t_j^*| > z) = o_p(n^{-1}) \tag{A19}$$

uniformly over z and

$$\mathrm{P}(J > z) - \mathrm{P}^*(J^* > z) = o_p(n^{-1}). \tag{A20}$$

Choosing $z = z^*_{\alpha j}$ or $z = J^*_\alpha$ in equations (A19) and (A20) produces the desired result. ∎

Bibliography

[1] Ahn, Seung C. and Peter Schmidt. "Efficient Estimation of Models for Dynamic Panel Data." *Journal of Econometrics* 68 (1994).

[2] Anderson, T.W. and C. Hsaio. "Formulation and Estimation of Dynamic Models using Panel Data." *Journal of Econometrics* 18 (1982).

[3] Back, Kerry and David P. Brown. "Implied Probabilities in GMM Estimators." *Econometrica* 61 (1993).

[4] Becker, Gary S. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76 (1968).

[5] Bekker, Paul A. "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica* 62 (1994).

[6] Beran, Rudolf. "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements." *Journal of the American Statistical Association* 83 (1988).

[7] Blundell, R., S. Bond, M. Devereux, and F. Schiantarelli. "Investment and Tobin's Q: Evidence from Company Panel Data." *Journal of Econometrics* 51 (1992).

[8] Bound, John and Richard B. Freeman. "What Went Wrong? The Erosion of Relative Earnings Among Young Black Men in the 1980s." *Quarterly Journal of Economics* 107 (1992).

[9] Brewin, Chris R., Bernice Andrews, Suzanna Rose, and Marilyn Kirk. "Acute Stress Disorder and Posttraumatic Stress Disorder in Victims of Violent Crime." *American Journal of Psychiatry* 156 (1999).

[10] Brown, Bryan W. and Whitney K. Newey. "Bootstrapping for GMM." MIT Department of Economics Working paper, 1992.

[11] Brown, Bryan W. and Whitney K. Newey. "Efficient Semiparametric Estimation of Expectations." MIT Department of Economics Working Paper, 1994.

[12] Card, David. "Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage." *Industrial and Labor Relations Review* 46 (1992).

[13] Card, David and Alan B. Krueger. *Myth and Measurement: The New Economics of the Minimum Wage.* Princeton, NJ: Princeton University Press, 1995.

[14] Cohen, Mark. "Pain, Suffering, and Jury Awards: A Study of the Cost of Crime to Victims." *Law and Society Review* 22 (1988).

[15] Cullen, Julie and Steven Levitt. "Crime, Urban Flight, and the Consequences for Cities." Working Paper No. 5737. Cambridge, MA: National Bureau of Economic Research, 1996.

[16] Cunningham, William C., John J. Strauchs, and Clifford W. Van Meter. "Private Security: Patterns and Trends." Research in Brief. Washington, DC: U. S. Department of Justice, National Institute of Justice, 1991.

[17] Dugan, Laura. "The Effect of Criminal Victimization on a Household's Moving Decision." *Criminology* 37 (1999).

[18] Feldstein, Martin. "Tax Avoidance and the Deadweight Loss of the Income Tax." Working Paper No. 5055. Cambridge, MA: National Bureau of Economic Research, 1995.

[19] Freeman, Richard B. "Crime and the Employment of Disadvantaged Youths." Working Paper No. 3875. Cambridge, MA: National Bureau of Economic Research, 1991.

[20] Freeman, Richard B. "Why Do So Many Young American Men Commit Crimes and What Might We Do About it?" *Journal of Economic Perspectives* 10 (1996).

[21] Freeman, Richard B. "The Economics of Crime" in *Handbook of Labor Economics, Volume 3*, O. Ashenfelter and D. Card, eds. New York, NY: Elsevier Science, 1999 (forthcoming).

[22] Freeman, Richard B. and William M. Rodgers III. "Area Economic Conditions and the Labor Market Outcomes of Young Men in the 1990s Expansion." Working Paper No. 7073. Cambridge, MA: National Bureau of Economic Research, 1999.

[23] Grogger, Jeff. "The Effect of Arrests on the Employment and Earnings of Young Men." *Quarterly Journal of Economics* 110 (1995).

[24] Grogger, Jeff. "Market Wages and Youth Crime." *Journal of Labor Economics* 16 (1998).

[25] Hahn, Jinyong. "A Note on Bootstrapping Generalized Method of Moments Estimators." *Econometric Theory* 12 (1996).

[26] Hall, Peter. *The Bootstrap and Edgeworth Expansion*. New York, NY: Springer-Verlag, 1992.

[27] Hall, Peter and Joel Horowitz. "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators." *Econometrica* 64 (1996).

[28] Hamermesh, Daniel S.. *Labor Demand*. Princeton, NJ: Princeton University Press, 1993.

[29] Hansen, Lars Peter. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 (1982).

[30] Holtz-Eakin, Douglas, Whitney K. Newey, and Harvey S. Rosen. "Estimating Vector Autoregressions with Panel Data." *Econometrica* 56 (1988).

[31] Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. "Wage Inequality and the Rise in the Returns to Skill." *Journal of Political Economy* 101 (1993).

[32] Juhn, Chinhui, Kevin M. Murphy, and Robert H. Topel. "Why Has the Natural Rate of Unemployment Increased Over Time?" *Brookings Papers on Economic Activities* 2 (1991).

[33] Katz, Lawrence F., and Kevin M. Murphy. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107 (1992).

[34] Killingsworth, Mark R.. *Labor Supply.* New York, NY: Cambridge University Press, 1983.

[35] Kling, Jeffrey R. "The Effect of Prison Sentence Length on the Subsequent Employment and Earnings of Criminal Defendants." Woodrow Wilson School Economics Discussion Paper No. 208 (February), 1999.

[36] Levitt, Steven D. "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation." Working Paper No. 5119. Cambridge, MA: National Bureau of Economic Research, 1995.

[37] Levitt, Steven D. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime." *American Economic Review* 87 (1997).

[38] Levitt, Steven D. "Why Do Increased Arrest Rates Appear to Reduce Crime: Deterrence, Incapacitation, or Measurement Error?" *Economic Inquiry* 36 (1998).

[39] Maasoumi, Esfandiar and P.C.B. Phillips. "On the Distribution of Instrumental Variables Estimators." *Journal of Econometrics* (1982).

[40] Michaels, Andrew J., Claire E. Michaels, Christina H. Moon, M.A. Zimmerman, Christopher Peterson, and J.L. Rodriguez. "Psychosocial Factors Limit Outcomes after Trauma." *Journal of Trauma* 44 (1998).

[41] Miethe, Terance D. and Mark C. Stafford. "Social Differentiation in Criminal Victimization: A Test of Routine Activities/Lifestyle Theories." *American Sociological Review* 52 (1987).

[42] Miller, Ted R., Mark A. Cohen, and Shelli B. Rossman. "Victim Costs of Violent Crime and Resulting Injuries." *Health Affairs* 12 (1993).

[43] Newey, Whitney K. and Daniel McFadden'. "Large Sample Estimation and Hypothesis Testing" in Handbook of Econometrics, R. Engle and D. McFadden, eds. Amsterdam: North-Holland, 1994.

[44] Nickell, Stephen. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49 (1981).

[45] Norris, Fran H. and Krzysztof F. Kaniasty. "The Psychological Experience of Crime: A Test of the Mediating Role of Beliefs in Explaining the Distress of Victims." *Journal of Social and Clinical Psychology* 10 (1991).

[46] Norström, Thor. "Theft Criminality and Economic Growth." *Social Science Research* 19 (1988).

[47] Owen, Art. "Empirical Likelihood Ratio Confidence Regions." *Annals of Statistics* 18 (1990).

[48] Pathe, Michele and Paul E. Mullen. "The Impact of Stalkers on their Victims." *British Journal of Psychiatry* 170 (1997).

[49] Perkins, Craig A., Patsy A. Klaus, Lisa D. Bastian, and Robyn L. Cohen. *Criminal Victimization in the United States, 1993: A National Crime Victimization Survey Report.* Washington, DC: U. S. Department of Justice, 1996.

[50] Piehl, Anne and John DiIulio. "Does Prison Pay? Revisited." *The Brookings Review* Winter (1995).

[51] Reuter, Peter, Robert MacCoun, and Patrick Murphy. *Money from Crime: A Study of the Economics of Drug Dealing in Washington, D.C.* Santa Monica, CA: The RAND Corporation, 1990.

[52] Rosenbaum, Paul R. and Donald B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1983).

[53] Rosenbaum, Paul R. and Donald B. Rubin. "Reducing Bias in Observational Studies using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79 (1984).

[54] Rosenbaum, Paul R. and Donald B. Rubin. "Constructing a Control Group using Multivariate Matching Methods that include the Propensity Score." *American Statistician* 39 (1985).

[55] Sargan, J.D. "Some Approximations to the Distribution of Econometric Criteria which are Asymptotically Distributed as Chi-Squared." *Econometrica* 48 (1980).

[56] Schmidt, Peter and Ann D. Witte. *An Economic Analysis of Crime and Justice: Theory, Methods, and Applications.* Orlando, FL: Academic Press, 1984.

[57] Schutzwohl, Matthias and Andreas Maercker. "Posttraumatic Stress Disorder Due to Criminal Victimization." *Zeitschrift fur Klinische Pscyhologie* 26 (1997).

[58] Schwartz, Joe and Thomas Exter. "Crime Stoppers." *American Demographics* 12 (1990).

[59] Staiger, Douglas and James H. Stock. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (1997).

[60] Tendler, Steward. "Report Predicts 40% Increase in Property Crime." *The Times (London)* (June 7, 1999).

[61] U.S. Dept. of Justice, Bureau of Justice Statistics. *Crime in the United States.* Washington, D.C.: U.S. Department of Justice, various years.

[62] U. S. Dept. of Justice, Bureau of Justice Statistics. *National Crime Surveys: National Sample, 1986-1991* [Computer file]. 5th ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 1993.

[63] U.S. Dept. of Justice, Federal Bureau of Investigation. *Uniform Crime Reporting Handbook.* Washington, D.C.: U.S. Government Printing Office, 1980.

[64] U.S. Dept. of Justice, Federal Bureau of Investigation. *Uniform Crime Reports: County Level Arrest and Offense Data,* 1986 [Computer file]. 2nd ICPSR ed. Ann Arbor, Mich.: Inter-university Consortium for Political and Social Research [producer and distributor], 1998.

[65] U.S. Dept. of Justice, Federal Bureau of Investigation. *Uniform Crime Reports: County Level Arrest and Offense Data,* 1989 [Computer file]. ICPSR ed. Washington, DC: U.S. Dept. of Justice, Federal Bureau of Investigation [producer], 1989. Ann Arbor, Mich.: Inter-university Consortium for Political and Social Research [distributor], 1997.

[66] U.S. Dept. of Justice, Federal Bureau of Investigation. *Uniform Crime Reports: County-Level Detailed Arrest and Offense Data,* 1992 [Computer file]. ICPSR ed. Ann Arbor, Mich.: Inter-university Consortium for Political and Social Research [producer and distributor], 1994.

[67] Viscusi, W. Kip. *Employment Hazards: And Investigation of Market Performance.* Cambridge, MA: Harvard University Press, 1979.

[68] Viscusi, W. Kip. "Occupational Safety and Health Regulations: Its Impact and Policy Alternatives" in *Research in Public Policy Analysis and Management,* Volume 2, J. Crecine, ed. Greenwich, CT: JAI Press, 1981.

[69] Viscusi, W. Kip and Michael J. Moore. "Workers' Compensation: Wage Effects, Benefit Inadequacies, and the Value of Health Losses." *The Review of Economics and Statistics* 69 (1987).

[70] Willis, Michael. "The Relationship between Crime and Jobs." University of California, Santa Barbara Working Paper, 1997.

THESIS PROCESSING SLIP

FIXED FIELD: ill. ______ name ______

index ______ biblio ______

▶ COPIES: (Archives) Aero (Dewey) Eng Hum

Lindgren Music Rotch Science

TITLE VARIES: ▶☐ ______

NAME VARIES: ▶☑ Michael

IMPRINT: (COPYRIGHT) ______

▶ COLLATION: ______

▶ ADD: DEGREE: ______ ▶ DEPT.: ______

SUPERVISORS: ______

NOTES:

cat'r: date:

page: J186

▶ DEPT: Econ

▶ YEAR: 2000 ▶ DEGREE: Ph.D.

▶ NAME: MAY, Sean Mi...