

UC Berkeley

Reports

Title

Establishing Causality in Welfare Research: Theory and Application: Interim Report of the California Welfare Reform Impact Study

Permalink

<https://escholarship.org/uc/item/9154r2d3>

Authors

Brady, Henry E.
Nicosia, Nancy
Seto, Eva Y.

Publication Date

2002-12-01

**ESTABLISHING CAUSALITY IN WELFARE RESEARCH:
THEORY AND APPLICATION**

**Interim Report
California Welfare Reform Impact Study**

**Henry E. Brady
Nancy Nicosia
Eva Y. Seto**

December, 2002

**Prepared for:
California Department of Social Services
and
The Administration for Children and Families
U.S. Department of Health and Human Services**

Points of view or opinions expressed in this document are those of the author(s) and do not necessarily represent the official position or policies of the Regents of the University of California, the California Department of Social Services, or the Administration for Children and Families, U.S. Department of Health and Human Services.

Table of Contents

Introduction

Chapter 1 Causality

1.1 Causality and Counterfactuals	Page 1- 1
1.2 Correlation and Manipulation	Page 1- 5
1.3 Four Theories of Causality	Page 1- 6
1.4 Summary of Four Theories of Causality	Page 1- 20

Chapter 2 Experimental and Non-Experimental Approaches For Determining Causality

2.1 Examples of Welfare Experiments	Page 2- 1
2.2 Assessing Randomized Experiments	Page 2- 2
2.3 Observational Studies	Page 2- 7
2.4 Quasi-Experiments	Page 2- 10
2.5 Natural Experiments	Page 2- 10
2.6 What are the Best Methods for Establishing Causality?	Page 2- 11

Chapter 3 Estimation, Forecasting, and Model Uncertainty

3.1 Estimation, Model Selection, and Hypothesis Testing	Page 3- 1
3.2 Consistency and Efficiency - Specification Error and Violating the Assumptions of OLS	Page 3- 7
3.3 Pooling/Panel Data: Correcting for Specification Error or Lack of Data	Page 3- 14
3.4 Forecasting and Model Checking	Page 3- 19
3.5 Model Uncertainty	Page 3- 23

Chapter 4 Data Description

4.1 Welfare Participation Variables	Page 4- 1
4.2 Employment Variables	Page 4- 8
4.3 County Demographics and Characteristic Variables	Page 4- 10

Chapter 5 Empirical Section: Estimation

5.1 Correlation and Bivariate Analysis	Page 5- 2
5.2 Multivariate Analysis	Page 5- 3
5.3 Panel/Pooled Data Analysis	Page 5- 8
5.4 Notes on Employment Variables and Coefficients	Page 5- 15
5.5 Summary	Page 5- 27

Chapter 6 Unconditional Out-Of-Sample Forecasting

6.1 Time Series Forecasting	Page 6- 1
6.2 Panel Data Forecasting	Page 6- 2
6.3 Varying the Estimation Sample: The Effect on Forecasts	Page 6- 5
6.4 Moving Forecasts and Estimation Sample	Page 6- 10
6.5 Optimal Estimation Lengths for Varying Forecast Years	Page 6- 13
6.6 Calculating the Minimum Required Impact	Page 6- 17
6.7 AMSFE: The Variance-Bias Tradeoff	Page 6- 19
6.8 Summary	Page 6- 20

Conclusions

Appendix

Bibliography

Introduction

The Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) of 1996 launched the greatest changes in American welfare policy since the 1935 Social Security Act created Aid to Dependent Children as a program for widows and their children. Among many other changes, PRWORA put time limits on the receipt of welfare, required that most recipients work, and eliminated the entitlement status of welfare. The “work-first” orientation of PRWORA was partly based upon field experiments undertaken in California that had shown that many welfare recipients could be moved quickly from welfare to work. But many elements of PRWORA, such as time limits, were essentially untested, and their inclusion in the legislation was the consequence of political sentiments rather than solid evidence about their effects. The evaluation of these features of welfare reform, therefore, had to take place after implementation, and the far-reaching nature of these changes has made it virtually impossible to use experimental methods.

This report is about the problems of establishing causal impacts when experimental methods are not available and when “observational data” must be used. These data are typically information about outcome variables such as the welfare caseload or the earnings of welfare recipients and the economic, demographic, and programmatic conditions that affect them. Because outcome variables are usually affected by many different factors, it is very difficult to separate out the impacts of program features from changes in the economy, in birth rates, or other factors that affect welfare outcomes. Statistical and econometric methods, such as regression analysis, have been developed to analyze these kinds of data and to separate out the impacts of each factor. These methods are very sophisticated, and they can be useful when the assumptions underlying them are met. Unfortunately, it is often hard to test these assumptions and to know whether or not they hold. One of the major goals of this report is to provide welfare researchers with a better sense of what these assumptions are and the likelihood that they hold with welfare data.

The work in this report takes advantage of the fact that the state of California has very rich data on 58 different counties. These data include welfare caseloads, welfare starts, welfare terminations, economic conditions, demographic trends, and other features of each county. In future work, we will use even richer data on individuals on welfare in California. In this report, “aggregate” data on California’s 58 counties are used to see how well models of caseloads can predict future caseloads.¹ If this prediction can be done well, then when policy changes are made, the difference between the predicted caseloads under the old policy and the actual caseloads under the new policy can be used as a measure of the impact, or the causal effect, of the new policy.

Roadmap to the Report

The report has three overall parts. The first two chapters discuss general issues about causality that inform the remainder of the report. Those interested in the empirical findings might want to

¹ We actually do not directly predict the number of caseloads. Rather we predict welfare starts and terminations which make it possible to predict future caseloads as the current caseload minus future terminations plus future starts.

skip these chapters. The next chapter discusses how statistical methods try to solve these problems, and those familiar with statistical methods might want to skip this chapter as well. The final three chapters present our empirical results using California data.

These three final chapters explore the feasibility of using county-level aggregate data on entrances and exits to welfare for evaluating the causal impact of welfare reform on welfare caseloads. Our approach relies upon being able to make predictions about what would have happened in the absence of some change. These predictions are sometimes called “counterfactuals” because they are counter to fact statements. They are easily identified because they describe what “would” have happened in the absence of some program changes. Thus, if we can successfully predict welfare caseloads in the future by assuming that present programs and trends continue, then when program changes occur, we can use these counterfactual predictions as statements about what would have happened to the caseload if the program changes had not occurred. Then any differences between the caseload that is predicted based upon past trends and the caseload that is observed after the program changes are implemented can be attributed to the program changes. The program changes can be considered the cause of the differences. Our analysis explores how well we can predict caseloads when there have been no program changes. If we do well in these circumstances, then we can probably assume that we would probably do well when there is a program change. Consequently, we can evaluate the impact of the program change.

This simple approach, however, is fraught with difficulties. The first two chapters on causality set out some of the theoretical limitations. Chapter 1 explores the strengths and weaknesses of the counterfactual definition of causality, and it suggests that we might like to know more in order to make reliable causal statements. This chapter shows that the counterfactual definition has great merit, but it can fail unless care is taken to define the counterfactual properly. The counterfactual approach also neglects two other aspects of good causal inference, the demonstration that the putative cause actually *manipulated* something to cause the outcome and the description of a theoretically compelling *mechanism* that links the cause with the effect. The last three chapters focus on the problem of defining the counterfactual correctly, but they also sometimes mention the importance of establishing manipulation and mechanism.

Chapter 2 shows how experimental methods use control groups to get a handle on counterfactual outcomes, and the chapter demonstrates the difficulties of getting comparable counterfactuals with observational studies. This chapter provides a picture of the strengths and limitations of experimental and observational methods. These general chapters are followed by Chapters 3 through 6 which explore how caseload forecasting methods can be used to evaluate welfare policies in California. Chapter 3 discusses specific statistical methods for estimation, forecasting, and evaluating model uncertainty. These methods are described in standard econometrics textbooks, but we have provided a synopsis of them that might be useful to those who are not familiar with modern time-series, time-series cross-sectional, and other econometric methods.

Chapter 4 describes the aggregate caseload data that is available for California’s 58 counties. Although this chapter is in one sense merely a recounting of the available data, it serves a much more fundamental purpose as well. The data that are available limit the kinds of counterfactual

predictions that can be spun out. If the data omit some important determinants or features of welfare caseloads, then the counterfactual predictions produced with our econometric models will be suspect.

Finally, Chapters 5 and 6 apply these methods to California caseload prediction and forecasting. Chapter 5 estimates a series of increasingly more sophisticated and complex models. Chapter 6 uses the best of these models to make unconditional out-of-sample forecasts of welfare starts and terminations in California. This chapter ends with a discussion of the magnitude of the policy impact that could be reliably detected with these methods.

Our results are frankly disappointing. Despite our best efforts, we are wary of the predictions that we make with our statistical models, and we worry that they would be unable to support a strong inference about the impact of program changes. Our predictions do not do a good job of matching caseloads when there are no program changes, and there is evidence that our models may be missing some important determinants of caseloads. In short, we are not sure that we can make very reliable counterfactual statements so that we cannot reliably evaluate program impacts.

Chapter 1 Causality

1.1 Causality and Counterfactuals

Why do we care about establishing causal connections? One answer is that humans depend upon causation all the time to explain what has happened to them, to make realistic predictions about what will happen, and to affect what happens in the future. Not surprisingly, we are inveterate searchers after causes. Almost no one goes through a day without uttering sentences of the form “*X caused Y*” or “*Y occurred because of X,*” even if the utterances are for the mundane purposes of explaining why the tree branch fell (“the high winds *caused* the branch to break and gravity *caused* it to fall”), why we will be late to work (“the traffic congestion *will cause* me to be late”), or why we won’t return a phone call (“*because* I do not want that person to bother me again”). All these statements have the same form in which a cause (*X*) leads to an effect (*Y*).¹

Social scientists typically deal with bigger and more contentious causal claims such as:

“Strict work requirements cause people to leave welfare faster.”

“A good economy causes people to get off welfare and to stay off welfare.”

“PRWORA caused the welfare caseload to decline.”

These are bigger causal claims, but the form of the statement is the same.² The goals are also the same. Causal statements explain events, allow predictions about the future, and make it possible to take actions to affect the future. Knowing more about causality can be useful for social science researchers.

Causal statements are so useful that most people cannot let an event go by without asking why it happened and offering their own “because”. They often enliven these discussions with counterfactual assertions such as “If the cause had not occurred, then the effect would not have happened.” Counterfactuals are closely related to causal statements. A counterfactual is a statement, typically in the subjunctive mood, in which a false or “counter to fact” premise is followed by some assertion about what would have happened if the premise were true. For example, from about 1993 to 2000, the American economy performed very well, and the performance of the economy is one possible explanation for the dramatic reduction in welfare caseloads that occurred from about 1995 onwards. Some observers, however, have argued that welfare reform did much of the work in reducing caseloads so that “if we had not had welfare reform, then there would not have been such a sharp decline in the welfare caseload.” The statement uses the subjunctive (“if we had not had welfare reform, then there would not have

¹ The word “because” suggests that an explanation is being offered as well as a causal process. One way of explaining an event is identify a cause for it.

² Some philosophers deny that causation exists, but we agree with the philosopher D.H. Mellors (1995) who says: “I cannot see why. I know that educated and otherwise sensible people, even philosophers who have read Hume, can hold bizarre religious beliefs. I know philosophers can win fame and fortune by gulling literary theorists and others with nonsense they don’t themselves believe. But nobody, however gullible, believes in *no* causation (page 1).”

been...”), and the premise is counter to the facts. The premise is false because welfare reform did occur in the real world as it unfolded. The counterfactual claim is that without welfare reform, the world would have proceeded differently, and we would not have had such a large caseload reduction (even with the strong economy.) Is this counterfactual true?

The truth of counterfactuals is closely related to the existence of causal relationships. The counterfactual claim made above implies that there is a causal link between the passage of PRWORA (the cause X) and the decline in welfare caseloads (the effect Y). The counterfactual, for example, would be true if PRWORA did *cause* the reductions in caseloads through its emphasis on welfare to work or on time limits. Therefore, if we had not made these changes in welfare, then we would not have had dramatic declines in welfare caseloads because welfare recipients would not have been encouraged to go to work and get off welfare.

Another way to think about this is to simply ask what would have happened in the *most similar world* in which PRWORA did not pass. Would the caseload declines still have happened? One way to answer this question would be to rerun the world with the cause eradicated so that PRWORA was not passed. The world would otherwise be the same. The economy would still be strong from 1993 to 2000, Bill Clinton would still be President during this period, and George W. Bush would win the November 2000 election. Everything else would be the same. If the caseload reductions did not occur in this world, then we would say that the counterfactual is true. Thus, the statement that welfare reform *caused* the caseload reductions is essentially the same as saying that in the *most similar world* in which welfare reform did not occur, the caseload reductions did not occur either. The existence of a causal connection can be checked by determining whether or not the counterfactual would be true in the most similar possible world where its premise is true. The problem, of course, is defining the most similar world and finding evidence for what would happen in it.

Consider the problem of definition first. Suppose, as was the case, that in the real world, PRWORA was enacted, and it was enacted in a world with a strong economy. Then consider two possible descriptions of the causal relationship between PRWORA and caseload reductions. In the economic sufficiency condition, the strong economy is sufficient for caseload reductions and PRWORA is not even necessary, so that in a world without PRWORA but with a strong economy there would still be caseload reductions. Thus PRWORA should not be held responsible for them. The economy would be the causal factor. In the interaction condition, the strong economy is only necessary for caseload reductions and the stimulus of PRWORA must interact with it to push people off welfare so that in a world without PRWORA but with a strong economy there would be no caseload reductions. Thus PRWORA should be held responsible for them; PRWORA is a causal factor.

But now consider the conditions under which PRWORA was enacted and what it means for our definition of a most similar possible world. Suppose that in a strong economy, the enactment of PRWORA was inevitable. Then a closest possible world without PRWORA would also be a world with a weak economy, and in this world, there would never be caseload reductions no matter what the causal mechanism. If the true relationship between PRWORA and caseload reductions is the economic sufficiency condition then the weak economy will insure that there are no caseload reductions. If the true relationship between RPWORA and caseload reductions

is the interaction condition then the lack of both PRWORA and a strong economy will doubly insure there are no caseload reductions. Thus, we would always conclude that PRWORA had an impact because we get caseload reductions when it is present and no caseload reductions when it is absent.

The problem here is obvious, even if the economic sufficiency condition is the truth, PRWORA comes bundled with a strong economy so that it is impossible to describe a closest possible world in which PRWORA is absent and the strong economy is present. Hence, we can be led astray into thinking that PRWORA is the causal factor when the economy is actually doing the work. This example demonstrates that defining the most similar possible world is a difficult task.

Beyond these definitional questions about most similar worlds, there is the problem of finding evidence for what would happen in the most similar world. We cannot rerun the world so that PRWORA does not pass. What can we do? Many philosophers have wrestled with this question, and we discuss the problem in detail later in the section on the counterfactual theory of causation.³ For now, we merely note that people act as if they can solve this problem because they assert the truth of counterfactual statements all the time.

Causality is at the center of explanation and understanding, but what, exactly, is it? And how is it related to counterfactual thinking? Four distinct theories of causality, summarized in Table 1.1, provide answers to these and other questions about causality. Philosophers debate which theory is the right one. For our purposes, we embrace them all, and we discuss them in more detail in the following pages. Our primary goal is developing better social science methods for studying the impacts of welfare programs, and our perspective is that all these theories capture some aspect of causality. Therefore, practical researchers can profit from drawing lessons from each one of them even though their proponents sometimes treat them as competing or even contradictory.

We believe that a really good causal inference should satisfy the requirements of all four theories. Causal inferences will be stronger to the extent that they are based upon finding all the following.

- (1) Constant conjunction of causes and effects required by the neo-Humean theory which requires this constant conjunction of cause and effect.
- (2) No effect when the cause is absent in the most similar world to where the cause is present as required by the counterfactual theory.
- (3) An effect after a cause is manipulated.
- (4) Activities and processes linking causes and effects required by the mechanism theory.

³ Standard theories of logic cannot handle counterfactuals because propositions with false premises are automatically considered true which would mean that all counterfactual statements, with their false premises, would be true, regardless of whether or not a causal link existed. Modal logics, which try to capture the nature of necessity, possibility, contingency, and impossibility, have been developed for counterfactuals (Lewis, 1973a). These logics typically judge the truthfulness of the counterfactual on whether or not the statement would be true in the most similar possible world where the premise is true. Problems arise, however, in defining the most similar world. These logics, by the way, typically broaden the definition of counterfactuals to include statements with true premises for which they consider the closest possible world to be the actual world so that their truth value is judged by whether or not their conclusion is true in the actual world.

Table 1.1
Four Theories of Causality

	Neo-Humean Regularity Theory	Counterfactual Theory	Manipulation Theory	Mechanisms and Capacities
Major Authors Associated with the Theory	Hume (1739); Mill (1888); Hempel (1965); Beauchamp & Rosenberg (1981)	Lewis (1973a,b; 1986)	Gasking (1955); Menzies & Price (1993); von Wright (1971)	Harre & Madden (1975); Cartwright (1989); Glennan (1996);
Approach to the Symmetric Aspect of Causality	Observation of constant conjunction and correlation	Consideration of truth of: “If the cause occurs then so does the effect” and “if the cause does not occur then the effect does not occur either”	Recipe that regularly produces the effect from the cause.	Consideration of whether there is a mechanism or capacity that leads from the cause to the effect.
Approach to the Asymmetric Aspect of Causality	Temporal precedence	Consideration of the truth of: “If the effect does not occur, then the cause may still occur.”	Observation of the effect of the manipulation	An appeal to the operation of the mechanism
Major problems solved.	Necessary connection.	Singular causation. Nature of necessity.	Common cause and causal direction.	Preemption.
Emphasis on Causes of Effects or Effects of Causes?	Causes of effects (E.g., Focus on dependent variable in regressions.)	Effects of causes (E.g., Focus on treatment’s effects in experiments.)	Effects of causes (E.g., Focus on treatment’s effects in experiments.)	Causes of effects (E.g., Focus on mechanism that creates effects.)
Studies with Comparative Advantage Using this Definition	Observational and causal modeling	Experiments; Case study comparisons; Counterfactual thought experiments	Experiments; Natural experiments; Quasi-experiments	Analytic models; Case studies

The claim that smoking causes lung cancer, for example, first arose in epidemiological studies that found a correlation between smoking and lung cancer. These results were highly suggestive to many, but this correlational evidence was insufficient to others (including one of the founders of modern statistics, R. A. Fisher). These studies were followed by experiments that showed that, at least in animals, the absence of smoking reduced the incidence of cancer compared to the incidence with smoking when similar groups were compared. But animals, some suggested, are not people. Other studies showed that when people stopped smoking (that is, when the putative cause of cancer was manipulated) the incidence of cancer went down as well. Finally, recent studies have uncovered biological mechanisms that explain the link between smoking and lung cancer. Taken together the evidence for a relationship between smoking and lung cancer now seems overwhelming.

In the next section, we introduce some basic ideas about causation that underlie the four theories, and the remainder of this chapter explains the four theories in much more detail.

1.2 Correlation and Manipulation

The modern study of causality begins with David Hume (1711-1776) who was writing at a time when the pre-eminent theory of causality was the existence of a necessary connection – a kind of “hook” or “force” – between causes and their effects so that a particular cause must be followed by a specific effect. Hume looked for the feature of causes that guaranteed their effects. He argued that there was no evidence for the necessity of causes because all we could ever find in events was the contiguity, precedence, and regularity of cause and effect. There was no evidence for any kind of hook or force. For Hume, the major feature of causation, beyond temporal precedence and contiguity, is simply the regularity of the association of causes with their effects, but there is no evidence for any kind of hook or necessary connection between causes and effects.⁴

The Humean analysis of causation became the predominant perspective in the nineteenth and most of the twentieth century, and it led in two directions both of which focused upon the logical form of causal statements. Some, such as the physicist Ernst Mach, the philosopher Bertrand Russell, and the statistician/geneticist Karl Pearson concluded that there was nothing more to causation than regularity so that the entire concept of causation should be abandoned in favor of functional laws or measures of association such as correlation which summarized the regularity.⁵ Others, such as the philosophers John Stuart Mill (1888), Karl Hempel (1965), and Tom Beauchamp and Alexander Rosenberg (1981) looked for ways to strengthen the regularity condition so as to go beyond mere accidental regularities. For them, true cause and effect regularities must be unconditional and follow from some lawlike statement. Their neo-Humean approach improved upon Hume’s theory, but, as we shall see, there appears to be no way to define lawlike statements in a way that captures all that we mean by causality.

What, then, do we typically mean by causality? In their analysis of the fundamental metaphors used to mark the operation of causality, the linguist George Lakoff and the philosopher Mark Johnson (1980a,b, 1999) describe prototypical causation as “the manipulation of objects by force, the volitional use of bodily force to change something physically by direct contact in one’s immediate environment. (1999, page 177)” Causes bring, throw, hurl, propel, lead, drag, pull, push, drive, tear, thrust, or fling the world into new circumstances. These verbs suggest that causation is forced movement, and for Lakoff and Johnson the “Causation Is Forced Movement

⁴ There are different interpretations of what Hume meant. For a thorough discussion see Beauchamp and Rosenberg (1981).

⁵ Bertrand Russell famously wrote that “the word ‘cause’ is so inextricably bound up with misleading associations as to make its complete extrusion from the philosophical vocabulary desirable.... The law of causality, like so much that passes muster among philosophers, is a relic of a bygone age, surviving like the monarchy, only because it is erroneously supposed to do no harm.” (Russell, 1918). Karl Pearson rejected causation and replaced it with correlation: “Beyond such discarded fundamentals as ‘matter’ and ‘force’ lies still another fetish amidst the inscrutable arcana of even modern science, namely the category of cause and effect. Is this category anything but a conceptual limit to experience, and without any basis in perception beyond a statistical approximation?” (Pearson, 1911, page vi) “It is this conception of correlation between two occurrences embracing all relationship from absolute independence to complete dependence, which is the wider category by which we have to replace the old idea of causation.” (Pearson, 1911, page 157).

metaphor is in a crucial way constitutive of the concept of causation (page 187).” Causation as forceful manipulation differs significantly from causation as the regularity of cause and effect because forceful manipulation emphasizes intervention, agency, and the possibility that the failure to engage in the manipulation will prevent the effect from happening. For Lakoff and Johnson, causes are forces and capacities that entail their effects in ways that go beyond mere regularity and that are reminiscent of the causal “hooks” rejected by Hume, although instead of hooks they emphasize manipulation, mechanisms, forces, and capacities.⁶

“Causation as regularity” and “causation as manipulation” are quite different notions, but each carries with it some essential features of causality. Two other notions of causality, the “counterfactual” and the “mechanistic,” also provide insights about it. Before going on to describe each of these theories, it is worthwhile reminding ourselves that causes are always defined in relation to what the philosopher John Mackie calls a “causal field” of other factors.

The existence of a causal field means that what people choose to consider the cause of an event depends upon how they define the causal field. Thus, an unfortunate cigarette smoker who causes an explosion in a house by igniting a gas leak will probably consider the causal field to be a situation where lighting a cigarette and no gas leak is the norm, hence the gas leak will be identified as the cause of the explosion. But an equally unfortunate cigarette smoker who causes an explosion at a gas station will probably consider lighting the cigarette to be the cause of the explosion and not the fact that gas fumes were present at the station.⁷ Similarly, a political scientist who studies great power politics may consider growing instability in the great power system to be the cause of World War I because a stable system could have weathered the assassination of Archduke Ferdinand, but an historian who studies the impact of assassination on historical events might argue that World War I was a prime example of how assassinations can cause bad consequences such as a world war. As Mackie notes, both are right, but “What is said to be caused, then, is not just an event, but an event-in-a-certain-field, and some ‘conditions’ can be set aside as not causing this-event-in-this-field simply because they are part of the chosen field, though if a different field were chosen, in other words if a different causal question were being asked, one of those conditions might well be said to cause this-event-in-that-other-field.” (Mackie, 1974, page 35)

1.3 Four Theories of Causality

Each of the four theories in Table 1.1 provides a different way to think about causality. The neo-Humean regularity theory depends upon observing the constant conjunction of supposed causes

⁶ As we shall show, two different theories of causation are conflated here. One theory emphasizes agency and manipulation. The other theory emphasizes mechanisms and capacities. The major difference is the locus of the underlying force that defines causal relationships. Agency and manipulation theories emphasize human intervention. Mechanism and capacity theories emphasize processes within nature itself.

⁷ Legal wrangling over liability often revolves around who should be blamed for an accident where the injured party has performed some action in a causal field. The injured party typically claims that the action should have been anticipated and its effects mitigated or prevented by the defendant in that causal field and the defendant argues that the action should not have been taken by the plaintiff or could not have been anticipated by the defendant.

and effects. The counterfactual theory requires the truth of counterfactuals such as “if the cause had not occurred then the effect would not have occurred either.” The manipulation theory considers what happens when a supposed cause is manipulated by a researcher. And the mechanisms and capacities theories look for activities and processes that link causes and effects. Each, we shall find, has something to contribute to our understanding of causality. The entries in the table describe the way that each theory takes care of some basic problems of establishing causality. These entries should become clearer as we proceed.

1.3.A Humean and Neo-Humean Theories of Causation

Lawlike Generalities and the Humean Regularity Theory of Causation – Humean and neo-Humean theories propose logical conditions that must hold for the constant conjunction of events to justify the inference that they have a cause-effect relationship. Specifically, Humeans have explored whether a cause must be sufficient for its effects, necessary for its effects, or something more complicated.

The classic definition shared by Hume, John Stuart Mill, and many others was that “ X is a cause of Y if and only if X is sufficient for Y .” That is, the cause must always and invariably lead to the effect. Certainly an X that is sufficient for Y can be considered a cause, but what about the many putative causes are not sufficient for their effect? Striking a match, for example, may be necessary for it to light, but it may not light unless there is enough oxygen in the atmosphere. Does this example show that striking a match is never a cause of a match lighting? Or being job-ready may be necessary to get a job and to get off welfare, but what if there are no jobs? Does this example demonstrate that being job-ready is never a cause of getting off welfare?

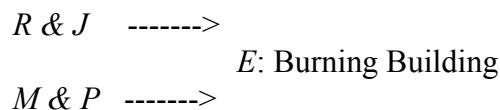
This leads to an alternative definition in which “ X is a cause of Y if and only if X is necessary for Y .” Under this definition, it is assumed that the cause (such as striking the match or being job-ready) must be present for the effect to occur, but it may not always be enough for the cause to actually occur (because there might not be enough oxygen or enough jobs). But how many causes are even necessary for their effects? If the match does not light after striking it, someone might use a blowtorch to light it so that striking the match is not even necessary for the match to ignite. Do we therefore assume that striking the match is never a cause of its lighting? Or someone may get off welfare without being job-ready because they get married to a partner who has a good job. Do we assume that being job-ready is never a cause of getting off welfare? Necessity and sufficiency seem unequal to the task of defining causation.⁸

These considerations led John Mackie to propose a set of conditions requiring that a cause be an insufficient [I] but necessary [N] part of a condition which is itself unnecessary [U] but exclusively sufficient [S] for the effect. These INUS conditions can be explained by an example. Consider two ways that the effect (E), which is getting off welfare, might occur. In one scenario a welfare recipient becomes job-ready and there are jobs available so she gets off welfare. In

⁸ And there are problems such as the following favorite of the philosophers: “If two bullets pierce a man’s heart simultaneously, it is reasonable to suppose that each is an essential part of a distinct sufficient condition of the death, and that neither bullet is *ceteris paribus* necessary for the death, since in each case the other bullet is sufficient.” (Sosa and Tooley, 1993, pp 8-9).

another, she becomes more marriageable and there are marriage partners available so she gets off welfare. A number of factors here are INUS conditions for getting off welfare. Being job-ready (*R*) and having jobs (*J*) available might be a path off welfare, or being more marriageable (*M*) and having marriage partners (*P*) available might be a path off welfare. Thus *R* and *J* together are exclusively sufficient [S] for getting off welfare and *M* and *P* together are exclusively sufficient [S] for getting off welfare. Furthermore, being job-ready and having jobs (*R&J*) are unnecessary [U] and being more marriageable and having marriage partners around (*M&P*) are unnecessary [U] because the person could have gotten off welfare with just one or the other combination of factors. Finally, *R*, *J*, *M*, or *P* alone is insufficient [I] to get off welfare even though *R* is necessary [N] in conjunction with *J* (or vice-versa) and *M* is necessary [N] in conjunction with *P* (or vice-versa). This formulation allows for the fact that no single cause is sufficient or necessary, but when experts say that being job-ready caused someone to get off welfare, they are saying in effect that being job-ready (*R*) is a condition such that it occurred, that the other conditions (*J*) which, conjoined with it, form a sufficient condition were also present, and that no other sufficient condition (such as *M&P*) of getting off welfare was present on this occasion. (Paraphrase of Mackie, 1965, page 245)

Figure 1.1



From the perspective of a practicing researcher, three lessons follow from the INUS conditions. First a putative cause such as *R* might not cause the effect *E* because *M&P* might be responsible. Hence, getting off welfare (*E*) will not always result from being job-ready (*R*) even though *R* could cause someone to get off welfare. Second, interactions among causes may be necessary for any one cause to be sufficient (*R* and *J* require each other and *M* and *P* require each other). Third, the relationship between any INUS cause and its effect might appear to be probabilistic because of the other INUS causes and variations in the causal field from one situation to another. For example, being job ready might appear to only sometimes cause someone to get off welfare depending upon whether or not jobs are available. In summary, the INUS conditions suggest the multiplicity of causal pathways and causes, the possibility of conjunctural causation (Ragin, 1987), and the likelihood that social science relationships will appear probabilistic even if they are deterministic.⁹ Even if causality is in some ultimate sense deterministic,¹⁰ the probabilistic nature of causality in practice means that we must use statistical methods to check for causality. We discuss statistical methods in Chapter 2 and especially Chapter 3, but for the remainder of this chapter we focus on the simpler case of deterministic causality.

INUS conditions reveal a lot about the complexities of causality, but as a definition of it, they turn out to be too weak – they do not rule out situations where there are common causes, and

⁹ These points are made especially forcefully in Marini and Singer (1988).

¹⁰ Chaos theory (Lorenz, 1963; Ruelle, 1990) suggests that even deterministic non-linear differential equations can lead to non-periodic behavior that appears random and that is extremely sensitive to the numerical values of starting points.

they do not exclude accidental regularities. The problem of common cause arises in a situation where, for example, more knowledge (K) gained through taking political science courses causes a welfare recipient to realize that she is eligible for another program (say Supplemental Security Income because of disability) which gets her off welfare (E) and simultaneously makes her more job-ready (R). That is, $K \rightarrow E$ and $K \rightarrow R$ (where the arrow indicates causation). Thus, the common cause of knowledge leads to the conjunction of job readiness (R) and getting off welfare (E), suggesting that job-readiness causes the departure from welfare even though the truth is that *knowledge is the common-cause of both*.

In some cases of common causes such as the rise in barometric pressure followed by the arrival of a storm, common sense tells us that the putative cause (the rise in barometric pressure) cannot be the real cause of the thunderstorm. But in the situation with an increase in knowledge from taking courses, the fact that being more job-ready has the capacity to get someone off welfare makes it less likely that we will realize that knowledge is the common cause of both the job-readiness and getting off welfare. We might be better off in the case where the political science course (as might be the case for some such courses!) clearly did not make someone more job-ready but only imparted some esoteric knowledge about eligibility for social programs and about causality. In that case we would probably reject the fantastic theory that knowledge about causality could get someone off welfare,¹¹ because knowledge about causality does not have the capacity to get someone off welfare, but the Humean theory would be equally confused by both situations because it could not appeal, within the ambit of its understanding, to causal capacities. For a Humean, the constant conjunction of esoteric knowledge about causality and getting off welfare suggests causation as much as the constant conjunction of job-readiness and getting off welfare.

Attempts to fix-up these conditions usually focus on trying to require “lawlike” statements that are unconditionally true, not just accidentally true. Since it is not unconditionally true that splitting wood causes fires, the presumption is that some such conditions can be found to rule-out this explanation. Unfortunately, no set of conditions seems to be successful.¹² Although the regularity theory identifies a necessary condition for describing causation, it basically fails because association is not causation and there is no reason why purely logical restrictions on lawlike statements should be sufficient to characterize causal relationships. Part of the problem is that there are many different types of causal laws and they do not fit any particular patterns. For example, one restriction that has been proposed to insure lawfulness is that lawlike statements should either not refer to particular situations or they should be derivable from laws that do not refer to particular situations. This would mean that Kepler’s first “law” about all planets moving in elliptical orbits around the sun (a highly specific situation!) was not a causal law before Newton’s laws were discovered, but it was a causal law after it was shown that it could be derived from Newton’s laws. But Kepler’s laws were always considered causal laws, and there seems to be no reason to rest their lawfulness on Newton’s laws. Furthermore, by this

¹¹ Actually, we believe that knowledge about causality could be very empowering, but for the purposes of this example, we suppose that it cannot get someone off welfare.

¹² For some representative discussions of the problems see (Harre and Madden, 1975, Chapter 2; Salmon, 1989, Chapters 1-2; Hausman, 1998, Chapter 3). Salmon (1989, page 15) notes that “Lawfulness, modal import [what is necessary, possible, or impossible], and support of counterfactuals seems to have a common extension; statements either possess all three or lack all three. But it is extraordinarily difficult to find criteria to separate those statements that do from those that do not.”

standard, almost all social science and natural science laws (e.g., plate tectonics) are about particular situations. In short, logical restrictions on the form of laws do not seem sufficient to characterize causality.

The Asymmetry of Causation – The regularity theory also fails because it does not provide an explanation for the asymmetry of causation. Causes should cause their effects, but INUS conditions are almost always symmetrical such that if *C* is an INUS cause of *E*, then *E* is also an INUS cause of *C*. It is almost always possible to turn around an INUS condition so that an effect is an INUS for its cause. One of the most famous examples of this problem involves a flagpole, the elevation of the sun, and the flagpole's shadow. The law that light travels in straight lines implies that there is a relationship between the height of the flagpole, the length of its shadow, and the angle of elevation of the sun. When the sun rises, the shadow is long, at midday it is short, and at sunset it is long again. Intuition about causality suggests that the length of the shadow is caused by the height of the flagpole and the elevation of the sun. But, using INUS conditions, we can just as well say that the elevation of the sun is caused by the height of the flagpole and the length of the shadow. There is simply nothing in the conditions that precludes this fantastic possibility.

The only feature of the Humean theory that provides for asymmetry is temporal precedence. If changes in the elevation of the sun precede corresponding changes in the length of the shadow, then we can say that the elevation of the sun causes the length of the shadow. And if changes in the height of the flagpole precede corresponding changes in the length of the shadow, we can say that the height of the flagpole causes the length of the shadow. But many philosophers reject making temporal precedence the determinant of causal asymmetry because it precludes the possibility of *explaining* the direction of time by causal asymmetry and it precludes the possibility of backwards causation. From a practical perspective, it also requires careful measures of timing that may be difficult in a particular situation.

Summary – This discussion reveals two basic aspects of the causal relation. One is a symmetrical form of association between cause and effect and the other is an asymmetrical relation in which causes produce effects but not the reverse. The Humean regularity theory, in the form of INUS conditions, provides a necessary condition for the existence of the symmetrical relationship, but it does not rule out situations such as common cause and accidental regularities where there is no causal relationship at all. From a methodological standpoint, it can easily lead researchers to presume that all they need to do is to find associations, and it also leads to an under emphasis on the rest of the requirement for a “lawlike” or “unconditional” relationship because it does not operationally define what that would really mean. A great deal of what passes for causal modeling suffers from these defects (Freedman, 1987, 1991, 1997, 1999)

The Humean theory does even less well with the asymmetrical feature of the causal relationship because it provides no way to determine asymmetry except temporal precedence. Yet there are many other aspects of the causal relation that seem more fundamental than temporal precedence. Causes not only typically precede their effects, but they also can be used to explain effects or to manipulate effects while effects cannot be used to explain causes or to manipulate them.

1.3.B Counterfactual Definition of Causation

In a book *On the Theory and Method of History* published in 1902, Eduard Meyer claimed that it was an “unanswerable and so an idle question” whether the course of history would have been different if Bismarck, then Chancellor of Prussia, had not decided to go to war in 1866. Speculation about Bismarck’s decision may be idle, but by some accounts, the Austro-Prussian-Italian War of 1866 paved the way for German and Italian unification (see, Wawro, 1996). In reviewing Meyer’s book in 1906, Max Weber agreed that “from the strict ‘determinist’ point of view” finding out what would have happened if Bismarck had not gone to war “was ‘impossible’ given the ‘determinants’ which were in fact present.” But he went on to say that “And yet, for all that, it is far from being ‘idle’ to raise the question what might have happened, if, for example, Bismarck had not decided for war. For it is precisely this question which touches on the decisive element in the historical construction of reality: the causal significance which is properly attributed to this individual decision within the totality of infinitely numerous ‘factors’ (all of which must be just as they are and not otherwise) if precisely this consequence is to result, and the appropriate position which the decision is to occupy in the historical account.” (Weber, 1906 [1978], p. 111). Weber’s review is an early discussion of the importance of counterfactuals for understanding history and making causal inferences. He argues forcefully that if “history is to raise itself above the level of a mere chronicle of noteworthy events and personalities, it can only do so by posing just such questions” as the counterfactual in which Bismarck did not decide for war.¹³

Lewis’s Counterfactual Theory of Causation – The philosopher David Lewis (1973b) has proposed the most elaborately worked out theory of how causality is related to counterfactuals. His theory requires the truth of two statements regarding two distinct events *X* and *Y*. Lewis starts from the presumption that *X* and *Y* have occurred so that the “counterfactual” statement:¹⁴ “If *X* were to occur, then *Y* would occur” is true. The truth of this statement is Lewis’s first condition for a causal relationship. Then he considers the truth-value of a second counterfactual:¹⁵ “If *X* were not to occur, then *Y* would not occur either.” If this is true as well, then he says that *X* causes *Y*. If, for example, Bismarck decided for war in 1866 and, as some historians argue, German unification followed because of his decision, then we must ask “If Bismarck had not decided for war, would Germany have remained divided?” The heart of Lewis’s theory is the set of requirements, described below, that he lays down for the truth of this kind of counterfactual.

Lewis’ theory has a number of virtues. It deals directly with singular causal events, and it does not require the examination of a large number of instances of *X* and *Y*. At one point in the philosophical debate about causation, it was believed that the individual cases such as “the hammer blow caused the glass to break” or “the assassination of Archduke Ferdinand caused World War I” could not be analyzed alone because these cases had to be subsumed under a general law (“hammer blows cause glass to break”) derived from multiple cases plus some particular facts of the situation in order to meet the requirement for a “lawlike” relationship. The

¹³ We are indebted to Richard Swedberg for pointing us towards Weber’s extraordinary discussion.

¹⁴ Lewis considers statements like this as part of his theory of counterfactuals by simply assuming that statements in the subjunctive mood with true premises and true conclusions are true. As noted earlier, most theories of counterfactuals have been extended to include statements with true premises by assuming, quite reasonably, that they are true if their conclusion is true and false otherwise.

¹⁵ This is a simplified version of Lewis’s theory based upon Lewis (1973a,b; 1986) and Hausman (1998, Chapter 6).

counterfactual theory, however, starts with singular events and proposes that causation can be established without an appeal to a set of similar events and general laws regarding them.¹⁶ The possibility of analyzing singular causal events is important for all researchers, but especially for those doing case studies who want to be able to say something about the consequences of Stalin succeeding Lenin as head of the Soviet Union or the impact of the butterfly ballot on the 2000 election.

The counterfactual theory also deals directly with the issue of *X*'s causal "efficacy" with respect to *Y* by considering what would happen if *X* did not occur. The problem with the theory is the difficulty of determining the truth or falsity of the counterfactual "If *X* were not to occur, then *Y* would not occur either." The statement cannot be evaluated in the real world because *X* actually occurs so that the premise is false, and there is no evidence about what would happen if *X* did not occur. It only makes sense to evaluate the counterfactual in a world in which the premise is true. Lewis's approach to this problem is to consider whether the statement is true in the closest possible world to the actual world where *X* does not occur. Thus, if *X* is a job-readiness program and *Y* is getting off welfare, then the closest possible world is one in which everything else is the same except that the person does not get the job-readiness program. If in this world, the person does not get off welfare, then the counterfactual is true, and the job-readiness program (*X*) causes the person to get off welfare (*Y*). The obvious problem with this approach is, as we showed earlier, identifying the closest possible world. If *X* is the assassination of Archduke Ferdinand and *Y* is World War I, is it true that World War I would not have occurred in the closest possible world where the bullet shot by the terrorist Gavrilo Princip did not hit the Archduke? Or would some other incident have inevitably precipitated World War I? And, to add to the difficulty, would this "World War I" be the same as the one that happened in our world?

Lewis' theory substitutes the riddle of determining the similarity of possible worlds for the neo-Humean theory's problem of determining lawlike relationships. To solve these problems, both approaches must be able to identify similar causes and similar effects. The Humean theory must identify them across various situations in the real world. This aspect of the Humean approach is closely related to John Stuart Mill's "Method of Concomitant Variation" which he described as follows: "Whatever phenomenon varies in any manner, whenever another phenomenon varies in some similar manner, is either a cause or an effect of that phenomenon, or is connected to it through some fact of causation." (Mill, 1888, page xxx)¹⁷ Lewis's theory must also identify similar causes and similar effects in the real world in which the cause does occur and in the many possible worlds in which the cause does not occur. This approach is closely related to Mill's "Method of Difference" in which: "If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances

¹⁶ In fact, many authors now believe that general causation (involving lawlike generalizations) can only be understood in terms of singular causation. "...general causation is a generalization of singular causation. Smoking causes cancer if (if and only if) smokers' cancers are generally caused by their smoking." (Mellors, 1995, pages 6-7). See also Sosa and Tooley, 1993. More generally, whereas explanation was once thought virtually to supersede the need for causal statements, many philosophers now believe that a correct analysis of causality will provide a basis for suitable explanations (see Salmon, 1990).

¹⁷ The Humean theory also has affinities with Mill's Method of Agreement which he described as follows: "If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon." (Mill, 1888, page 280)

differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon (Mill, 1888, page 280).¹⁸”

In addition to identifying similar causes and similar effects, the Humean theory must determine if the conjunction of these similar causes and effects is accidental or lawlike. This task requires understanding what is happening in each situation and comparing the similarities and differences across situations. Lewis’s theory must identify the possible world where the cause does not occur that is most similar to the real world. This undertaking requires understanding the facts of the real world and the laws that are operating in it. Consequently, assessing the similarity of a possible world to our own world requires understanding the lawlike regularities that govern our world.¹⁹ It seems as if Lewis has simply substituted one difficult task, that of establishing lawfulness, for the job of identifying the most similar world.

The Virtues of the Counterfactual Definition of Causation – Lewis has substituted one difficult problem for another, but the reformulation of the problem has a number of benefits. The counterfactual approach provides new insights into what is required to establish causal connection between causes and effects. The counterfactual theory makes it clear that establishing causation does not require observing the universal conjunction of a cause and an effect.²⁰ One observation of a cause followed by an effect is sufficient for establishing causation if it can be shown that in a most similar world without the cause, the effect does not occur. The counterfactual theory proposes that causation can be demonstrated by simply finding a most similar world in which the absence of the cause leads to the absence of the effect. Consequently, comparisons, specifically the kind of comparison advocated by John Stuart Mill in his “Method of Difference,” have a central role in the counterfactual theory as they do in the analysis of case studies.

Lewis’s theory provides us with a way to think about the causal impact of singular events such as the badly designed butterfly ballot in Palm Beach County, Florida that led some voters in the 2000 Presidential election to complain that they mistakenly voted for Reform Party candidate Patrick Buchanan when they meant to vote for Democrat Al Gore. The ballot can be said to be causally associated with these mistakes if in the closest possible world in which the butterfly ballot was not used the vote for Buchanan was lower than in the real world. Ideally this closest possible world would be a parallel universe in which the same people received a different ballot, but this, of course, is impossible. The next best thing is a situation where similar people employed a different ballot. In fact, the butterfly ballot was only used for election day voters in Palm Beach County. It was not used by absentee voters. Consequently, the results for the absentee voting can be considered a surrogate for the closest possible world in which the butterfly ballot was not used, and in this absentee voting world, voting for Buchanan was dramatically lower, suggesting that at least 2000 people who preferred Gore – more than enough

¹⁸ Mill goes on to note that the Method of Difference is “a method of artificial experiment.” (Page 281). Notice that for both the Method of Concomitant Variation and the Method of Difference, Mill emphasizes the association between cause and effect and not the identification of which event is the cause and which is the effect. Mill’s methods are designed to detect the symmetric aspect of causality but not its asymmetric aspect.

¹⁹ Nelson Goodman makes this point in a 1947 article on counterfactuals, and James Fearon (1991) discusses its implications for counterfactual thought experiments in political science. Also see Tetlock and Belkin (1996).

²⁰ G. H. von Wright notes that the counterfactual conception of causality shows that the hallmark of a lawlike connection is “*necessity and not universality*.” (von Wright, 1971, page 22)

to give the election to Gore – mistakenly voted for Buchanan on the butterfly ballot.

The difficult question, of course, is whether the absentee voting world can be considered a good enough surrogate for the closest possible world in which the butterfly ballot was not used.²¹ The counterfactual theory does not provide us with a clear sense of how to make that judgment.²² But the framework does suggest that we should consider the similarity of the election-day world and the absentee voter world. To do this, we can ask whether election day voters are different in some significant ways from absentee voters, and this question can be answered by considering information on their characteristics and experiences. In summary, the counterfactual perspective allows for analyzing causation in singular instances, and it emphasizes comparison, which seems difficult but possible, rather than the recondite and apparently fruitless investigation of the lawfulness of statements such as “All ballots that place candidate names and punch-holes in confusing arrangements will lead to mistakes in casting votes.”

1.3.C Experimentation and the Manipulation Theory of Causation

The manipulation theory of causation rests on a simple insight. When some factor has been manipulated and a change subsequently occurs, then the factor that has been manipulated is an excellent candidate for the causal factor.²³ It is hard to exaggerate the importance of this insight. Although philosophers are uncomfortable with manipulation and agency theories of causality because they put people (as the manipulators) at the center of our understanding of causality, there can be little doubt about the power of manipulation for determining causality. Agency and manipulation theories of causation (Gasking, 1955; von Wright, 1975; Menzies and Price, 1993) elevate this insight into their definition of causation. For Gasking “the notion of causation is essentially connected with our manipulative techniques for producing results” (1955, pages 483), and for Menzies and Price “events are causally related just in case the situation involving them possesses intrinsic features that *either* support a means-end relation between the events as is, *or* are identical with (or closely similar to) those of another situation involving an analogous pair of means-end related events.” (Menzies and Price, 1993, pages 197). These theories focus on establishing the direction of causation, but Gasking’s metaphor of causation as “recipes” also suggests an approach towards establishing the symmetric, regularity aspect of causation. Causation exists when there is a recipe that regularly produces effects from causes.

Perhaps our ontological definitions of causality should not employ the concept of agency

²¹ For an argument that the absentee votes are an excellent surrogate, see Wand et al., “The Butterfly Did It,” *American Political Science Review*, December, 2001.

²² In his book on counterfactuals, Lewis only claims that similarity judgments are possible, but he does not provide any guidance on how to make them. He admits that his notion is vague, but he claims it is not ill understood. “But comparative similarity is not ill-understood. It is vague—very vague—in a well-understood way. Therefore it is just the sort of primitive that we must use to give a correct analysis of something that is itself undeniably vague.” (Lewis, 1973a, page 91). In later work Lewis (1979, 1986) formulates some rules for similarity judgments, but they do not seem very useful to us and to others (Bennett, 1984).

²³ It might be more correct to say that the cause is buried somewhere among those things that were manipulated or that are associated with the manipulation. It is not always easy, however, to know what was manipulated as in the famous Hawthorne experiments in which the experimenters thought the treatment was reducing the lighting for workers but the workers apparently thought of the treatment as being treated differently from all other workers. (Franke and Kaul, 1978; Jones, 1992) Part of the work required for good causal inference is clearly describing what was manipulated and unpacking it to see what feature caused the effect.

because most of the causes and effects in the universe go their merry way without human intervention, and even our epistemological methods often discover causes, as with Newtonian mechanics or astrophysics, where human manipulation is impossible. Yet our epistemological methods cannot do without agency because human manipulation appears to be the best way to identify causes, and many researchers and methodologists have fastened upon interventions as the way to pin-down causation. Herbert Simon, for example, bases his definition of causality on experimental systems because “in scientific literature the word ‘cause’ most often occurs in connection with some explicit or implicit notion of an experimenter’s intervention in a system.” (Simon, 1952, page 518). When full experimental control is not possible, Thomas Cook and Donald T. Campbell recommend “quasi-experimentation,” in which “an abrupt intervention at a known time” in a treatment group makes it possible to compare the impacts of the treatment over time or across groups (Cook and Campbell, 1986, page 149). The success of quasi-experimentation depends upon “a world of probabilistic multivariate causal agency in which some manipulable events dependably cause other things to change.” (Page 150). John Stuart Mill suggests that the study of phenomena which “we can, by our voluntary agency, modify or control” makes it possible to satisfy the requirements of the Method of Difference (“a method of artificial experiment”) even though “by the spontaneous operations of nature those requisitions are seldom fulfilled.” (Mill, 1888, pages 281, 282). Sobel champions a manipulation model because it “provides a framework in which the non-experimental worker can think more clearly about the types of conditions that need to be satisfied in order to make inferences” (Sobel, 1995, page 32). David Cox claims that quasi-experimentation “with its interventionist emphasis seems to capture a deeper notion” (Cox, 1992, page 297) of causality than the regularity theory.

There are those who dissent from this perspective, but even they acknowledge “wide agreement that the idea of causation as consequential manipulation is stronger or ‘deeper’ than that of causation as robust dependence.” (Goldthorpe, 2001, page 5). This account of causality is especially compelling if the manipulation theory and the counterfactual theory are conflated, as they often are, and viewed as one theory. Philosophers seldom combine them into one perspective, but all the methodological writers cited above (Simon, Cook and Campbell, Mill, Sobel, and Cox) conflate them because they draw upon controlled experiments, which combine intervention and control, for their understanding of causality.

One of the strengths of experiments is that experimenters typically worry about whether their manipulations actually occur. In physical experiments, researchers test their apparatus or experimental set-up to make sure that it has the intended consequences. In psychological experiments, researchers often include “manipulation checks” to see if the “manipulated independent variable was actually enacted and/or perceived by the subject as intended (Sawyer, Lynch, and Brinberg, 1995).” More generally, the statistician Don Rubin argues that “the motto ‘no causation without manipulation’ is a critical guideline for clear thinking in empirical studies for causal effects (Rubin, 1986, 962).” This guideline forces those doing observational studies to clarify whether they intend to make a causal statement when they say that education or sex is associated with some dependent variable such as the length of welfare spells. If they do intend to make a causal statement, then they must describe the experiment or manipulation that could be undertaken to show how these characteristics affect welfare spells. They must describe a specific educational program or a specific sex change program that might affect the length of welfare spells. This description must go beyond describing a change in the value of the

independent variable. It must describe a concrete set of actions that would change the value of variable.

One of the advantages of program evaluation is that programs are typically manipulations that involve specific actions at a particular time and place. These manipulations may be fuzzily described (as with some job readiness or job training programs), but they always involve attempts to do something and to change attitudes or behaviors. The program itself is an empirical referent for the manipulation, and there is an identifiable moment when the person enters the program, an identifiable set of actions that occur, and an identifiable moment when the person leaves the program. This situation is much different than when characteristics are associated with an outcome variable such as spell length. At best, the association of individual characteristics with an outcome variable leaves implicit the manipulations that might be undertaken to move someone from one characteristic (e.g., male) to another (female). At worst, the natural variation in a characteristic (e.g., education level) is the result of so many different processes that it usually makes no sense to let the natural variation be a proxy for a program that manipulates the variable. For example, a program that tries to provide welfare mothers with more education to improve their job prospects (such as the California Cal-Learn program that provided support services and monetary incentives and sanctions to encourage welfare mothers to finish high school) is much different than the natural processes whereby welfare mothers attain a certain level of education. And there is no reason to believe that the impact of the additional education provided by Cal-Learn has anything like the same impact as we observe in a regression of job success on education and other variables in the population.

1.3.D Preemption and the Mechanism Theory of Causation

Preemption – Experimentation’s amalgamation of the lessons of counterfactual and manipulation theories of causation produces a powerful technique for identifying the effects of manipulated causes. Yet, in addition to the practical problems of implementing the recipe correctly, the experimental approach does not deal well with two related problems. It does not solve the problem of causal preemption which occurs when one cause acts just before and preempts another, and it does not so much explain the causes of events as it demonstrates the effects of manipulated causes. In both cases, the experimentalists’ focus on the impacts of manipulations in the laboratory instead of on the causes of events in the world, leads to a failure to explain important phenomena, especially those phenomena which cannot be easily manipulated or isolated.

The problem of preemption illustrates this point. The following example of preemption is often mentioned in the philosophical literature. A man takes a trek across a desert. His enemy puts a hole in his water can. Another enemy, not knowing the action of the first, puts poison in his water. Manipulations have certainly occurred, and the man dies on the trip. The enemy who punctured the water can thinks that she caused the man to die, and the enemy who added the poison thinks that he caused the man to die. In fact, the water dripping out of the can preempted the poisoning so that the poisoner is wrong. This situation poses problems for the counterfactual theory because one of the basic counterfactual conditions required to establish that the hole in the water can caused the death of the man, namely the truth of the counterfactual “if the hole had not been put in the water can, the man would not have died,” is false even though the man did in fact

die of thirst. The problem is that the man would have died of poisoning if the hole in the water can had not preempted that cause, and the “back-up” possibility of dying by poisoning falsifies the counterfactual.

The preemption problem is a serious one, and it can lead to mistakes even in well-designed experiments. Presumably the closest possible world to the one in which the water can has been punctured is one in which the poison has been put in the water can as well. Therefore, even a carefully designed experiment will conclude that the puncturing of the can did not kill the man crossing the desert because the unfortunate subject in the control condition would die (from poisoning) just as the subject in the treatment would die (from the hole in the water can). The experiment alone would not tell us how the man died. A similar problem could arise in medical experiments. Arsenic was once used to cure venereal disease, and it is easy to imagine an experiment in which doses of arsenic “cure” venereal disease but kill the patient while the members of the control group without the arsenic die of venereal disease at the same rate. If the experiment simply looked at the mortality rates of the patients, it would conclude that arsenic had no medicinal value because the same number of people died in the two conditions.

In both these instances, the experimental method focuses on the effects of causes and not on explaining effects by adducing causes. Instead of asking why the man died in his trek across the desert, the experimental approach asks what happens when a hole is put in the man’s canteen and everything else remains the same. The method concludes that the hole had no effect. Instead of asking what caused the death of the patients with venereal disease, the experimental method asks whether giving arsenic to those with venereal disease had any net impact on mortality rates. It concludes that it did not. In short, experimental methods do not try to explain events in the world so much as they try to show what would happen if some cause were manipulated. This does not mean that experimental methods are not useful for explaining what happens in the world, but it does mean that they sometimes miss the mark.

Mechanisms, Capacities, and the Pairing Problem –The preemption problem is a vivid example of a more general problem with the Humean account that requires a solution. The general problem is that constant conjunction of events is not enough to “pair-up” particular events even when preemption is not present. Even if we know that holes in water cans generally spell trouble for desert travelers, we still have the problem of linking a particular hole in a water can with a particular death of a traveler. Douglas Ehring notes that:

Typically, certain spatial and temporal relations, such as spatial/temporal contiguity, are invoked to do this job. [That is, the hole in the water can used by the traveler is obviously the one that caused his death because it is spatially and temporally contiguous to him.] These singularist relations are intended to solve the residual problem of causally pairing particular events, a problem left over by the generalist core of the Humean account. (Ehring, 1997, page 18)

Counterfactual theories, because they can explain singular causal events, do not suffer so acutely from this “pairing” problem, but the preemption problem shows that remnants of the difficulty remain even in counterfactual accounts (Ehring, 1997, Chapter 1). In both the desert traveler and arsenic examples, the counterfactual account cannot get at the proper pairing of causes and

effects because there are two redundant causes to be paired with the same effects. Something more is needed.

The solution in both these cases seems obvious, but it does not follow from the neo-Humean, counterfactual, or manipulation definitions of causality. The solution is to inquire more deeply into what is happening in each situation in order to describe the capacities and mechanisms that are operating. An autopsy of the desert traveler would show that the person died of thirst, and an examination of the water can would show that the water would have run out before the poisoned water could be imbibed. An autopsy of those given arsenic would show that the signs of venereal disease were arrested while other medical problems, associated with arsenic poisoning, were present. Further work might even show that lower doses of arsenic cure the disease without causing death. In both these cases, deeper inquires into the mechanism by which the causes and effects are linked would produce better causal stories.

But what does it mean to explicate mechanisms and capacities?²⁴ “Mechanisms” we are told by Machamer, Darden, and Craver (2000, page 3) “are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions.” The crucial terms in this definition are “entities and activities” which suggest that mechanisms have pieces. Glennan (1996, page 52) calls them “parts,” and he requires that it should be possible “to take the part out of the mechanism and consider its properties in another context (page 53).” Entities, or parts, are organized to produce change. For Glennan (page 52), this change should be produced by “the interaction of a number of parts according to direct causal laws.” The biological sciences abound with mechanisms of this sort such as the method of DNA replication, chemical transmission at synapses, and protein synthesis. But there are many mechanisms in the social sciences as well including markets with their methods of transmitting price information and bringing buyers and sellers together, electoral systems with their routines for bringing candidates and voters together in a collective decision-making process, the diffusion of innovation through social networks, the two-step model of communication flow, weak ties in social networks, dissonance reduction, reference groups, arms races, balance of power, etc. (Hedstrom and Swedberg, 1998). As these examples demonstrate, mechanisms are not exclusively mechanical, and their activating principles can range from physical and chemical processes to psychological and social processes. They must be composed of appropriately located, structured, and oriented entities which involve activities that have temporal order and duration, and “an activity is usually designated by a verb or verb form (participles, gerundives, etc.)” (Machamer, Darden, and Craver, 2000, page 4) which takes us back to the work of Lakoff and Johnson (1999) who identified a “Causation Is Forced Movement metaphor.”

Mechanisms provide another way to think about causation. Glennan argues that “two events are causally connected when and only when there is a mechanism connecting them” and “the necessity that distinguishes connections from accidental conjunctions is to be understood as deriving from a underlying mechanism” which can be empirically investigated (page 64). These

²⁴ These approaches are not the same, and those who favor one often reject the other (see, e.g., Cartwright, 1989 on capacities and Machamer, Darden, and Craver, 2000 on mechanisms). But both emphasize “causal powers” (Harre and Madden, 1975, Chapter 5) instead of mere regularity or counterfactual association. We focus on mechanisms because we believe that they are somewhat better way to think about causal powers, but in keeping with our pragmatic approach, we find much that is useful in “capacity” theories.

mechanisms, in turn, are explained by causal laws, but there is nothing circular in this because these causal laws refer to how the *parts* of the mechanism are connected. The operation of these parts, in turn, can be explained by lower level mechanisms. Eventually the process gets to a bedrock of fundamental physical laws which Glennan concedes “cannot be explained by the mechanical theory (page 65).”

Consider explaining social phenomena by examining their mechanisms. Duverger’s law, for example, is the observed tendency for just two parties in simple plurality single-member district elections systems (such as the United States). The entities in the mechanisms behind Duverger’s law are voters and political parties. These entities face a particular electoral rule (single district plurality voting) which causes two activities. One is that voters often vote strategically by choosing a candidate other than their most liked because they want to avoid throwing their vote away on a candidate who has no chance of winning and because they want to forestall the election of their least wanted alternative. The other activity is that political parties often decide not to run candidates when there are already two parties in a district because they anticipate that voters will spurn their third party effort.

These mechanisms underlying Duverger’s law suggest other things that can be observed beyond the regularity of two party systems being associated with single member plurality-vote electoral systems that led to the law in the first place. People’s votes should exhibit certain patterns and third parties should exhibit certain behaviors. And a careful examination of the mechanism suggests that in some federal systems that use simple plurality single-member district elections we might have more than two parties, seemingly contrary to Duverger’s Law. Typically, however, there are just two parties in each province or state, but these parties may differ from one state to another, thus giving the impression, at the national level, of a multi-party system even though Duverger’s Law holds in each electoral district.²⁵

Or consider meteorological²⁶ and physical phenomena. Thunderstorms are not merely the result of cold fronts hitting warm air or being located near mountains, they are the results of parcels of air rising and falling in the atmosphere subject to thermodynamic processes which cause warm humid air to rise, to cool, and to produce condensed water vapor. Among other things, this mechanism helps to explain why thunderstorms are more frequent in areas, such as Denver, Colorado, near mountains because the mountains cause these processes to occur – without the need for a cold air front. Similarly, Boyle’s law is not merely a regularity between pressure and volume, it is the result of gas molecules moving within a container and exerting force when they hit the walls of the container. This mechanism for Boyle’s law also helps to explain why temperature affects the relationship between the pressure and volume of a gas. When the temperature increases, the molecules move faster and exert more force on the container walls. Mechanisms like these are midway between general laws on the one hand and specific descriptions on the other hand, and activities can be thought of as causes which are not related to lawlike generalities.²⁷ Mechanisms typically explicate observed regularities in terms of lower

²⁵ This radically simplifies the literature on Duverger’s law (see Cox, 1997 for more details).

²⁶ The points in this paragraph, and the thunderstorm example, come from Dessler (1991).

²⁷ Jon Elster says: “Are there lawlike generalizations in the social sciences? If not, are we thrown back on mere description and narrative? In my opinion, the answer to both questions is No. The main task of this essay is to explain and illustrate the idea of a *mechanism* as intermediate between laws and descriptions.” (Elster, 1998, page 45)

level processes, and the mechanisms vary from field to field and from time to time. Moreover, these mechanisms “bottom-out” relatively quickly – molecular biologists do not seek quantum mechanical explanations and social scientists do not seek chemical explanations of the phenomena they study.

Finally, consider welfare programs where we know that education is one of the most powerful predictors of short welfare spells and success in getting off welfare. What mechanisms account for this? Does education make people more job-ready, more marriageable, more knowledgeable about welfare programs, or more confident? What are the exact mechanisms by which education leads to short welfare spells? Without knowing these mechanisms, it is hard to feel confident about causal claims about education and welfare.

When an unexplained phenomenon is encountered in a science, “Scientists in the field often recognize whether there are known types of entities and activities that can possibly accomplish the hypothesized changes and whether there is empirical evidence that a possible schemata is plausible.” They turn to the available types of entities and activities to provide building blocks from which to construct hypothetical mechanisms. “If one knows what kind of activity is needed to do something, then one seeks kinds of entities that can do it, and vice versa.” (Machamer, Darden, and Craver, page 17)

Mechanisms, therefore, provide a way to solve the pairing problem, and they leave a multitude of traces that can be uncovered if a hypothesized causal relation really exists. For example, those who want to subject Max Weber’s hypothesis about the Reformation leading to capitalism do not have to rest content with simply correlating Protestantism with capitalism. They can also look at the detailed mechanism he described for how this came about, and they can look for the traces left by this mechanism (Hedstrom and Swedberg, 1998, page 5; Sprinzak, 1972).²⁸

Multiple Causes and Mechanisms – In this section, the need to solve the problem of preemption and the pairing problem led to a consideration of mechanisms. Many different authors have come to a similar conclusion about the need to identify mechanisms (Cox, 1992; Simon and Iwasaki, 1988; Freedman, 1991; Goldthorpe, 2001), and this approach seems commonplace in epidemiology (Bradford Hill, 1965) where debates over smoking and lung cancer or sexual behavior and AIDS have been resolved by the identification of biological mechanisms that link the behaviors with the diseases.

1.4 Summary of Four Theories of Causality

Four causal theories and two fundamental features of causality have been described. One of the features is the symmetric association between causes and effects. The other is the asymmetric fact that causes produce effects, but not the reverse. Table 1.1 summarizes how each theory identifies these two aspects of causality.

²⁸ Hedstrom and Swedberg (1998) and Sorenson (1998) rightfully criticize causal modeling for ignoring mechanisms and treating correlations among variables as theoretical relationships. But it might be worth remarking that causal modelers in political science have been calling for more theoretical thinking (Achen, 1983, Bartels and Brady, 1993) for at least two decades, and a constant refrain at the annual meetings of the Political Methodology Group has been the need for better “micro-foundations.”

Regularity and counterfactual theories do better at capturing the symmetric aspect of causation than its asymmetric aspect. Regularity theories rely upon the constant conjunction of events and temporal precedence to identify causes and effects. Their primary tool is essentially the “Method of Concomitant Variation” proposed by John Stuart Mill in which the causes of a phenomenon are sought in other phenomena which vary in a similar manner. Counterfactual theories rely upon elaborations of the “Method of Difference” to find causes by comparing instances where the phenomenon occurs and instances where it does not occur to see in what circumstances the situations differ. Counterfactual theories suggest searching for surrogates for the closest possible worlds where the putative cause does not occur to see how they differ from the situation where the cause did occur. This strategy leads naturally to experimental methods where the likelihood of the independence of assignment and outcome, which insures one kind of closeness, can be increased by rigid control of conditions or by randomly assigning treatments to cases. None of these methods is foolproof because none solves the pairing problem or gets at the connections between events, but experimental methods typically offer the best chance of achieving closest possible worlds for comparisons.

Causal theories that emphasize mechanisms and capacities provide guidance on how to solve the pairing problem and how to get at the connections between events. And the growing interest in mechanisms in the social sciences (Hedstrom and Swedberg, 1998) is providing a basis for opening up the black box of the Humean regularity and the counterfactual theories.

The other major feature of causality, the asymmetry of causes and effects, is captured by temporal priority, manipulated events, and the independence of causes. Each notion takes a somewhat different approach to distinguishing causes from effects once the unconditional association of two events (or sets of events) has been established. Temporal priority simply identifies causes with the events that came first. If growth in the money supply reliably precedes economic growth, then the growth in the money supply is responsible for growth. Manipulation theories identify the manipulated event as the causally prior one. If a social experiment manipulates work requirements and finds that greater stringency is associated with faster transitions off welfare, then the work requirements are presumed to cause these transitions. Finally, one event is considered the cause of another if a third event can be found that satisfies the INUS conditions for a cause and that varies independently of the putative cause. If being job-ready varies independently of being in a location where jobs are available, and both satisfy the INUS conditions for getting off welfare, then both must be causes of getting off welfare. Or if education levels of voters vary independently of their getting the butterfly ballot, and both satisfy INUS conditions for mistakenly voting for Buchanan instead of Gore, then both must be causes of those mistaken votes.

Chapter 2 Experimental and Non-Experimental Approaches For Determining Causality

The preceding chapter outlined four theories of causality. In this chapter, we consider the degree to which various designs meet the standards of these theories. We start with experimental methods because they are often considered the “gold-standard” for making causal inferences. We show why experimental methods are considered so powerful, but we also discuss their limitations. In addition, we explore alternatives to experimental methods. We begin by defining what we mean by an experiment.

Experiments involve observing the impact of some manipulation, often called the “treatment,” under strict conditions that control for other factors that might affect the outcome of the experiment. In classic scientific experimentation as many circumstances as possible are physically controlled so that the observed outcome must be the impact of the treatment. Often the outcome produced by the treatment is compared to either an implicit or explicit baseline situation called the “control.” Thus, the outcome of an experiment in which some portion of the brain is stimulated electrically is compared with the situation where no stimulation was present. Or the outcome of an experiment in which a catalyst is introduced into a beaker is compared with the well-known (typically inert) results where no catalyst is present. In these circumstances, the only significant difference between the treatment and the control is the cause so that any differences in the outcome must be the result of the cause.

In many circumstances, however, it is almost impossible to control all of the factors that might affect the outcome. In welfare experiments, for example, there is no way to find two identical families and to provide one with the treatment – such as some new welfare program – and the other without the treatment. There are just too many ways that families can differ. The statistician R.A. Fisher proposed a solution to this problem in the 1920s (Fisher, 1935). He suggested that some fraction, typically half, of the experimental units (families in our example and agricultural plots in his experiments) should be randomly assigned to treatment. And, ideally, there should be a great number of these units. With random assignment to treatment there is no reason to suspect that the entities that get the treatment are any different, on average, from those control entities that do not. Consequently, if they exhibit differences (on average) on the outcome of interest such as length of welfare spell or size of plants at harvest time, then these differences must be due to the treatment. Before showing how randomization satisfies both the counterfactual and manipulation theories of causation, it is worth describing some welfare experiments to provide some examples of what can be done with experiments.

2.1 Examples of Welfare Experiments

The use of experiments as a way to get estimates of the effect of a social policy began back in the 1960’s. The U.S. Office of Economic Opportunity sponsored the first negative income tax experiment in 1968. A negative income tax is a subsidy for low-income employment. A national negative income tax program was not considered politically feasible unless concerns about the work disincentive effects could be alleviated. In total there were four Negative Income Tax Experiments undertaken. The most comprehensive and well known of these was the Seattle and Denver Income Maintenance Experiments (SIME/DIME).

One of the most well known randomized experiments done in California was the study of the Greater Avenues for Independence (GAIN) program. The Family Support Act of 1988 created the federal Job Opportunities and Basic Skills Training (JOBS) Program. California implemented the JOBS program with the GAIN program beginning in 1989. The Manpower Demonstration Research Corporation conducted an evaluation of the GAIN program beginning in 1988. The study was conducted in six of California's 58 counties. The GAIN program was county administered, and thus the programs differed among the six research counties. Participants were randomly assigned to one of two treatment statuses after attending an orientation meeting. The experimental group was given access to the services of the GAIN program and the control group was not. GAIN serviced participants in two ways. Counties separated GAIN participants into two groups, those in need of basic education and those who were deemed job ready. Those in need of education were sent to classes. Participants who were job ready were sent to job placement. The control group was not subject to the mandatory GAIN participation requirements, but could seek similar services from other programs.

Under a federal waiver, California also undertook a project to assess the effects of changes in Aid to Families with Dependent Children (AFDC). In December 1992, the Assistance Payments Demonstration Project began. This project continued and after July 1993 became the Work Pays Demonstration Project. The California Work Pays Demonstration Project selected 30,000 cases in four California counties. Experimental cases had the same treatment as non-research cases and were given reductions in aid amounts and more lenient allowances for work. Control cases had their aid payments frozen at December 1992 levels.

California also undertook an evaluation of the Cal-Learn program beginning in 1994. The Cal-Learn program was designed to provide incentives for pregnant and parenting teens to return to or stay in school. The Cal-Learn evaluation used a two-way factorial research design involving 3 treatment states and 1 control state. The Cal-Learn program involved two main components, financial incentives and case management services. Teens could receive either bonuses and sanctions or case management or both treatments. The factorial research design allowed for analysis of the effects of each intervention separately as well as in combination. This meant that policy makers did not have to only consider doing one treatment or the other but also could look at how the two components could work with each other.

2.2 Assessing Randomized Experiments

Randomized experiments are excellent research designs because they provide a setting in which counterfactual possibilities can be clearly stated and in which these counterfactual outcomes can be tested while ruling out confounding factors. Experiments do this by manipulating the treatment through randomization which insures that on average the outcome, conditional on the treatment, is independent of all other factors that affect the outcome of the experiment. Thus, experiments have two features which seem appealing for making inferences. First the cause is actively manipulated to create the experimental and control groups. Second, the simultaneous existence of an experimental and control group means that there are two "possible worlds" side-by-side – one in which the experimental entities get the treatment and another in which they do not. Moreover, with proper randomization, the characteristics of these two worlds should be, on

average, the same. In the following section, we show how this provides a way of determining the truth or falsity of the counterfactual that is at the center of the counterfactual theory of causality.

Controlled Experiments and Closest Possible Worlds – As we showed in Chapter 1, the difficulties with the counterfactual definition are identifying the characteristics of the closest possible world in which the putative cause does not occur and finding an empirical surrogate for this world. For the butterfly ballot discussed in that chapter, sheer luck led a team of researchers to discover that the absentee ballot did not have the problematic features of the butterfly ballot.¹ But how can we find surrogates in other circumstances?

One answer is controlled experiments. Experimenters can create mini-closest-possible worlds by finding two or more situations and assigning putative causes (called “treatments”) to some situations but not to others (which get the “control”). If in those cases where the cause *C* occurs, the effect *E* occurs, then the first requirement of the counterfactual definition is met: when *C* occurs, then *E* occurs. Now, if the situations which receive the control are not different in any significant ways from those that get the treatment, then they can be considered surrogates for the closest possible world in which the cause does not occur. If in these situations where the cause *C* does not occur, the effect *E* does not occur either, then the second requirement of the counterfactual definition is confirmed: in the closest possible world where *C* does not occur, then *E* does not occur. The crucial part of this argument is that the control situation, in which the cause does not occur, must be a good surrogate for the closest possible world to the treatment. Experiments provide a way of insuring that this will be so.

*Problems with the Counterfactual Definition*² – The counterfactual definition of causation leads to substantial insights about causation, and it provides useful recipe that can be implemented through experimentation, for determining when causation has occurred. It also leads to two significant problems. Using the counterfactual definition the direction of causation cannot be established, and two effects of a common cause can be mistaken for cause and effect. Consider, for example, an experiment as described above. In that case, in the treatment group, when *C* occurs, *E* occurs, and when *E* occurs, *C* occurs. Similarly, in the control group, when *C* does not occur, then *E* does not occur, and when *E* does not occur, then *C* does not occur. In fact, there is perfect observational symmetry between cause and effect which means that the counterfactual definition of causation as described so far implies that *C* causes *E* and that *E* causes *C*. The same problem arises with two effects of a common cause because of the perfect symmetry in the situation. Consider, for example, a rise in the mercury in a barometer and thunderstorms. Each is an effect of high pressure systems, but the counterfactual definition would consider them to be causes of one another.³

¹ For the story of how the differences between the election day and absentee ballot were discovered, see Brady et al, 2001a.

² This section relies heavily upon Hausman, 1998, especially Chapters 4-7 and Lewis, 1973b.

³ Thus, if barometric pressure rises, thunderstorms occur and vice-versa. Furthermore, if barometric pressure does not rise, then thunderstorms do not occur and vice-versa. Thus, by the counterfactual definition, each is the cause of the other. (To simplify matters, we have ignored the fact that there is not a perfectly deterministic relationship between high pressure systems and thunderstorms.)

These problems bedevil Humean and counterfactual theories. If we accept these theories in their simplest forms, we must live with a seriously incomplete theory of causation that cannot distinguish causes from effects and that cannot distinguish two effects of a common cause from real cause and effect. That is, although the counterfactual theory can tell whether two factors *A* and *B* are causally connected⁴ in some way, it cannot tell whether *A* causes *B*, *B* causes *A*, or *A* and *B* are the effects of a common cause (sometimes called spurious correlation). The reason for this is that the truth of the two counterfactual conditions amounts to a particular pattern of the crosstabulation of the two factors *A* and *B*. In the simplest case where the columns are the absence or presence of the first factor (*A*) and the rows are the absence or the presence of the second factor (*B*), then the same diagonal pattern is observed for situations where *A* causes *B* or *B* causes *A*, or for *A* and *B* being the effects of a common cause. In all three cases, we either observe the presence of both factors or their absence. It is impossible from this kind of symmetrical information, which amounts to correlational data, to detect causal asymmetry or spurious correlation. The counterfactual theory, like the Humean regularity theory, only describes a necessary condition, the existence of a causal connection between *A* and *B*, for us to say that *A* causes *B*.

Solving the Problems of Causal Direction and Common Cause Through Manipulation –

Requiring temporal precedence can solve the problem of causal direction by simply choosing the phenomenon that occurs first as the cause, but it cannot solve the problem of common cause because it would lead to the ridiculous conclusion that since the mercury rises in barometers before storms, this upward movement in the mercury must cause thunderstorms. For this and other reasons, David Lewis rejects using temporal precedence to determine the direction of causality, but his solution to the problem is not very convincing. Moreover, from the perspective of a practicing researcher, temporal precedence seems to be a much easier way to establish the direction of causation. But it has its own limitations including the difficulty of identifying what comes before what in many situations. Sometimes this is just the difficulty of measuring events in a timely fashion – when, for example, did Protestantism become fully institutionalized and did it precede the institutionalization of capitalism? Does the increase in the money supply really precede economic upturns?⁵

But identifying what comes before what can also involve deep theoretical difficulties regarding the role of expectations (Sheffrin, 1983), intentions, and human decision-making. Consider, for example, the relationship between educational attainment and marriage timing. “Among women who leave full-time schooling prior to entry into marriage, there are some who will leave school and then decide to get married and others who will decide to get married and then leave school in anticipation of the impending marriage.” (Marini and Singer, 1988, page 377). In both cases, leaving school will precede marriage, but in the first case leaving school preceded the decision to marry and in the second case leaving school came after the decision to get married. Thus the timing of observable events cannot always determine causality, although the timing of intentions (to marry in this case) can determine causality. Unfortunately, it may be hard to get data on the

⁴ As implied by this paragraph, there is a causal connection between *A* and *B* when either *A* causes *B*, *B* causes *A*, or *A* and *B* are the effects of a common cause. (See Hausman, 1998, pages 55-63).

⁵ The appropriate lag length in the relationship between money and economic output continues to be debated in economics, and it has led to the “established notion that monetary policy works with long and variable lags (Abdullah and Rangazas, 1988, page 680).”

timing of intentions. Finally, there are philosophical qualms about using temporal precedence to determine causal priority. Clearly, from a practical and theoretical perspective, it would be better to have a way of establishing causal priority that did not rely upon temporal precedence.

One solution is to have a clear-cut intervention that manipulates the putative cause. An intervention simplifies the job of establishing causal priority by appeal to the manipulation theory of causation. It also provides a way, especially in the experimental context where the intervention is random assignment of treatment, to reject the possibility that the supposed cause and effect are merely the result of a common cause. The cause must be the manipulation.⁶

The Strengths of Experiments – The combination of intervention and control in experiments makes them especially effective ways to identify causal relationships. If experiments only furnished closest possible worlds, then the direction of causation would be indeterminate without additional information. If experiments only manipulated factors, then accidental correlation would be a serious threat to valid inferences about causality. Both features of experiments do substantial work.

Any approach to determining causation in non-experimental contexts that tries to achieve the same success as experiments must recognize both these features. Many of the methodologists cited in Chapter 1 (Simon, Cook and Campbell, Mill, Sobel, and Cox) conflated them. There are even psychological experiments which suggest that when considering alternative explanations for some phenomenon, people typically consider nearby worlds in which individual agency figures prominently. When asked to consider what could have happened differently in a vignette involving a drunken driver and a new route home from work, subjects focus on having taken the new route home instead of on the factors that lead to drunken driving. They choose a cause and a closest possible world in which *their* agency matters. But there is no reason why the counterfactual theory and the manipulation theory should be combined in this way. The counterfactual theory of causation emphasizes possible worlds without considering human agency and the manipulation theory of causation emphasizes human agency without saying anything about possible worlds. Experiments derive their strength from combining both theoretical perspectives, but it is all too easy to overlook one of these two elements in generalizing from experimental to observational studies.⁷

The Limitations of Experiments – Experiments satisfy the requirements for two basic approaches to causality, the counterfactual and manipulation theories. Why, then, don't we do them more

⁶ It might be more correct to say that the cause is buried somewhere among those things that were manipulated or that are associated with the manipulation. It is not always easy, however, to know what was manipulated as in the famous Hawthorne experiments in which the experimenters thought the treatment was reducing the lighting for workers but the workers apparently thought of the treatment as being treated differently from all other workers. Part of the work required for good causal inference is clearly describing what was manipulated and unpacking it to see what feature caused the effect.

⁷ Some physical experiments actually derive most of their strength by employing such powerful manipulations that no controls are needed. At the detonation of the first atom bomb, no one doubted that the explosion was the result of nuclear fission and not some other uncontrolled factor. Similarly, in what might be an apocryphal story, it is said that a Harvard professor who was an expert on criminology once lectured to a class about how all social science evidence suggested that rehabilitating criminals simply did not work. A Chinese student raised his hand and politely disagreed by saying that during the Cultural Revolution, he had observed cases where criminals had been rehabilitated. Once again, a powerful manipulation may need no controls.

often? Probably the biggest reason is that they are hard to do, especially in the social sciences. The limitations are both practical and ethical. The practical limitations are the same as those which held-up the first experimental test of Newton's theories of orbiting satellites until almost 300 years after their formulation. Sputnik was expensive and complicated and so are most social and biological experiments. The ethical limitations have to do with the unacceptability of randomly assigning people to families, political parties, or guerilla groups. But there are other limitations as well.

Consider, for example, experimental tests of Boyle's Law which says that Pressure (P) times Volume (V) is proportional to Temperature (T). What would happen if a "Boyle's Law" experiment varied T and measured P but the apparatus was set-up, unbeknownst to the investigator, so that V would adjust. If V adjusted enough, then P might not vary at all. Alternatively, P might vary somewhat as T was adjusted, but V might vary as well. If only P and T were being measured, then the experiment might grossly underestimate the possible impacts of T . There are no violations of physical laws here, and there is no failure of the experimental method. The method would be giving a true and correct rendition of what occurs under the experimental circumstances which just happen to allow V to vary. One of the problems, then, with experiments is that they only tell us what happens under the conditions that happen to exist in the experiment. These conditions may be very specialized and even idiosyncratic.

Now consider an experiment which tries to increase the employment P of people by providing them with some training T requiring substantial study and reading. Suppose that V measures the violence of this group of people, and suppose that the treatment affects some subjects by causing them to become employed (increasing P) but it affects others by getting them frustrated and more violent (increasing V) because they cannot seem to learn. In fact, suppose that for any individual, the relationship among these three variables is exactly the same as the gas law so that PV is always proportional to T , but for some people only P can vary, for others only V can vary, and for still others both can vary but to a different extent depending upon the person. Under these circumstances, the treatment, T , could lead to highly variable employment outcomes depending upon the mix of people in the program. In some instances the program might seem, on average, to get people employed, and in others it might seem to harm them by decreasing their employment possibilities. The inferences in each case would be correct for the population that was studied, but it would be wrong to generalize from it. In fact, if the experimenter had also measured V , then it would become clear that there was a structural, lawlike relationship among P , V , and T .

This example might seem far-fetched, but American social policy has already generalized from a series of experiments that might have been misleading in just this way. Through a series of experiments in California, welfare researchers concluded that a "work-first" welfare program was much better than a "training-first" program. The work-first program was based upon job attachment theory which presumes that welfare recipients have the skills to be good workers, but they have forgotten (or never learned) the habits required to get jobs. Consequently, getting welfare recipients tied to jobs is the best way to move them out of poverty. Job attachment theory has much different implications than human capital theory which supposes that welfare recipients lack the skills to be good workers. According to this theory, "job-training" is needed to provide welfare recipients with the skills they need to get jobs. Recent work by Hotz, Imbens

and Klerman (2001) suggest that work-first programs did well because they were implemented in counties where jobs were plentiful and the welfare recipients had relatively high education and past job experience, but if these same programs had been implemented in counties with a less job-ready population, then work-first might not have been so successful. Indeed, it might just frustrate welfare recipients who try to get jobs even though they are not suited for them.

There are other problems with experiments as well. Heckman (1992) has argued that randomization may affect participation decisions so that the people who get involved in a randomized experiment may differ from those who would get involved in a full-scale program. The assumption that there is no effect of randomization on participation decisions “is not controversial in the context of randomized agricultural experimentation” (p. 227) which is where the Fisher’s experimental model was developed. This model is the intellectual basis for modern social experiments, but it may require some modification with human subjects. Heckman also argues that experiments are “inflexible vehicle for predicting outcomes in environments different from those used to conduct the experiment.” (p. 227).

Nevertheless, randomized experiments are still the gold standard for making valid inferences. LaLonde (1986) and Fraker and Maynard (1987) have shown that when experimental data is analyzed using standard observational methods, the results are quite different, and there are good reasons to believe that the experimental methods are typically more trustworthy. Although critics (Heckman and Hotz, 1989) using better observational methods provide evidence that “tempers the recent pessimism about non-experimental evaluation procedures” (p. 863), experimental methods still seem like a powerful way to make reliable inferences. In fact, a recent report that asked whether non-experimental methods could match experimental methods in the evaluation of welfare-to-work programs concluded that the best non-experimental methods did not work well enough to replace random assignment (Bloom, Michalopoulos, Hill, and Lei, 2002).

2.3 Observational Studies

There are cases when experiments are either impractical or unethical. When this is true, there is no other choice but to use observational studies. Observational studies can still produce reliable inferences if we are careful to consider alternative causes and to rule out competing explanation. Statistical methods, such as regression analysis, have been developed to make it possible to compare outcomes from situations with the treatment to those without the treatment. These methods, described in detail in Chapter 3, try to “control” for all other observed factors that might affect the outcome. But there are substantial perils from using these methods. In regression analysis, for example, the left-hand side variable is the outcome variable Y^* and the right-hand-side variables are the control variables (or covariates) X^* and some measure of the treatment and control such as a dummy variable D^* . The use of this technique has led to two difficulties. First, unlike correlation analysis, which is inherently symmetrical, regression analysis is inherently asymmetrical. One variable has to be chosen as the left-hand-side or “dependent” variable and others have to be chosen as the right-hand-side or “independent” variables. It is all too easy for researchers to fall into the easy assumption that the left-hand-side variable is the effect of the right-hand-side causes. Yet, once outside of the experimental paradigm, there is no guidance about which variable should be considered the outcome or effect.

Observational studies very seldom have any built-in asymmetry that suggests the proper dependent variable, but experiments do have this asymmetry which comes from one variable being manipulated. Second, all the right-hand-side variables are treated symmetrically in regression. Yet, the experimental framework treats covariates and treatments asymmetrically. If the requirements for good causal inference hold for the assignment of the treatment D^* and the outcome Y^* when the covariates X^* are controlled (perhaps through random assignment), it does not follow that we can interchange D^* and X^* . Once again, it is important to recognize that experiments identify putative causes through their manipulation of them.

Where does this leave observational studies? One, rather weak, answer is that we have no choice but to use them to answer many questions for which experiments are either impractical or unethical. A better answer is that observational studies can still produce reliable inferences if we are careful to consider alternative causes and to rule out competing explanations. The basic tool for this is disciplined comparisons where we try to find as many ways as possible to compare one situation with another in order to rule out competing explanations. We can offer a number of tools for improving this process. They range from better theory (models that provide mechanisms and explanations), through better research design (thinking about the inference problem and employing natural experiments and matching), to improved model-building (better model selection, more concern with model uncertainty, and model replication through the use of multiple data-sets). Chapter 3 considers the issues regarding improved model-building. Here we focus on better theory and better research design.

Better Theory: Models that provide mechanisms and explanations – One of the major flaws in many observational studies (and experiments as well) is that there is often very little theory to help guide the inferential task. Yet, most observational studies must make a passel of assumptions – what variables to include, the functional form of relationships, the way error enters the model – that can affect the ultimate inference. One of the best things that social scientists can do is to develop better models that will provide guidance about these decisions. These models should pay special attention to the “social mechanisms” (Hedstrom and Swedberg, 1998) that generate and explain events. At the simplest level, this means that researchers should not be happy with regression “models” that simply throw variables into a regression. It is nowhere near enough to know that job training programs increase the work effort of welfare recipients, that the possession of civic skills increases political participation, or that proportional voting systems increase the number of political parties. Researchers must seek to understand the exact mechanisms by which training increases work effort, civic skills increase participation, and proportional voting methods lead to more parties. These mechanisms must include detailed descriptions of the decision-making problem facing individuals and the way that they solve this problem. For example, if some candidates in American presidential primaries gain “momentum” from winning early primaries (Bartels, 1988), then models of the individual level processes (e.g., increased name recognition or strategic voting) that might lead to momentum should be developed (Brady, 1996) and experiments should be undertaken to see whether these processes actually occur.

This call for better theory may seem utopian, but social scientists have developed theories that help guide the research process. Political scientists, for example, have gained considerable understanding of electoral systems not only through detailed empirical studies (Lijphart, 1994)

but also through sophisticated modeling (Cox, 1997) which helps to explain empirical regularities. Sociologists and others have developed theories of mass political behavior (Lichbach, 1995, 1996) which explain the actions of rebel's and cooperators alike. Economic theory provides guidance about both macro-economic and micro-economic phenomena. And labor economists and sociologists now know a lot more about the mechanisms that might affect getting on welfare, getting off welfare, and getting jobs.

Better Research Design: Thinking about the inference problem – Researchers can never worry enough about the validity of their inferences. In his book on *How Experiments End* (1987), Peter Galison argues that experiments end when researchers believe they have a result that will stand up in court because they cannot think of any credible ways to make it go away. A lot of the work of inference is trying to think of ways to make results go away, and researchers should think hard about this before, during, and after a study. Researchers who never experience sleepless nights worrying “what if I am wrong?” should probably rethink their research strategies.

General frameworks for thinking about inference can help to generate lists of generic threats to inference. Fisher's classic *The Design of Experiments* (1935) is all about setting up experiments in ways that will make the results stand up in court. The classic handbook for observational studies, Campbell and Stanley's *Experimental and Quasi-Experimental Designs for Research* (1966, see also Cook and Campbell, 1979), provides an extraordinarily fertile list of threats to validity for many different kinds of research designs. All researchers should be familiar with the Campbell-Stanley-Cook lists.

In the past 25 years, Rubin and his collaborators (Rubin, 1974, 1978, 1990; Holland and Rubin, 1988; Holland, 1986; Rosenbaum and Rubin, 1983) have developed an elegant generalization of the Neyman framework for inference that covers experiments and observational studies. The central focus of this work has been a careful explication of the assignment or selection method (see also Heckman, 1979, Heckman and Robb, 1985). This framework has led to concrete methods for improving causal inference such as the use of propensity scores for matching (Rosenbaum and Rubin, 1983) and the analysis of the conditions under which path modeling can be successful (Holland, 1988). Every empirical researcher should become familiar with this framework. Manski (1993, 1995) has explored what can be inferred from observational studies when there are problems of extrapolation, selection, simultaneity, mixing, or reflection. An understanding of these generic problems should also be part of every researcher's tool-kit.

Familiarity with this literature should enable researchers to develop better designs for their research which control for some of the major threats to valid inferences. Time-series studies, for example, make it possible to determine whether putative causes really change before their supposed effects. Time-series cross-sectional studies add the possibility of comparing across different units to see if the same results occur. Life-history data provide more and better controls for individual differences. None of these designs is foolproof, but they can provide some confidence that major sources of confounding have been controlled.

2.4 Quasi-Experiments

When full experimental control is not possible, Thomas Cook and Donald T. Campbell recommend “quasi-experimentation,” in which “an abrupt intervention at a known time” in a treatment group makes it possible to compare the impacts of the treatment over time or across groups (Cook and Campbell, 1986, page 149). The success of quasi-experimentation depends upon “a world of probabilistic multivariate causal agency in which some manipulable events dependably cause other things to change.” (Page 150). A quasi-experiment is similar to an experiment because the putative cause is manipulated for some entities but not for some others, but it is different from an experiment because there is no random assignment of entities to the manipulation.

With quasi-experiments, the analyst cannot rule out pre-existing differences between those entities that get the treatment and those that do not get it. Consequently, care must be taken to control for such differences in order to obtain valid measures of causal effects. One way to do this is to use difference-in-difference estimators.

Difference-in-difference estimates successively take two differences of the outcome measure when data are observed on multiple entities, some of which get the treatment and some of which do not, over time periods which are before and after the introduction of the manipulation. The first difference is between the outcome for each unit after the time when the manipulation is introduced and the outcome before the manipulation is introduced. It is simply the before and after comparison for each unit. For those units that get the manipulation, this difference might be considered the impact of the manipulation because it describes what kinds of changes occurred for each unit on the outcome variable before and after the manipulation. But it is also possible that some other factors might have been operating on both the treated and untreated units so that the outcome variable might have changed for both sets of units, whether or not they received the treatment. To adjust for this possibility, a second difference is taken between the first difference for the treated and the first difference for the untreated units. If the other factors operate equally on both the treated and untreated units, then this second difference will subtract them out, leaving the true impact of the treatment. This method is clearly not foolproof when, for example, the other factors act differently on the treated and the untreated units or when there is little temporal variation in the outcome variable, but it can adjust for certain types of confounding factors.

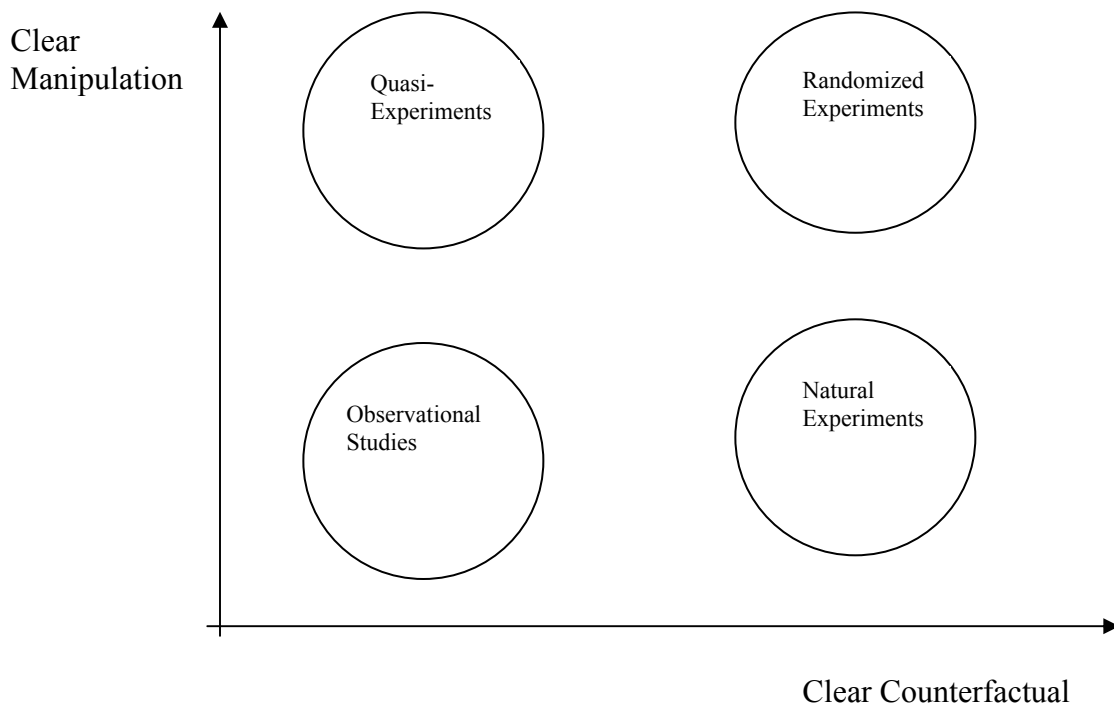
2.5 Natural Experiments

Better Research Design: Employing natural experiments and using matching – Partly as a result of doubts about observational studies, researchers have increasingly looked for “natural experiments” in which essentially random events provide some inferential leverage akin to randomized experiments. For example, the Vietnam War draft lottery randomly selected some people to enter the military which makes it possible to determine the impact of military service on future earnings (Angrist, 1990), and miscarriages occur almost at random so that they can be used to study the consequences of teenage childbearing on mother’s incomes (Hotz, Mullin, Sanders, 1997). This approach has been used to determine the consequences of workers’ compensation on injury duration (Meyer, Viscusi, and Durbin, 1995), past disenfranchisement on

future voting (Firebaugh and Chen, 1995), uncertainty on decision-making (Metrick, 1995), ballot form on voting choices (Wand, Shotts, Sekhon, Mebane, Herron, and Brady, 2001), the draft lottery and veteran status on lifetime earnings (Angrist, 1990), the minimum wage on total employment (Card and Krueger, 1994), the number of children in a family on their future life prospects (Rosenzweig and Wolpin, 1980), and political parties on voting behavior in the U.S. and Confederate Houses during the Civil War (Jenkins 1999).

Figure 2.1 provides a summary of the relationship among randomized experiments, quasi-experiments, natural experiments, and observational studies. Randomized experiments have both a clear manipulation and clear counterfactual through randomization. Observational studies often have neither. Quasi-experiments rely upon real manipulations, but there can be substantial uncertainty about the proper counterfactual. Finally, natural experiments, because they rely upon cases where random events have occurred, typically have clear counterfactuals, but the manipulation, because it is not truly under the control of the researcher, may be far from what the researcher actually cares about. The Vietnam War lottery, for example, makes it possible to determine the impact of military service on future earnings, but Vietnam War military service is a decidedly different manipulation from peacetime military service which might be the researcher's true interest.

Figure 2.1: Relationship Among Experiments and Observational Studies



2.6 What are the Best Methods for Establishing Causality?

Causal Inference with Experimental and Observational Data – Now that we know what causation is and we have reviewed a number of research designs, what lessons can we draw for doing empirical research? Table 1.1 in Chapter 1 shows that each theory provides sustenance for different types of studies and different kinds of questions. Regularity and mechanism theories

tend to ask about the causes of effects while counterfactual and manipulation theories ask about the effects of imagined or manipulated causes. The counterfactual and manipulation theories converge on experiments, although counterfactual thought experiments flow naturally from the possible worlds approach of the counterfactual theory. Regularity theories are at home with observational data, and the mechanical theory thrives on analytical models and case studies.

Which method, however, is the best method? Clearly the gold-standard for establishing causality is experimental research, but even that is not without flaws. When they are feasible, well done experiments can help us construct closest possible worlds and explore counterfactual conditions. But we still have to assume that there is no preemption occurring which would make it impossible for us to determine the true impact of the putative cause, and we also have to assume that there are no interactions across units in the treatment and control groups and that treatments can be confined to the treated cases. If, for example, we are studying the impact of a skill training program on the tendency for welfare recipients to get jobs, we should be aware that a very strong economy might preempt the program itself and cause those in both the control and treatment conditions to get jobs simply because employers did not care much about skills. As a result, we might conclude that skills do not count for much in getting jobs even though they might matter a lot in a less robust economy. Or if we are studying electoral systems in a set of countries with a strong bimodal distribution of voters, we should know that the voter distribution might preempt any impact of the electoral system by fostering two strong parties. Consequently, we might conclude that single-member plurality systems and proportional representation systems both led to two parties, even though this is not generally true. And if we are studying some educational innovation that is widely known, we should know that teachers in the “control” classes might pick-up and use this innovation thereby nullifying any effect it might have.

If we add an investigation of mechanisms to our experiments, we might be able to develop safeguards against these problems. For welfare recipients, we could find out more about their job search efforts, for party systems we could find out about their relationship to the distribution of voters, and for teachers we could find out about their adoption of new teaching methods.

Once we go to observational studies, matters get much more complicated. Spurious correlation is a real danger. There is no way to know whether those cases which get the treatment and those which do not differ from one another in other ways. It is very hard to be confident that either independence of assignment and outcome or conditional independence of treatment and assignment holds. Because nothing has been manipulated, there is no surefire way to determine the direction of causation. Temporal precedence provides some information about causal direction, but it is often hard to obtain and interpret it.

The Causality Checklist – Table 2.2 summarizes the questions that a researcher should ask about causality. Not every question can be answered for every study. But a careful review of them can help to refine causal thinking. Some of the questions are general ones that merely ask the researcher to be clear about the events and inferences that are being considered. Others ask about the symmetric and asymmetric aspects of causation. Seldom will any specific study be able to answer all of them, but it would be worrisome if a research program failed to address almost all of them at one time or another. In Chapter 7, we shall review this checklist to see how well our own research has met the standards set out in this and the preceding chapter.

Table 2.2 Causality Checklist

General Issues

- What is the “cause” (C) event? What is the “effect” (E) event?
- What is the exact causal statement of how C causes E?
- What is the corresponding counterfactual statement about what happens when C does not occur?
- What is the causal field? What is the context or universe of cases in which the cause operates?
- Is this a physical or social phenomenon or some mixture?
- What role, if any, does human agency play?
- What role, if any, does social structure play?
- Is the relationship deterministic or probabilistic?

Neo-Humean Theory

- Is there a constant conjunction (i.e., correlation) of cause and effect?
- Is the cause necessary, sufficient or INUS?
- What are other possible causes, i.e., rival explanations?
- Is there a constant conjunction after controls for these other causes are introduced?
- Does the cause precede the effect? In what sense?

Counterfactual Theory

- Is this a singular conjunction of cause and effect?
- Can you describe a closest possible (most similar) world to where C causes E but C does not occur? How close are these worlds?
- Can you actually observe any cases of this world (or something close to it, at least on average)? Again, how close are these worlds?
- In this closest possible world, does E occur in the absence of C?
- Are there cases where E occurs but C does not occur? What factor intervenes and what does this tell us about C causing E?

Manipulation Theory

- What does it mean to manipulate your cause? Be explicit. How would you describe the cause?
- Do you have any cases where C was actually manipulated? How? What was the effect?
- Is this manipulation independent of other factors that influence E?

Mechanism and Capacities Theories

- Can you explain, at a lower level, the mechanism(s) by which C causes E?
- Do the mechanisms make sense to you?
- What other predictions does this mechanism lead to?
- Does the mechanism solve the pairing problem?
- Can you identify some capacity that explains the way the cause leads to the effect?
- Can you observe this capacity when it is present and measure it?
- What other outcomes might be predicted by this capacity?
- What are possible preempting causes?

Chapter 3 Estimation, Forecasting, and Model Uncertainty¹

3.1 Estimation, Model Selection, and Hypothesis Testing

3.1.A Estimating Programmatic Impacts

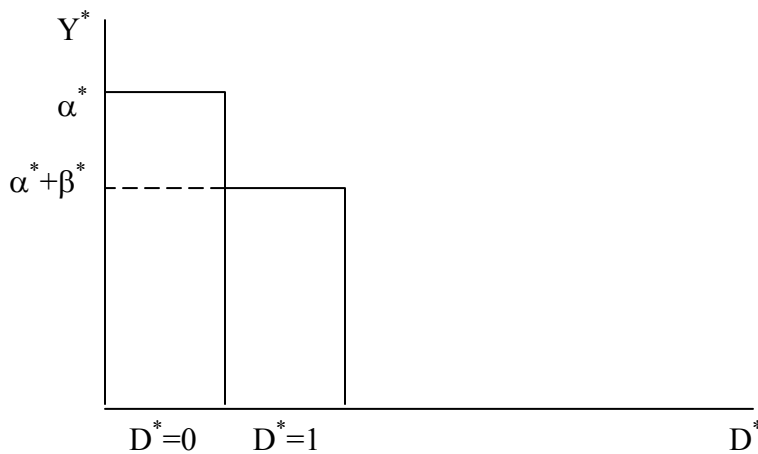
Model estimation is generally used to determine if there exists a *causal* relationship between two variables, and where one exists, the direction and significance of that relationship.

Suppose that we want to examine the impact of a particular treatment (D^*) such as job training on welfare caseloads (Y^*). We expect that the implementation of job training for welfare recipients will reduce the caseload by increasing job skills and hence employment prospects for participants. We can examine the impact by comparing time periods in which treatment was not imposed to time periods in which treatment was imposed. Our postulated bivariate² model could take the following linear form:

$$(1) Y^*_t = \alpha^* + \beta^* D^*_t + \varepsilon^*_t$$

where Y^* is our dependent variable, D^* is the exogenous (i.e., independent) treatment variable, and ε^* is a stochastic error term. The treatment D^* is a 0/1 variable: it is equal to 1 after job training treatment is implemented and 0 before the treatment is implemented. The subscript t indexes the observations across time. The parameters α^* and β^* are estimated by the model. The α^* parameter is the intercept while the parameter β^* measures the *causal effect* of the treatment D^* on welfare caseloads (Y^*).³

Figure 3.1



¹ This chapter is based on the following econometric texts: Pindyck & Rubinfeld (1991), Greene (1997), Judge et al. (1988) and Chatfield (JASSA 1995).

² Bivariate indicates that there are only two regressors: a constant and one additional explanatory variable.

³ Alternatively, we could have variation across individuals (i) such that $Y^*_i = \alpha^* + \beta^* D^*_i + \varepsilon^*_i$. In this case, the comparison would be welfare participation across two distinct groups: one treated group which received job training ($D^*_i=1$) and a control group which did not receive job training ($D^*_i=0$). The impact (β^*) would still measure the impact of the job training program, but would use variation across individuals rather than variation across time to identify the effect.

If the data are comprised of state level caseload levels across time, then the parameter β^* measures the change in caseloads after the implementation of the job training treatment. In Figure 3.1, the treatment has a negative impact on caseloads (Y^*) as caseload levels are reduced by β^* after the treatment is implemented. If there were no other factors affecting the caseload during this time period, then we could say that this effect was *causal*. In Section 3.2.B, we will discuss how to estimate the impact of a program when there are additional factors which may also affect the outcome of interest.

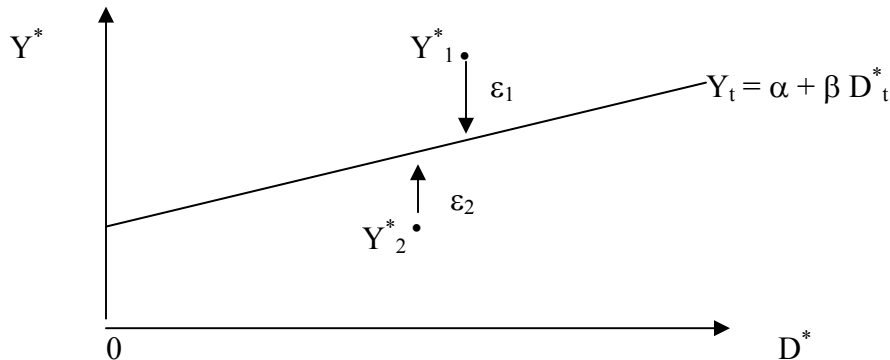
3.1.B OLS: Minimizing the Sum of Squared Residuals

The most common estimation technique for this type of model is ordinary least squares (OLS). OLS estimates the true parameters (α^* , β^*) of the model by minimizing the sum of squared residuals (SSR). The estimated parameters, (α , β), are then used to form the predicted values for caseloads (Y). The predictions are calculated using the estimated parameters and the actual values of the explanatory variable (job training) so that $Y_t = \alpha + \beta D_t^*$. The residual (ε) is defined as the difference between the actual (Y^*) and predicted caseloads (Y). The number of observations in the sample is N .

$$\text{Minimize SSR} = \sum_{t=1}^N \varepsilon_t^2 = \sum_{t=1}^N (Y_t^* - Y_t)^2$$

As shown in Figure 3.2, the model's predicted caseloads (Y) lie on the line $Y_t = \alpha + \beta D_t^*$, where α and β are the parameters estimated by OLS. The residuals measure the distance from the actual values (Y_1^*, Y_2^*) and their corresponding predicted values (Y_1, Y_2) which lie on a line.

Figure 3.2



Although it is possible to use a variety of distance measures, the square of the residual is used in order to avoid situations (as shown in Figure 3.2) where positive and negative residual values would eliminate each other. By taking into account both over- and under-predictions symmetrically, OLS minimizes the overall prediction error. Using the sum of the square of the residual, rather than the sum of the absolute values of the residuals, OLS overweights the large residuals in the minimization process.

Using this distance minimization technique, the parameters are estimated as follows.

The slope coefficient (β^*) measures the covariance of the caseloads and job training variables and reports the correlation normalized by the variance in the caseloads. In our model, this slope coefficient reports the impact of implementing job training on the welfare caseload.

$$\beta = \text{Cov}(D^*, Y^*) / \text{Var}(Y^*)$$

The intercept or constant term (α^*) is estimated as the difference between the expected value of the caseload, $E(Y^*)$, and the portion predicted by the slope coefficient multiplied by the expected value of the independent variable, $\beta E(D^*)$.

$$\alpha = E(Y^*) - \beta E(D^*)$$

Since OLS minimizes the sum of square residuals, we expect the residuals from this estimation to be zero. If the assumptions of OLS hold, we also expect the residuals to be white noise – that is, they are distributed without any discernible pattern.

However, in graphing the residuals, we may find that there is a discernible pattern. Often this is the result of a violation of at least one of the OLS assumptions and indicates the presence of heteroskedasticity or serial correlation. Heteroskedasticity will be discussed in Section 3.2.D and serial correlation in Section 3.2.E.

3.1.C OLS: Best Linear Unbiased Estimators

For this linear model, OLS estimates are best linear unbiased estimates (BLUE) if the following assumptions hold:

1. D^* and Y^* are related linearly
2. The values of D^* are non-stochastic⁴
3. The error term is independently and identically distributed (i.i.d.):
 - i. with mean zero and constant variance: $\varepsilon^* \sim (0, \sigma^2)$
 - ii. and uncorrelated errors: $E(\varepsilon^*_t, \varepsilon^*_{ts}) = 0$ for all $t \neq s$
4. The error term and explanatory variable are uncorrelated: $E(D^*, \varepsilon^*) = D E(\varepsilon^*) = 0$

If these assumptions hold, OLS minimizes the sum of squared residuals subject to estimating the true parameters (α^*, β^*) of the model. That is, among the set of possible linear unbiased parameter estimates, OLS estimates are *efficient* in that they have the smallest variance.

Unbiasedness requires that the expected value of the estimated parameters (β) be equal to the true value (β^*). This means that the expected value of our estimated impact (or parameter) is equal to its true value.

$$\text{Unbiasedness}^5: E(\beta) = \beta^*$$

⁴ This assumption can be relaxed such that D^* need only be pre-determined rather than non-stochastic.

⁵ In essence, the distribution of the estimated parameter should collapse to the true value as the sample becomes asymptotically large. A related notion is that of consistency. While some estimators may not be unbiased in small

OLS minimizes the variance of the estimator conditional upon the required characteristic of unbiasedness. That is, given that both β_1 and β_2 are unbiased estimators of β^* , OLS selects β_1 over β_2 if β_1 has a smaller variance:

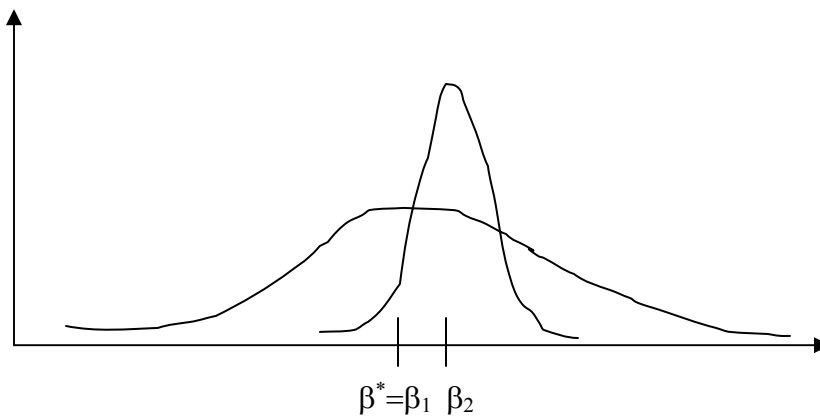
$$\text{Efficiency: } \text{Var}(\beta_1) < \text{Var}(\beta_2).$$

It is important to note that there is a trade-off between unbiasedness and efficiency. The formula for calculating the mean squared error (MSE) of an estimated parameter captures this trade-off. As the bias decreases, the variance increases. And vice versa. The MSE of an estimated parameter is calculated as:

$$\text{MSE}(\beta) = E(\beta - \beta^*)^2 = [\text{Bias}(\beta)]^2 + \text{Var}(\beta)$$

It is possible to have a biased estimator with a smaller variance than an unbiased estimator. In Figure 3.3, β_1 is unbiased because it is centered about the true parameter β^* , but the shape of its distribution indicates that it has a much larger variance than the biased estimator β_2 . It is important to note that while β_2 is biased, more of its distribution is located closer to the true value of the parameter than for the unbiased parameter β_1 . Determining which is the *better* estimate depends on the goal of the estimation. OLS avoids this trade-off by concentrating solely on unbiased estimators.

Figure 3.3



3.1.D Goodness of Fit

Given that the above assumptions hold and OLS is the proper estimation procedure, there is still the issue of model selection. That is, given the linear specification and required assumptions, the decision remains as to which explanatory variables should be included in the model.

samples, the estimators may become consistent as the sample becomes very large (i.e., *asymptotically*). For example, while $[\sum_{i=1}^N (\beta_i + 1)]/N$ is a biased estimator of the true parameter β^* in small samples, it becomes consistent as the sample increases because the bias decreases as the sample size (N) becomes large.

One way to choose among a series of specifications is to compare measures of goodness of fit. One such measure is the R-squared (R^2).

The R^2 measures the amount of the variation in the dependent variable that is explained by variation in the independent variables included as explanatory factors. In our example, it is the percent of variation in caseload over time that is explained by the implementation of the job training program.

$$R^2 = \frac{\sum_{t=1}^N Y_t^2 / \sum_{t=1}^N Y_t^{*2}}{\sum_{t=1}^N Y_t^2} = 1 - \frac{\sum_{t=1}^N \varepsilon_t^2 / \sum_{t=1}^N Y_t^2}{\sum_{t=1}^N Y_t^2}$$

An easy method for increasing the R^2 of a model is to increase the number of explanatory variables in the model. Increasing the number of explanatory variables can never decrease the *unadjusted* R^2 because adding additional variables never decreases the amount of explained variation. As a result, comparing the unadjusted R^2 s for two models favors specifications with more explanatory variables because the unadjusted R^2 calculation does not take into account the number of explanatory variables when determining goodness of fit.

Given this incentive for increasing the number of explanatory variables, the preferred measure of goodness-of-fit is often an R^2 adjusted to penalize for attempts at over-fitting the model. The *adjusted R-squared* reduces the goodness of fit measure based on the number of explanatory variables by correcting for the fewer degrees of freedom.⁶ The adjusted R^2 is calculated as follows, where k is the number of included explanatory variables and N is the number of observations:

$$R_a^2 = 1 - [(N-1)/(N-k)] (1-R^2)$$

Use of the adjusted R^2 allows a reasonable comparison of models with different numbers of explanatory variables.

3.1.E Hypothesis Testing: T-tests and F-tests

In addition to parameter estimates, estimation techniques such as OLS provide a standard error for each estimated parameter. The standard error is an indication of the precision with which parameters are estimated. Standard errors can be used to form confidence intervals and test hypotheses. Two standard hypothesis tests are the t-test and the F-test.

3.1.E.i T-tests

T-tests are the most common form of hypothesis testing and are used to determine the significance of individual parameters. Recall that in our caseload model, the parameter of interest was the coefficient on the job training variable. We want to know whether job training has any impact on welfare caseloads. The estimate of our impact of job training was β .

The null hypothesis (H_0) for a t-test is that the parameter of interest is equal to zero, which would mean that job training has no impact on caseloads. Even though caseloads may have fallen when job training was implemented, job training may not be the cause of the decline because it may have

⁶ The available degrees of freedom ($N-k$) are negatively related to the number of explanatory variables (k).

resulted from some other change, such as an improved economy, that occurred at the same time as the implementation of job training. If so, we will accept the null hypothesis that job training has no impact on caseloads. The alternative hypothesis (H_a) is that job training did affect caseloads so that the parameter is *significantly* different from zero.

$$H_0: \beta = 0$$

$$H_a: \beta \neq 0$$

In order to reject the null hypothesis that job training has no impact, we must find not only a parameter estimate which differs from zero ($\beta \neq 0$), but we must also find that the value of zero does not fall within the confidence interval constructed from the parameter value and the standard error. As we are checking that the parameter differs significantly from zero on both sides of the distribution (i.e., above and below), this is a two-tailed test.

A t-statistic is calculated using both the parameter estimate (β) and the corresponding standard error (s_β); the resulting t-statistic is compared to a critical value from the t-distribution.

$$\text{t-statistic} = (\beta/s_\beta)$$

A confidence interval can be constructed using the estimated parameter value, the standard error of the estimate, and the critical value (t_c) to determine the size of the interval. The formula for the interval is:

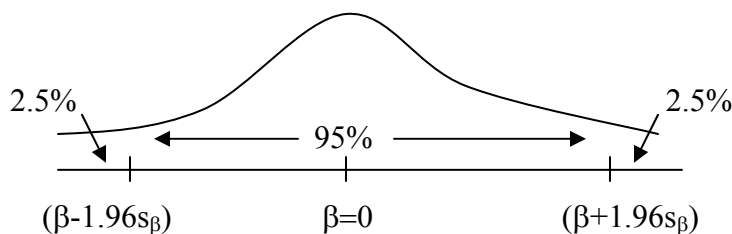
$$\beta \pm t_c s_\beta$$

The standard confidence interval is 95%. In constructing a 95% confidence interval, each tail of the distribution would contain 2.5% of the values. Using a t-distribution, the critical value for a 95% two-tailed confidence interval is 1.96 and the constructed interval is:

$$(\beta - 1.96 s_\beta) < \beta^* < (\beta + 1.96 s_\beta).$$

Figure 3.4 shows the distribution for the estimated value of the parameter under the null hypothesis that $\beta^*=0$. If the estimated value for the parameter lies in either of the 2.5% tails of the distribution, we can reject the null hypothesis that job training has no impact on welfare caseloads.

Figure 3.4



Note that a 95% confidence interval indicates that only 5% of the time would we reject the null when it is true.⁷

$$\text{Probability}[(\beta - 1.96 s_{\beta}) \leq \beta^* \leq (\beta + 1.96 s_{\beta})] = 95\%$$

From this formula it is straightforward to show that a t-statistic greater in absolute value than 1.96 would require us to reject the null hypothesis that the true parameter is equal to zero at the 5% significance level.

3.1.E.ii F-tests

While t-tests allow us to determine the statistical significance of each variable individually, an F-test allows us to test whether all of the explanatory variables excluding the constant are jointly significantly different from zero. An F-test is useful when we want to test the impact of two treatments jointly. Suppose that job training was implemented simultaneously with a decrease in benefits. In that case, we could include both job training and benefit levels in the equation. We could then test the significance of job training and benefit levels separately using a t-test for each or we could test their joint significance using an F-test.

The f-statistic has an f-distribution – it is distributed with k-1 and N-k degrees of freedom, where N is the number of observations in the sample and k is the number of explanatory variables included in the model.

$$f\text{-statistic} = F[k-1 \ N-k] = (R^2/(k-1))/((1-R^2)/(N-k))$$

The null hypothesis for the F-test is that the slope coefficients are jointly zero. As with the t-statistic, a large f-statistic indicates rejection of the null hypothesis. The calculated f-statistic must be larger than the critical value from a table of the f-distribution for the relevant degrees of freedom in order to reject the null.

3.2 Consistency and Efficiency – Specification Error and Violating the Assumptions of OLS

For the following discussion, we assume a multivariate regression specification. This model is similar to Model (1), but includes additional explanatory variables.

$$(2) Y_t^* = \alpha^* + \gamma^* X_t^* + \beta^* D_t^* + \varepsilon_t^*$$

X^* can be a single explanatory variable or a matrix of explanatory variables and γ^* can be a single or a vector of parameters. As before, the index t can be interpreted as indexing time periods.

⁷ That is, false negatives occur only 5% of the time.

3.2.A Inclusion of Irrelevant Variables

In Model (2), we have included additional explanatory variables. In some cases, those variables may not be relevant to the estimation in that they do not explain any of the variation in welfare caseloads. For example, including the rate at which polar ice caps melt could be included in the regression as the second regressor. However, its inclusion will likely not affect the predicted caseload nor the estimation of the impact of job training because the rate at which polar ice caps melt is likely not correlated with caseload or job training.

Including irrelevant variables in the list of explanatory variables does not harm the consistency⁸ of the parameter estimates. Irrelevant variables will generally have insignificant parameter estimates – that is, using a t-test, we cannot reject the null hypothesis that the parameter for the irrelevant variable is zero.

Note, however, that statistically insignificant parameter estimates may also occur if the explanatory variable, though relevant, is highly correlated with other included explanatory variables. This issue is termed multicollinearity and will be discussed in Section 3.2.C.

3.2.B Omitted Variables

The bivariate form (Model (1)) used in Section 3.1.A is over-simplified, but can be easily generalized to a multivariate form of Model (2). In the multivariate form, additional explanatory variables are added. X^* can be a single independent variable or a matrix of independent variables which we postulate may also affect welfare caseloads. For example, we may include economic conditions in X^* such as the unemployment rate. Changes in the unemployment rate may cause welfare caseloads to fluctuate. In order to avoid attributing the fluctuations due to economic conditions to contemporaneous changes to the availability of job training, it is necessary to include the unemployment rate in our estimation of the impact of job training. γ^* is the corresponding parameter measuring the relationship between the economic conditions (X^*) and welfare caseloads.

$$(2) Y_t^* = \alpha^* + \gamma^* X_t^* + \beta^* D_t^* + \varepsilon_t^*$$

Under certain conditions, the inclusion of the variable measuring the economic condition may change the estimate of the impact (β^*) of the job training program. If the implementation of the job training program is correlated with the unemployment rate, then the estimate of the impact of job training from Model (2) may well differ from the parameter estimate from Model (1). For example, if job training programs were implemented at a particular time because economic conditions and job availability had increased, then estimation of the program impact without including controls for economic conditions can overestimate the effect of the program. There would have been a change in welfare caseload due to changing economic conditions - even in the absence of the job training program. Some individuals would have exited welfare for the labor market even without the job training program. Model (1) however attributes the entire change in caseload to the job training program. Model (2) allows us to allocate a portion of the change to

⁸ Recall that consistency is an asymptotic notion relating to the unbiasedness of the estimator.

economic conditions while still estimating the impact of the job training program. Including variables measuring economic conditions in the model is an attempt to construct a better counterfactual by adjusting for those factors that might cause the periods after the implementation of the job training program to differ in relevant ways from the periods before the implementation of the program.

As described above, omitting relevant explanatory variables can cause the estimated parameters to be inconsistent (i.e., biased). The inconsistency of parameter estimates follows if the omitted variable is correlated with any of the included explanatory variables.⁹ The direction and extent of the bias caused by the omission of relevant variables can be determined. For example, if the true model has two relevant explanatory variables (D^* , X^*), but only one variable (D^*) is included in the regression, then the estimated parameter for the included variable (β) would be biased. The size and direction of the bias depends on the correlation between the included (D^*) and omitted (X^*) variables. We can calculate the biased estimated parameter as:

$$\beta = \beta^* + \gamma (\text{Cov}(D^*, X^*) / \text{Var}(D^*))$$

As shown in the above equation, if the omitted and included variables are positively correlated, the parameter estimate for the included variable will be bias upward. However, if they are negatively correlated, the estimate will be biased downward. The size of the bias is positively related to the correlation between the omitted and included variables and inversely related to the variance of the included variable. Unlike other estimation problems, bias due to omitted variables is generally not eliminated as the sample becomes larger.¹⁰

In our example, failure to account for the improvement in the unemployment rate causes an overestimate in the impact of job training on welfare caseloads. Omitting the unemployment rate attributes too much of the decline in welfare caseloads to job training.

3.2.C Multicollinearity and the Inclusion of Lags

In order for OLS to work properly in a multivariate regression setting, the matrix of explanatory variables (D^* , X^*) must have full rank. For example, in a model with two explanatory variables, D^* and X^* , the two variables must not be a linear combination of each other.¹¹ OLS will be unable to estimate parameters for both variables when they are perfectly collinear. If the model contains two perfectly collinear explanatory variables, one of the variables must be removed from the model.

However, in cases where the independent variables are highly, though not perfectly collinear, the solution is less clear. In such cases, the model may be estimated, but the resulting parameters must be interpreted with care. The parameters for the collinear variables are likely to be imprecisely estimated because there is little independent variation across the two variables – that is, it will be difficult to estimate the effect of the treatment D^* holding economic conditions X^* constant because the

⁹ In the rare cases where the omitted and included variables are not correlated, the parameter estimates can be consistently estimated. However, as stated, such a situation is unlikely to occur.

¹⁰ However, there are techniques such as fixed effects, which can eliminate some forms of omitted variable bias. (See Section 3.3.B).

¹¹ That is, they must not be *perfectly collinear*.

collinearity indicates that the data contain little information on this situation. Re-estimating the regression with and without each of the two highly correlated variables will give some indication of the extent of the problem.

Multicollinearity is of particular concern when one or more of the explanatory variables are simply lags of the explanatory variables. Including lags of variables becomes necessary when the impact of a variable is not immediate.

In Model (2), we included the contemporaneous unemployment rate in our regression to control for changing economic conditions. However, suppose we think that improvements in economic conditions take time to affect welfare caseloads. Perhaps it is not only the current unemployment rate, but also the previous four months of changing unemployment that impacts welfare caseloads. Using our standard model, we now include not only the current value of the unemployment rate (X_t^*), but also four lags.

$$(3) Y_t^* = \alpha^* + \beta^* D_t^* + \gamma_0^* X_t^* + \gamma_1^* X_{t-1}^* + \gamma_2^* X_{t-2}^* + \gamma_3^* X_{t-3}^* + \gamma_4^* X_{t-4}^* + \varepsilon_t^*$$

If the lagged variable is highly correlated with itself over time (autocorrelated), then including multiple lags of the same variable in the regression may cause imprecise estimates of the parameters.¹²

A solution to multicollinearity due to autocorrelation is to specify the form or structure of lags. There are two common forms of lag structure: geometric lag and polynomial distributed lag. Once the lag structure has been specified, OLS can be used to estimate the model.

3.2.C.i Geometric Lags

The first approach, the geometric lag, assigns positive weights to each of the lagged explanatory variables. The values of these weights (w) decline geometrically with time.

$$Y_t^* = \alpha^* + \gamma^* w^0 X_t^* + \gamma^* w^1 X_{t-1}^* + \gamma^* w^2 X_{t-2}^* + \gamma^* w^3 X_{t-3}^* + \gamma^* w^4 X_{t-4}^* + \beta^* D_t^* + \varepsilon_t^*$$

Since each the weights are positive and less than one, $0 < w < 1$, the weight given to earlier lags diminishes over time such that $w^0 < w^1 < w^2 < w^3 < w^4$. Note that the weight given to the current value of X^* is one because any value of w raised to the zero power is simply 1. So the current value is fully incorporated. For a given weight, $w = 1/2$, the current value has a weight of 1, the first lag has a weight of $1/2$, the second lag a weight of $1/4$, the third lag a weight of $1/8$, and the fourth lag a weight of $1/16$.

The above specification can also be rewritten in either of the following forms:

$$Y_t^* = \alpha^* + \gamma^* X_t^* + \gamma^* w^1 X_{t-1}^* + \gamma^* w^2 X_{t-2}^* + \gamma^* w^3 X_{t-3}^* + \gamma^* w^4 X_{t-4}^* + \beta^* D_t^* + \varepsilon_t^*$$

$$Y_t^* = \alpha^* + \gamma^* \sum_{i=0}^4 w^i X_{t-i}^* + \beta^* D_t^* + \varepsilon_t^*$$

¹² A second problem with including a substantial number of lags is the resulting loss in available degrees of freedom.

3.2.C.ii Polynomial Distributed Lags

Alternatively, the structure of the lags could be specified using a polynomial distributed lag. This method is more flexible than the geometric lag structure because it does not require that weights decline over time. Instead, the polynomial distributed lag model allows the structure to be specified as a continuous function, which in turn can be approximated by a polynomial function.

Assume a model with five terms of the explanatory variable X^* .¹³

$$Y_t^* = \alpha^* + \gamma^* (w^0 X_t^* + w^1 X_{t-1}^* + w^2 X_{t-2}^* + w^3 X_{t-3}^* + w^4 X_{t-4}^*) + \beta^* D_t^* + \varepsilon_t^*$$

The superscripts on the weights here do not represent powers, as they did in the previous section for geometric lags. For this approach, it is necessary to specify a functional form for the weights, specifically a polynomial function.

$$w^i = c_0 + c_1 i + c_2 i^2 + c_3 i^3 \quad \text{for } i = 0, 1, 2, 3, 4$$

Substituting in the polynomial function for the weights into the model yields:

$$\begin{aligned} Y_t^* = & \alpha^* + \gamma^* c_0 X_t^* + \\ & \gamma^* (c_0 + c_1 + c_2 + c_3) X_{t-1}^* + \\ & \gamma^* (c_0 + 2c_1 + 4c_2 + 8c_3) X_{t-2}^* + \\ & \gamma^* (c_0 + 3c_1 + 9c_2 + 27c_3) X_{t-3}^* + \\ & \gamma^* (c_0 + 4c_1 + 16c_2 + 64c_3) X_{t-4}^* + \beta^* D_t^* + \varepsilon_t^* \end{aligned}$$

By gathering terms, we can rewrite the model as:

$$\begin{aligned} Y_t^* = & \alpha^* + \gamma^* c_0 (X_t^* + X_{t-1}^* + X_{t-2}^* + X_{t-3}^* + X_{t-4}^*) + \\ & \gamma^* c_1 (X_{t-1}^* + 2X_{t-2}^* + 3X_{t-3}^* + 4X_{t-4}^*) + \\ & \gamma^* c_2 (X_{t-1}^* + 4X_{t-2}^* + 9X_{t-3}^* + 16X_{t-4}^*) + \\ & \gamma^* c_3 (X_{t-1}^* + 8X_{t-2}^* + 27X_{t-3}^* + 64X_{t-4}^*) + \beta^* D_t^* + \varepsilon_t^* \end{aligned}$$

It is important to note that our formula for the weights is a polynomial function of degree three, which we are using to specify a structure for a five-period lag. The degree of the polynomial function must be less than the number of lags in order to decrease the number of parameters. The original specification had five parameters related to X^* ; this number is reduced to four parameters when our third-degree polynomial function is used to weight the lags.

3.2.D Heteroskedasticity

Suppose again that we want to model the impact of job training, but that we now have data on caseloads for each of the 58 California counties for one period of time. Some counties have job training in place while others do not. Our variation is now across counties rather than across time. The subscript i now indexes each county.

¹³ The following example is taken from Pindyck & Rubinfeld.

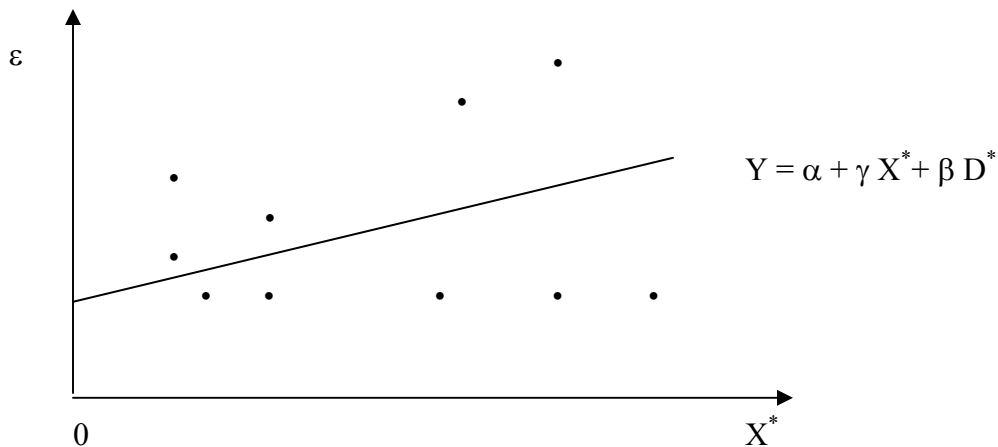
$$(4) Y_i^* = \alpha^* + \gamma^* X_i^* + \beta^* D_i^* + \varepsilon_i^*$$

The welfare caseload for each county (Y^*) can be modeled as a function of county population (X^*) and whether the county had a job training program (D^*).

In samples with a wide variance among observations, the variance of the errors may fluctuate in a well-defined manner among observations. This is a violation of the OLS assumption of constant variance.

$$\begin{aligned} \text{Var}(\varepsilon_i^*) &= \sigma_i^* \\ \sigma_i^* &\neq \sigma^* \text{ for all } i \end{aligned}$$

Recall that residuals from an OLS regression should have no discernible pattern if the OLS assumptions are not violated. But if the constant variance assumption is violated, there may be a pattern in the residuals. For example, in graphing the residuals from Model (4), we might find that the observations for counties with larger populations have larger variances in their error terms. The graph below shows how the distribution of the residuals (ε) relative to the county population (X^*). The residuals, the distance between the actual value of the dependent variable and the OLS predicted value, increase with the size of the independent variable.



Despite the violation of the constant variance assumption, the parameter estimates are still consistent because the explanatory variables and the error term remain uncorrelated. However, the OLS estimators are no longer efficient. The standard errors of the OLS parameters will be under-estimated because OLS fails to take into account the non-constant variance.

Recall that the t-tests used to test hypotheses are based on the standard errors as well as the parameter estimates. If the presence of heteroskedasticity causes the standard errors to be under-estimated, the calculated t-statistics are too large and the significance of the parameters will be over-estimated. Hypothesis tests based on these standard errors may consequently be incorrectly evaluated. As a

result, we may incorrectly infer a significant impact of job training on caseload where one does not exist.

To correct for this type of heteroskedasticity, the variables in the original multivariate model must be transformed by normalizing all the variables in the regression by county population (X^*).

$$Y^*/X^* = \alpha^* (1/X^*) + \gamma^* (D^*/X^*) + \beta^* (X^*/X^*) + \varepsilon^*/X^*$$

The error from the transformed model now has constant variance and the model can be estimated via OLS.

It is possible to test for the existence of heteroskedasticity using White's test.¹⁴ The null hypothesis for White's test is homoskedasticity or constant variance. Rejection of the null indicates the existence of heteroskedasticity. Should the test determine that heteroskedasticity exists, it is possible to correct the standard errors to conduct reliable hypothesis testing. However, the correction depends on the type of heteroskedasticity identified.

There are three types of heteroskedasticity: known variance, unknown variance, and variance that covaries with an independent variable. In the first case, there is sufficient prior knowledge regarding the variances such that weighted least squares can be used to resolve the estimation issues. With an unknown variance, a solution does not exist unless we have multiple observations for each set of observations of the independent variable. In the final case, there is a relationship between the variance and one of the independent variables. The estimation correction again involves weighted least squares. We will focus on this case in the example below.

3.2.E Error Structure – Serial Correlation

Recall that OLS requires that the error terms be uncorrelated across observations.¹⁵ If, for example, we have a time-series sample such that t indexes time rather than individuals, this condition requires that the error terms in each time period be uncorrelated with the error terms from previous time periods. In some cases, particularly with time series, the errors are likely to be serially correlated. For example, a positive error in the previous period may be related to a positive shock this period. In such cases, the disturbance term tends to have a lingering impact.

The simplest case of serial correlation is first-order autocorrelation, where the impact of the stochastic shock affects the current time period and the next time period only.

$$Y_t^* = \alpha^* + \gamma^* X_t^* + \beta^* D_t^* + \varepsilon_t^*$$

$$\varepsilon_t^* = \rho \varepsilon_{t-1}^* + v_t^* \quad \text{where } 0 < \rho < 1$$

Here the current error term, ε_t^* is a function of the current stochastic term v_t^* and last period's (i.e., the lagged) error term, ε_{t-1}^* . Substituting the error structure into the original model, it is easy to see that

¹⁴ Other tests can also be applied including Bartlett's Test, Goldfeld-Quandt, and the Breusch Pagan LM test.

¹⁵ In Model (4), this means: $E(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$; while in Model (2), this means: $E(\varepsilon_t, \varepsilon_s) = 0$ for all $t \neq s$.

the current value of the dependent variable is a function of the current stochastic error and the error lagged one period.

$$Y_t^* = \alpha^* + \gamma^* X_t^* + \beta^* D_t^* + \rho \varepsilon_{t-1}^* + v_t^*$$

Like heteroskedasticity, serial correlation does not affect the consistency or unbiasedness of the parameter estimates, but it does affect the efficiency with which they are estimated. Standard errors obtained from using OLS on serially correlated data will be underestimated and adversely affect hypothesis tests as was the case with heteroskedasticity.

A simple test for the presence of serial correlation is the Durbin-Watson.

$$DW = [\sum_{t=2}^T (\varepsilon_t - \varepsilon_{t-1})^2] / [\sum_{t=1}^T \varepsilon_t^2]$$

The null hypothesis for this test is no serial correlation. In general, a Durbin-Watson value near the value of 2 indicates that we fail to reject the null. If, however, the Durbin-Watson statistic indicates the presence of serial correlation, there are a number of corrections available for dealing with serially correlated error structures.

3.3 Pooling/Panel Data: Correcting for Specification Error or Lack of Data

3.3.A Why OLS Fails to be Efficient and/or Consistent in Panel Data

Often sufficient data are not available to estimate the model either because the time period is not long enough or there are not enough sampled individuals (e.g. counties). One way to circumvent this issue is to add more data - often this takes the form of pooling the data. In our job training example, pooling the data involves creating a panel where we follow multiple counties across multiple time periods. In this way, we create two dimensions over which the sample may vary: time and county. In addition to solving for lack of data, pooling data also allows for the use of more sophisticated analysis techniques.

It is important to note that the additional variation across counties introduces the need for more sophisticated techniques. When constructing a counterfactual for time-series analysis, we used comparisons of the period after the change to the period before the change. With panel data, the comparisons are both across time and counties. This situation is similar to the difference-in-difference technique introduced in Chapter 2. While the additional variation is useful, it may introduce new confounding factors. We will discuss these issues and potential solutions later in this section.

While we can estimate models using OLS on panel data, the assumptions necessary to insure that OLS is consistent and efficient are much stronger in a panel setting because of the two dimensions along which the data now vary. The conditions on uncorrelated error terms now apply across individuals and time periods. The errors must be uncorrelated across counties as discussed in Section 3.1.B:

$$E(\varepsilon_{it}^*, \varepsilon_{jt}^*) = 0 \text{ for } i \neq j$$

and the error term must now also be uncorrelated for each county across all time periods as discussed in Section 3.2.E on serial correlation:

$$E(\varepsilon_{it}^*, \varepsilon_{it+1}^*) = 0 \text{ for all } t$$

Failure of OLS assumptions requires that we search for alternative estimation methods. However, it is important to note that even when OLS assumptions are not violated, OLS estimation may still no longer be efficient because it fails to take advantage of the fact that we now have multiple observations on the same individual.

3.3.B Alternatives to OLS: Random Effects and Fixed Effects

For both efficiency and consistency concerns, alternative estimation techniques must be used with panel data. Alternatives to OLS include random effects and fixed effects estimation.

In the pooled data, the regression now takes the following form:

$$(5) Y_{it}^* = \alpha^* + \gamma^* X_{it}^* + \beta^* D_{it}^* + \varepsilon_{it}^*$$

Here we follow the counties (i) over several time periods (t). We observe changes in caseloads, job training treatment, and economic conditions over time for each county.

However, now the error term is no longer simply a stochastic term. Instead, the error term has up to three components: $\varepsilon_{it}^* = \alpha_i^* + \omega_t^* + \upsilon_{it}^*$. The components allow the error to have a component specific to individual counties (α_i^*), a component specific to time periods (ω_t^*), and a stochastic error term (υ_{it}^*).

$$Y_{it}^* = \alpha^* + \gamma^* X_{it}^* + \beta^* D_{it}^* + \alpha_i^* + \omega_t^* + \upsilon_{it}^*$$

Like OLS, random effects assumes that the correlation between the explanatory variables (D^*, X^*) and the stochastic part of the error term (ε_{it}^*) is zero. Because of the structure of the error term, this assumption now requires three elements for each explanatory variable:

$$\begin{aligned} E(\alpha_i, D^*) &= 0, E(\alpha_i, X^*) = 0 \\ E(\upsilon_{it}, D^*) &= 0, E(\upsilon_{it}, X^*) = 0 \\ E(\omega_t, D^*) &= 0, E(\omega_t, X^*) = 0 \end{aligned}$$

If these assumptions hold, then random effects is both consistent and efficient. It is important to note, however, that both OLS and random effects will estimate similar values for the parameters if the assumptions hold.¹⁶

Likewise, fixed effects will also consistently estimate the parameters of the model. However, unlike OLS and random effects, fixed effects does not assume zero correlation between the explanatory variables and the error term (ε_{it}). Instead, under fixed effects, we assume that there is some part (α_i) of the error term which covaries with the explanatory variables, but that this part is non-varying within the individual county (i.e., not varying with time in our example) and hence can be differenced out of

¹⁶ In this case, OLS is consistent though not efficient.

the regression. In our specification, it means that some counties may inherently have higher caseloads for reasons that do not vary over time.

The specification for this model is as follows:

$$\begin{aligned} Y_{it}^* &= \alpha^* + \gamma^* X_{it}^* + \beta^* D_{it}^* + \varepsilon_{it}^* \\ \varepsilon_{it}^* &= \alpha_i^* + v_{it}^* \end{aligned}$$

or to combine the two equations:

$$Y_{it}^* = \alpha^* + \gamma^* X_{it}^* + \beta^* D_{it}^* + \alpha_i^* + v_{it}^*$$

For example, assume that there is some unobservable characteristic which varies across counties, but remains constant over time for each individual county. Because the characteristic is unobservable, we do not have an explanatory variable with which to condition out the factor. For example, it may be the case that certain counties always have higher caseloads despite similar populations and job training programs. This difference may be based on some characteristic which we cannot observe or quantify and which does not vary over time. If that is the case, then a fixed effects model will allow us to differentiate between this unobservable element and the effects of the included explanatory variables.

Using OLS, this issue would lead to an omitted variable bias. However, using fixed effects, we can remove the time-invariant unobservable characteristic. This can be accomplished by demeaning the variables in the equation (e.g. $Y_{it}^* = Y_{it}^* - \underline{Y}_i$) or by including indicator variables for each county so that, in effect, the intercept is allowed to vary at the county level. The I_i variables are indicator variables isolating the observations for each county, and the α_i terms are the county-specific intercepts.¹⁷

$$Y_{it}^* = \alpha^* + \gamma^* X_{it}^* + \beta^* D_{it}^* + \alpha_i^* I_i + v_{it}^*$$

In this case, fixed effects assumes that the explanatory variables (X^* , D^*) are uncorrelated with only one component of the error (v_{it}^*), rather than the entire error (ε_{it}^*). If this error specification is true, then neither OLS nor random effects will yield consistent estimators because of the implicit omitted variable (i.e., the unobservable characteristic). Of the three, only fixed effects yields consistent parameter estimates.

While panel data have the benefit of increasing the number of observations and consequently the precision of the estimates of the impacts of common variables, they may also complicate the estimation. To see how this is so, we must review what we did in the county-level multivariate regressions where we used variation across time within a single county to estimate the impact of job training on caseloads. In effect, these regressions allowed us to compare the observed caseload in one set of circumstances where there is job training with a predicted caseload that would have occurred without job training. That is, the regressions make it possible to construct a counterfactual. One way to think about these regressions is that they take the observed caseload

¹⁷ When a constant is included in the regression, one of the individual dummies must be dropped to avoid collinearity. In this case, the α_i term must be added to the constant α to calculate the individual-specific intercept.

when there is no job training and adjust it for factors affecting caseload, other than job training itself, that change between the period when there is no job training program and the period when job training is implemented. This adjusted value is what the caseload would have been without job training. If it differs significantly from the observed caseload when there is job training, then we can say that job training has an impact.

Panel models have more statistical power than multivariate models for single counties because there is variation across both time and counties and because it is assumed that many of the control variables operate similarly from one county to another. Hence, the impact of these control variables can be more precisely estimated by using information from all the counties. As a result, more accurate counterfactuals can be produced which make it possible to get better estimates of the impact of the job training program. This argument works perfectly if the control variables act the same across counties. But, counties may differ in ways that can undermine the results of the analysis. Perhaps, for example, economic variables have more impact on welfare caseloads in some counties than in others, but the panel model falsely assumes that the impact is the same. As a result, the counterfactuals estimated from the panel model for a specific county, say one in which the economic variables have no impact, will incorrectly assume that economic variables do have some impact. Thus, the comparison of the observed caseload with a counterfactual prediction will involve an erroneous counterfactual -- one in which economic variables are assumed to have an impact. There is no simple solution to this problem, although in some cases such differences between counties may be overcome by using fixed effects models if the characteristics are fixed over time. For example, if social norms such as the stigmatization of being on welfare are constant over time within a county but vary across counties, then a fixed effects model will control for these differences.

In addition to adding a county-specific indicator variable (or demeaning the variables) in a fixed effects model, a time-invariant characteristic can also be removed by first-differencing the data.¹⁸ It is worthwhile to check the results of the first-differencing versus the fixed effects specification. Both specifications should yield similar results. If the results differ, there is evidence of some specification error such as the possibility that the omitted characteristic may not be time-invariant. If the unobservable characteristics are time-varying, fixed effects may not be sufficient to yield consistent estimates.

3.3.C Choosing between Random Effects and Fixed Effects

While both random effects and fixed effects may be preferable to OLS in estimating models with panel data, choosing between the two estimation techniques requires information on the specification of the error term.

The Hausman test is a specification test for choosing between fixed effects and random effects. The null hypothesis is that the random effects specification is the correct specification. Recall that random effects requires that the random effects (α_i) and the explanatory variables (D^* , X^*) be uncorrelated.¹⁹

¹⁸ First differencing the data involves subtracting a county's data at time t from the same county's data at time $t-1$ (e.g. $Y_{it}^* - Y_{it-1}^*$).

¹⁹ This follows from the assumption that X^* and ε^* are uncorrelated since $\varepsilon_{it}^* = \alpha_i^* + \nu_{it}^*$

By comparing estimates and efficiency measures (variance measures), this assumption can be tested. Under the null hypothesis of no correlation, the two estimates should be similar. In this case, random effects and fixed effects models are both consistent, but the random effects model is more efficient. However, if the null is rejected, the fixed effects model is consistent, but the random effects model is not.

3.3.D Individual-Specific and Time-Specific Intercepts

As shown above, the error specification in panel data can include individual (e.g. county) and time components in addition to a stochastic error term. These components allow the model to incorporate county-specific and time-specific effects. These components shift the mean response (intercept), but do not interact with the slope parameters.

Individual-specific effects vary across individuals, but are time-invariant within each individual. For example, year of birth varies across individuals, but does not vary for a single individual over time. A more interesting example is that used in Section 3.3.B where each county has a time-invariant omitted unobserved characteristic. Both of these factors can be eliminated by including an indicator variable for each individual or county. The second example is more relevant to fixed effects because there is no other means to correct for the omitted characteristic because it is unobserved.²⁰

Time-specific effects vary by the observation period (e.g. year or month), but are constant across individuals. An example of a time-specific effect is an occurrence that affects everyone in the sample equivalently and is particular to a specific time-period such as a change in economic opportunities due to a recession in a particular year that affects the mean accession level, but does not interact with the explanatory variables. Including an indicator variable for each time period (T) allows the researcher to control for such factors easily.

$$Y_{it}^* = \alpha^* + \gamma^* X_{it}^* + \beta^* D_{it}^* + \alpha_i^* I_i + \tau_t^* T_t + v_{it}^*$$

Including indicator variables for individual and time-specific omitted variables in a panel dataset, is a simple way to control for unobserved characteristics or events. To avoid multicollinearity, the indicator variable for one county must be excluded from the regression if a constant is included in the regression. The results for the indicator variables are then relative to the excluded county whose county-specific constant is the regression constant (α^*). Other counties have county-specific constants equal to the regression constant plus their county-specific intercept: $\alpha^* + \alpha_i^*$.

3.3.E Individual- and Time-Specific Slope Coefficients

While this section thus far has focused on differences in means across counties and time-periods, it is also possible that individuals or groups of individuals vary in their response to explanatory variables. That is, the slope coefficients may also vary across counties.

²⁰ By contrast, we can control for the observable differences in year of birth by including that variable in the regression. This is not possible for unobservable characteristics because we lack a variable which quantifies the characteristic.

For example, we may have two types of counties: agricultural and urban. Due to different opportunities available in each county, the counties may vary not only in the mean or intercept of caseloads, but also in the response to changes in program availability. In order to estimate the responses separately, we can interact the job training treatment with an indicator variable identifying the county type.

$$Y_{it}^* = \alpha^* + \gamma^* X_{it}^* + \beta^* D_{it}^* + \beta_i^* I_i D_{it}^* + \varepsilon_{it}^*$$

The model now includes the new term, $\beta_i^* I_i D_{it}^*$, where I_i identifies the status of each individual and is equal to one for agricultural counties and zero for urban counties. This model now estimates an overall response to changes in job training availability (β). This is the total impact of job training on welfare participation for the excluded group of urban counties ($I_i=0$). In addition to the overall response (β^*), agricultural counties also experience an additional response (β_i), which is added to the overall response, so that the impact of job training on welfare participation for agricultural counties is ($\beta^* + \beta_i^*$). Note that β_i^* may be positive or negative. If the coefficient β^* is negative, then overall job training reduces welfare caseloads. However, if the county-specific slope coefficient (β_i^*) is positive though smaller in absolute value than β^* , it indicates that welfare participation in agricultural counties is less impacted by job training than in urban counties. The reverse is true if the county-specific slope coefficient is positive.

In this example, we have divided our counties into two types and estimated slope coefficients separately for each group. In the extreme, a panel allows us to estimate a separate slope coefficient for each individual county provided that we have enough observations over time for each county.

Although we focus on differences across counties or groups of counties in this example, similar analysis can be done across specific time-periods or groups of time-periods. If we suspect that the response to changes in job training treatment is changing over time due to related legislative changes, we may want to differentiate responses before and after the legislation.

3.4 Forecasting and Model Checking

Using estimated parameters from the techniques described above, it is possible to predict values of the dependent variable outside the estimation sample. For a given sample, researchers may estimate the model on all the available observations, and then forecast periods outside of the available sample. However, it can be useful to estimate the model on a subsample of the data and then use the remaining observations to check the forecasting accuracy of the model.

Both of these types of forecasting are called out-of-sample forecasting because the forecasts are outside the estimation period. The distinction between out-of-sample and in-sample forecasting is critical. Recall that OLS chooses the parameter estimates to minimize errors over the estimation period so that within sample forecasting will be quite accurate. However, the out-of-sample errors may be quite large because observations during those time periods (if available) were not included in the minimization. We will discuss the potential problems in Section 3.4.B after our introduction to forecasting.

3.4.A Constructing Forecasts and Forecast Types

Forecasts are simply values of the dependent variable calculated using the parameters estimated using the estimation sample (α, β) and current values of the independent variable. Note that both the forecasted dependent variable and the independent variable are out-of-sample.

For example, assume that we estimated a simple model of welfare caseloads over the time period from 1969 to 1985; assume that caseloads are simply a function of one independent variable, job training. We would now like to forecast accessions for 1986. Using the parameter values estimated from 1969 to 1985 and the value for benefit levels in 1986, we can calculate accessions for 1986:

$$Y_{1986} = \alpha + \beta D_{1986}$$

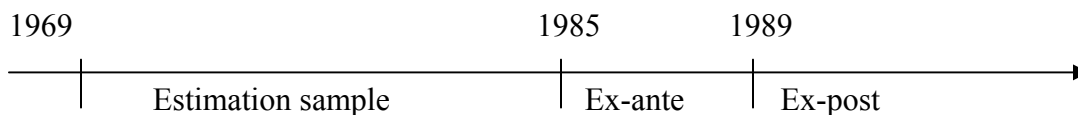
It should be clear that forecasts are constructed in a manner very similar to the predicted values for the estimation sample. An important distinction is that the parameters used in the calculation are estimated over a time period which does *not* include the forecast period(s).

While all forecasts share this characteristic, there are two dimensions along which forecasts may vary: availability of data for the dependent variable and availability of data for the independent variable(s). The resulting four forecasting types are: ex-ante forecasts, ex-post forecasts, unconditional forecasts, and conditional forecasts.

The first dimension is the time-period over which the forecasts are made. *Ex-ante forecasting* refers to forecasting of the dependent variable for periods where its values are already known, but which are outside the estimation sample. *Ex-post forecasting* predicts values for periods which have not yet occurred.

For example, assume that we have annual data for our independent and dependent variables from 1969 to 1989. We estimate a model over the time period from 1969 to 1985. Using the estimated coefficients from this regression, we can then forecast the dependent variable beyond the estimation period.

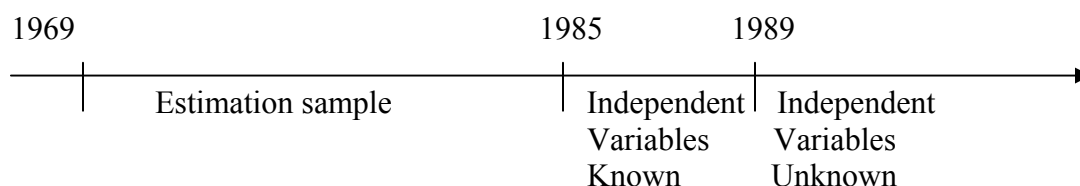
Ex-ante forecasting can predict the dependent variable from 1986 to 1989 – the final year for which values of the dependent are available. *Ex-post forecasting* refers to prediction of values after 1990.



The major distinction between ex-ante and ex-post forecasting is that we can immediately check the precision of ex-ante forecasts because we know and can compare it to the actual outcome. For ex-post forecasts, we would have to wait for the actual values to occur before a comparison could be made.

The second characteristic along which forecasts vary is the availability of data for the independent variables.

Unconditional forecasting indicates that values for the forecasted time period(s) are known for all of the explanatory variables. *Conditional forecasting* indicates that values for some or all of the explanatory variables are unknown. In this case, values for the unknown explanatory variables must be estimated and the forecasts for the dependent variable are conditional on forecasts for the independent variable.²¹



In the above diagram, the estimation period is again 1969 to 1985. As before, we have data through 1989; in particular, the values of the independent variables are known (X^*) through 1989. As a result, forecasts through 1989 are unconditional because values for the explanatory variables are known. However, forecasts after 1989 are conditional because they require estimation (or predictions) of values for the explanatory variables (X) before forecasts can be made for the dependent variable.

$$\begin{array}{ll} \text{Unconditional Forecasting (X known):} & Y_{1986} = \alpha + \beta X^*_{1986} \\ \text{Conditional Forecasting (X predicted):} & Y_{1986} = \alpha + \beta X_{1986} \end{array}$$

3.4.B Model Checking: Forecast Error

Forecasts, like predicted values, offer the researcher the opportunity to evaluate the model. Comparison of forecasted values with actual values, already available or soon to be available, allows the researcher to check the accuracy of predictions by calculating measures of forecast error. One possible measure, which parallels OLS' minimization of the sum of the square residuals, minimizes the sum of the squared errors over the forecasted time periods (T) only.

$$\text{Minimize Forecast Error} = \sum_{t=1}^T \epsilon_t^2 = \sum_{t=1}^T (Y_t^* - Y_t)^2$$

A related summary measure is the average mean squared forecast error (AMSFE).

$$\text{AMSFE} = \frac{\sum_{t=1}^T (Y_t^* - Y_t)^2}{T} = \frac{\sum_{t=1}^T (\epsilon_t)^2}{T}$$

Such calculations are possible with *ex-ante* forecasts since the actual values for the dependent variable are known when the forecast is made. With *ex-post* forecasts, calculation of the forecast error cannot be done until the actual values become known.

²¹ The forecasts are now dependent on an earlier forecast which is also subject to estimation error. As a result, the standard errors of the forecast for the dependent variable are generally larger for conditional forecasts than for unconditional forecasts.

3.4.C Sources of Forecast Error

There are three main sources of forecasting error: insufficient sample size, differences between the estimation and forecast sample, and structural changes in the model.

3.4.C.i *Insufficient Sample Size*

Often model estimation is imprecise because the sample size, the number of observations, is too small. The imprecise estimation of model parameters will then lead to imprecise forecasts. This source of error is straightforward to correct – simply find more observations.²² As the number of observations increases, the estimates will become more accurate.

3.4.C.ii *Estimation vs. Forecast Sample*

The second source of forecast error was mentioned briefly in Section 3.4.A. In discussing the calculation of forecasts, we noted that the parameters used to calculate the forecast are estimated over a different sample than that which is forecast. While this is standard practice, there are occasions when problems with the forecasts can result.

In particular, it is important to note that while OLS selects parameters in order to minimize the sum of squared residuals, it cannot and does not minimize the out-of-sample forecast error. Within the estimation sample, OLS will ensure²³ small errors between predicted and actual values. However, because the out-of-sample observations were not included in the estimation, the difference between the forecasts and the actual values can be quite high.

3.4.C.iii *Structural Changes*

In our ex-ante forecasting example, we estimated a model of welfare caseloads as a function of the job training programs from 1969 to 1985. We then used the resulting parameters to forecast caseloads for 1986 to 1989. If the regression had a high R^2 (or goodness of fit), we would expect that our ex-ante forecasts would be very similar to the actual values for caseloads.

However, it is important to note that our parameters measure the *relationship* between accessions and benefits over a different time period than our forecasts. If the relationship between the dependent and independent variables changes after 1985, we may find that our forecast error is quite large despite the fact the sum of squared residuals (in-sample) was small.

For example, assume that legislation in 1986 increases the income levels at which people are eligible for benefits, but leaves job training unchanged. We may see an increase in actual caseloads in 1986 due to increased eligibility, but our forecasted caseloads will not reflect this change because our model, estimated only until 1985, does not account for this change in policy.

²² Clearly, finding additional observations may not always be possible. However, theoretically, this source of error is easily remedied.

²³ This hinges on the requirement that the assumptions of OLS estimation hold.

3.5 Model Uncertainty

All of the above methodologies rely on the fundamental belief that we know the true underlying model or equivalently, the process that generated the data. Ideally, the researcher should postulate a model based on either theory or some previous beliefs about the process that generated the dependent variable. Of course, given that we often choose the best model among a significant number of alternative models, this ideal generally does not hold.

In fact, rather than starting with knowledge of the underlying process and then estimating the resulting theoretical model on the data, we often run several models and use the results to choose a final theoretical model based on what our findings indicate about the significance of our parameters and goodness-of-fit of the regression. Although a common practice, this is technically an incorrect approach to modeling and violates the assumptions behind our estimation and inference methodologies.

3.5.A Pre-Testing and Selecting Models

It should not be surprising that our models find highly significant results or have high R^2 statistics when we make an effort to select the model which performs best specifically on these criteria. Likewise, the practice of selecting models based on pre-testing models on the final dataset also creates problems for inference.

3.5.A.i Pre-test

In practice, researchers tend to test various specifications on the same dataset that is used for the final estimation. Theoretically, a pre-test should be estimated on a separate dataset or on a subsample of the final dataset. The pre-test data should not be used again in the final estimation of the model.

A simple example of this potential bias involves the estimation of a slope parameter β_2 . We have a model where Y^* is dependent on a variable, X^* , and potentially on a second variable, Z^* . We have two postulated models since we are debating whether to include the second explanatory variable.

$$\begin{aligned} Y^* &= \alpha^* + \beta^* D^* + \gamma^* X^* + \mu^* Z^* + \varepsilon_{it}^* \\ Y^* &= \alpha^* + \beta^* D^* + \gamma^* X^* + \varepsilon_{it}^* \end{aligned}$$

If upon estimating the multivariate model, we find that the estimated coefficient μ is not significant, we may drop the variable Z^* from the equation and re-estimate the model with a single explanatory variable. This will increase our *adjusted* R^2 because the irrelevant variable Z^* has been eliminated and the degrees of freedom has increased.

This process introduces two sources of potential bias. First, it is important to recall that proper use²⁴ of OLS gives us an unconditional estimate of μ^* . However, the practice of first testing whether μ is non-zero and significant before deciding whether to include it in the final specification gives us an estimate which is *conditional* on the parameter being significantly different from zero.

²⁴ The term proper use indicates that the assumptions of OLS are not violated.

$$\mu_{pt}^* = \begin{cases} \mu & \text{if } \mu \text{ is significant} \\ 0 & \text{otherwise} \end{cases}$$

This *pre-test* estimator, μ_{pt}^* , is now conditioned on finding a significant response and consequently is no longer equal to the unconditional OLS estimator μ .

A second issue arises in the selection of a model from a series of alternatives. By comparing alternative models and selecting the group of explanatory variables with the highest R^2 or which maximizes some other criteria, we may in fact bias the parameter estimates by eliminating or including explanatory variables based on the estimation results rather than prior beliefs and theory.

Related selection issues which may lead to inappropriate inferences are sample selection and data transformation. Sample selection involves the down-weighting or dropping of outliers without proper theoretical foundations. Data transformation includes transforming explanatory variables in order to create normality or to decrease their variance.

The fundamental problem with both pre-tests and selection is that reporting only the results from our preferred OLS regression fails to incorporate the model selection process into the inferences drawn from the results. The fundamentals of OLS theory are thus violated and the resulting inferences drawn from OLS estimation may be inappropriate when pre-test and selection are conducted on the same dataset as the final analysis.

Particularly with smaller samples, the variance of the parameter will increase due to additional uncertainty resulting from model selection. This bias, however, generally becomes less significant as the sample size increases. The reduction in bias as the sample size increases follows from the asymptotic distribution of the parameter which is not affected by *consistent* procedures for model selection.²⁵

3.5.B Reporting Alternative Specifications and Model Averaging

To account for the fact that models are being selected after the results are known, some authors report a series of specifications and their results. The benefit of this practice is that it gives some indication of the robustness of the results – specifically, their robustness to changes in the specification.

While reporting results from multiple specifications is an improvement, Chatfield (JASSA 1995) argues for a more Bayesian approach. He argues that estimation results from multiple specifications should also be averaged into a single result.

The specifications are not weighted equally in the averaging of the results. Rather, the weight attributed to each result should be specified *before* the estimation results are known in order to identify the researchers prior beliefs about the plausibility of the specification.

²⁵ There are some selection procedures that will affect the distribution, but often the resulting distribution can be calculated.

Chapter 4 Data Description

In the previous chapter we discussed various analytical techniques available to welfare researchers. In this chapter, we describe data that can be used with the analytical techniques to examine causal relationships in future chapters. While the data we use do not pertain to a particular experiment, with clearly defined control and experimental groups, we still can examine the effects of changes in AFDC generosity during the time period prior to the enactment of welfare reform legislation. Chapter 5 uses these data to estimate models of AFDC-FG and AFDC-U accession and termination rates. Chapter 6 considers forecasting models.

The analysis presented in the next two chapters uses many types of data available from state agencies covering a wide variety of areas. State agencies collect data to administer various programs. Rarely do these agencies explicitly design data systems to collect data merely for research purposes. These administrative data systems, however, can collect valuable information for the examination of causal relationships in welfare.

4.1 Welfare participation variables

In California, the agency directly in charge of administering welfare, specifically AFDC¹, is the Department of Social Services. California's AFDC program is administered at the county level using statewide policies. This differentiation means that there is an opportunity to collect welfare information both from the county and from the state. Using state data sources ensures some consistency among counties and gives us the opportunity to examine panels of counties.

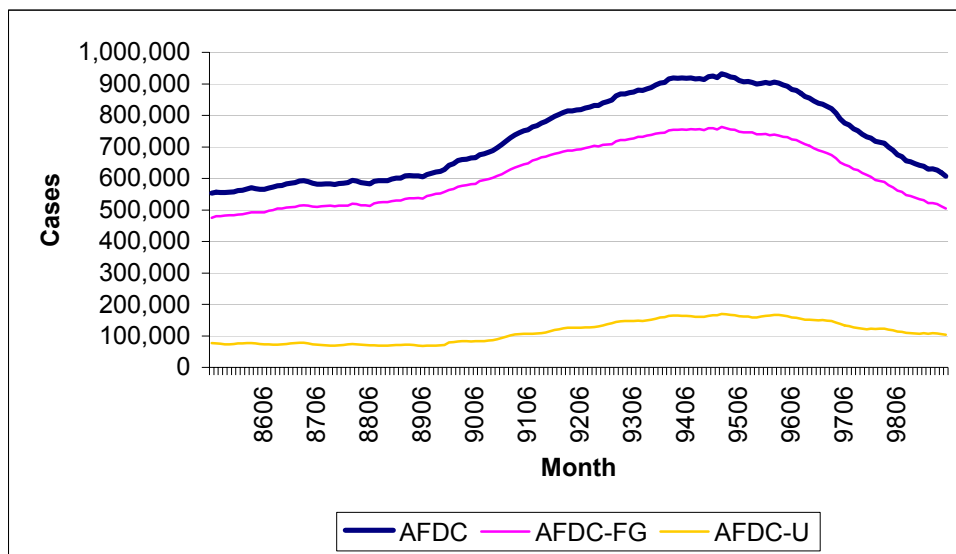
4.1.A. Monthly Caseload Reports - CA-237 Report

Each county reports county level caseload numbers each month to the California Department of Social Services on form CA-237. While we have data from these CA-237 forms up to the present, we restrict ourselves to county level welfare caseload figures from 1985 to 1999 to ensure consistency². We see in Figure 4.1 that from 1985 through 1996 (the pre-TANF period) California has experienced remarkable growth in AFDC caseloads. While both the one-parent (AFDC-FG) and two-parent (AFDC-U) AFDC programs have experienced large increases in caseloads during the early 1990's, the growth in the two-parent AFDC-U program has been the most noteworthy. AFDC-U caseloads more than doubled between 1989 and 1995 to over 160 thousand cases. Between 1996 and 1999, AFDC caseloads in both the FG and U programs have been declining. By 1999 the AFDC-FG caseloads had returned to 1988 levels and AFDC-U caseloads returned to 1991 levels. In general, however, AFDC-U cases are a smaller proportion of the total AFDC caseload, fluctuating between 11 percent and 18 percent, with a mean of 15%.

¹ In January 1998, California converted all AFDC cases to CalWORKs, California's TANF program. For simplicity, throughout this report we refer to AFDC/CalWORKs solely as AFDC.

² As California transitioned between AFDC and CalWORKs the CA-237 form underwent several changes. The first of these changes happened in July 1999. For consistency of data, we use reports from a single version of the form.

Figure 4.1 California AFDC Caseloads



4.1.B. Monthly Accessions and Terminations Reports - CA-237 Report

Not only do counties report both AFDC-FG and AFDC-U caseload numbers to the state on the monthly CA-237 reports, they also differentiate between continuing cases, accessions and terminations. This allows researchers to determine which component of the caseload is affected by the explanatory variables under consideration. We find analysis that differentiates between accessions and terminations to examine effects on caseloads more illuminating.

Previous work on accessions and terminations (or openings and closings) has been hampered by data collection difficulties. In state reporting, openings are considered only to be the cases that enter the caseload through a formal application process. As CBO (1993) reports, this severely undercounts the number of openings since many cases do not join AFDC with a formal application. In California, this process manifests in a separation of openings into those who begin with a formal application process and those who rejoin the rolls after a short break (less than one year). Those who return to AFDC are labeled as “restorations”. Our definition of accessions is more encompassing than just new case openings. We consider accessions to be new openings plus restorations plus transfers (either from other programs or from other counties). This definition of accessions does not suffer from the same consistent undercounting found in previous studies.

A month by month comparison of constructed caseload³ and actual caseload for each county finds the difference to be mostly noise. The mean difference is extremely close to zero (-0.34 for AFDC-U and -0.64 for AFDC-FG) with median and mode both equal to zero. See Figures 4.2 and 4.3 which show differences over time for AFDC-U and AFDC-FG respectively for the entire state. While it does seem that differences seem to be larger in more recent time periods, we do not know why. These differences when expressed as a percentage of the caseload are small. The

³ ($CC_t = AC_{t-1} + A_t - T_t$, where CC is constructed caseload, AC is actual caseload, A is accessions as described above, T is terminations, and t is a time subscript)

median differences for the AFDC-U caseload is less than 2% of caseload and for AFDC-FG is less than 1% of the caseload.

Figure 4.2: Monthly Adjustments to Reconcile AFDC-U Caseload

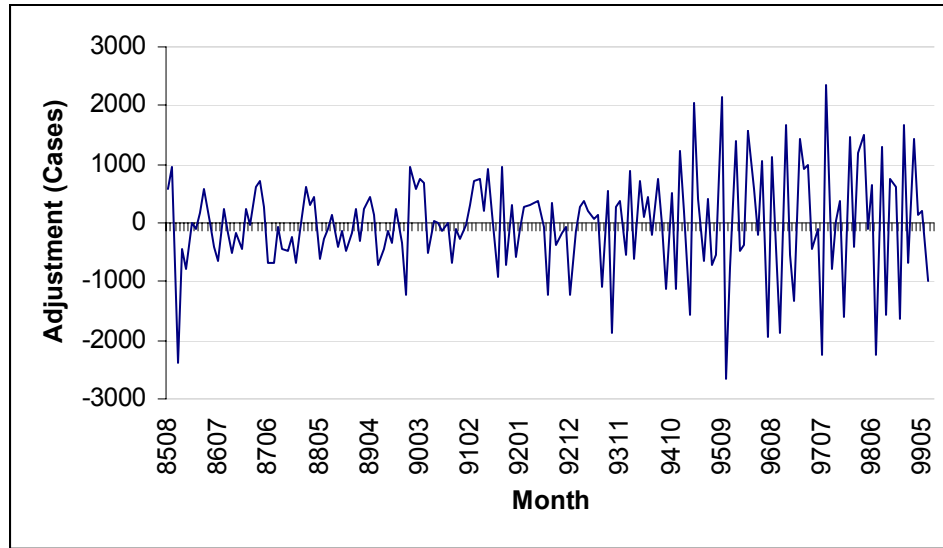
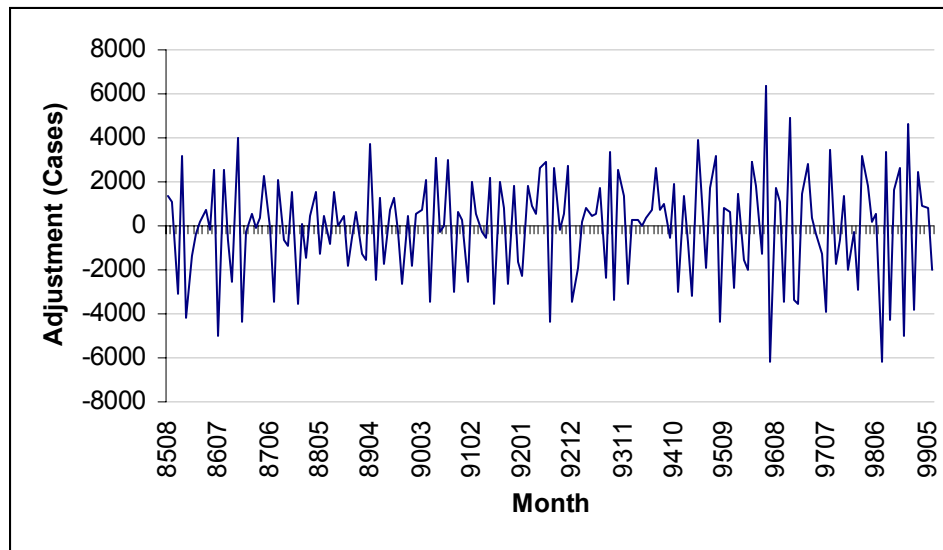


Figure 4.3: Monthly Adjustments to Reconcile AFDC-FG Caseload



Taking these month to month differences over the entire time period covered in this report (July 1985 though June 1999) we find that our constructed caseload⁴ tends to overstate the caseload by a small number of cases. The mean difference is 82 cases for AFDC-U and 95 cases for AFDC-

⁴ ($CCC_t = AC_{8507} + \sum_{k=8508}^t (A_k - T_k)$, where CCC is cumulated constructed caseload, AC_{8507} is actual caseload in July 1985, A is accessions summed over the period from August 1985 to time t, T is terminations summed over the same time period, and t is a time subscript)

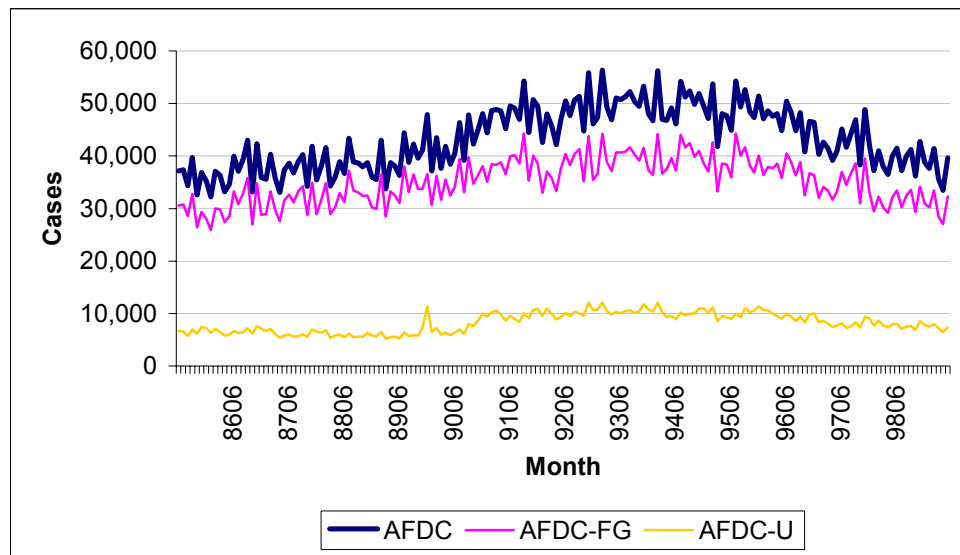
FG. Median differences are 11 and 2 for AFDC-U and AFD_FG respectively. We find that while over time our constructed caseload seems to be always higher than the actual caseload, this isn't true for all counties. We find that some counties overstate caseload, some counties understate caseload and some are mostly correct. As we would expect, we find that Los Angeles County (a county where our constructed measure is typically higher than actual caseload) drives any statewide caseload figures.

While we would prefer that this difference did not exist, we find that we are not able to completely recreate caseloads with accessions and terminations. Differences may be adjustments to caseload figures that are also reported on the CA-237 reports that cannot be separated between accessions and terminations. These adjustments allow the CA-237 reports for each county to reconcile caseload monthly. The CA-237 reports do not state any reason for the adjustment.

An analysis of these adjustments finds no real behavior of interest. The adjustments differ over time and by county. Running the final models we describe in Chapter 5, we find the only clear behavior is that counties differ in the amount of adjustment they report.

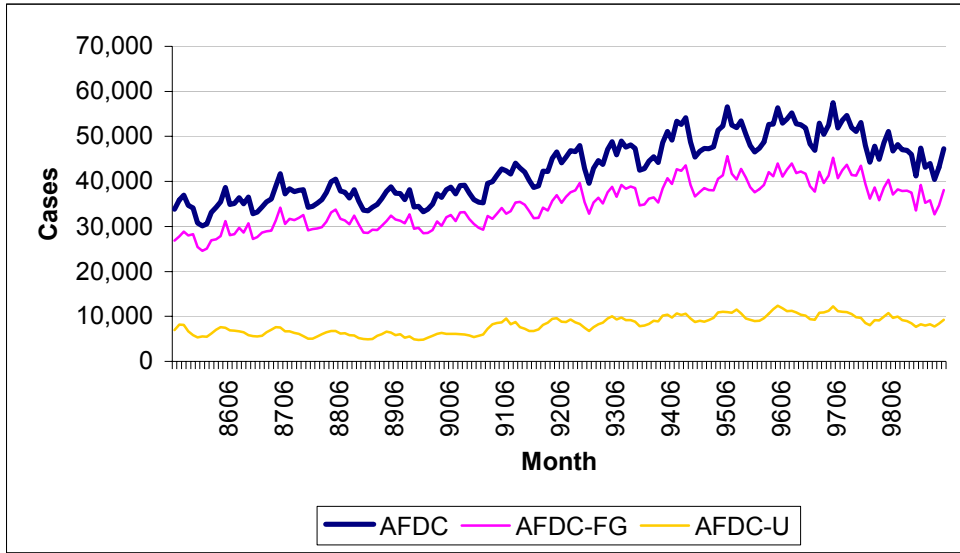
We believe that the components of interest in caseload are not the static part of the caseload that remains on but rather the flows into and out of the caseload. Our definitions of accessions and terminations account for these flows. Figures 4.4 and 4.5 show accession and termination levels for California.

Figure 4.4 Accessions in California⁵



⁵ There is a significant spike in AFDC-U accessions in January 1990. This sudden increase in accessions is due to the conversion of Refugee Demonstration Project AFDC cases to standard AFDC cases. The majority of these cases were AFDC-U cases rather than AFDC-FG cases. Numerous attempts to control for this spike, including dummy variables or smoothing over the transition of these cases did not have any effect on the regressions run in Chapter 5.

Figure 4.5 Terminations in California



In our analyses we use county level accession and termination rates. Accession rates are defined as the number of accessions divided by the population at risk of becoming on aid. The exact specification of population at risk is discussed below. Termination rates are defined as the number of terminations divided by caseload. Figure 4.6 and 4.7 show accession and termination rates for the state overall. We can see that rates show a different picture than levels. Figure 4.7 shows that the termination rate falls until 1992. This fall in termination rate is more a result of caseloads increasing at a faster rate than terminations rising. Figures 4.5 and 4.1 confirm that in fact during this time both terminations and caseload were increasing.

Figure 4.6 California Accession Rate

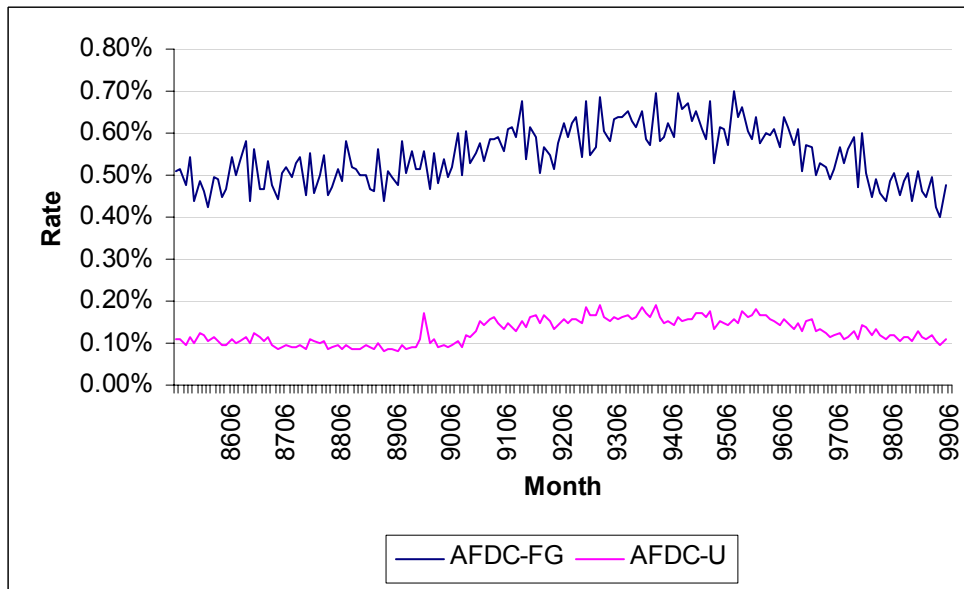
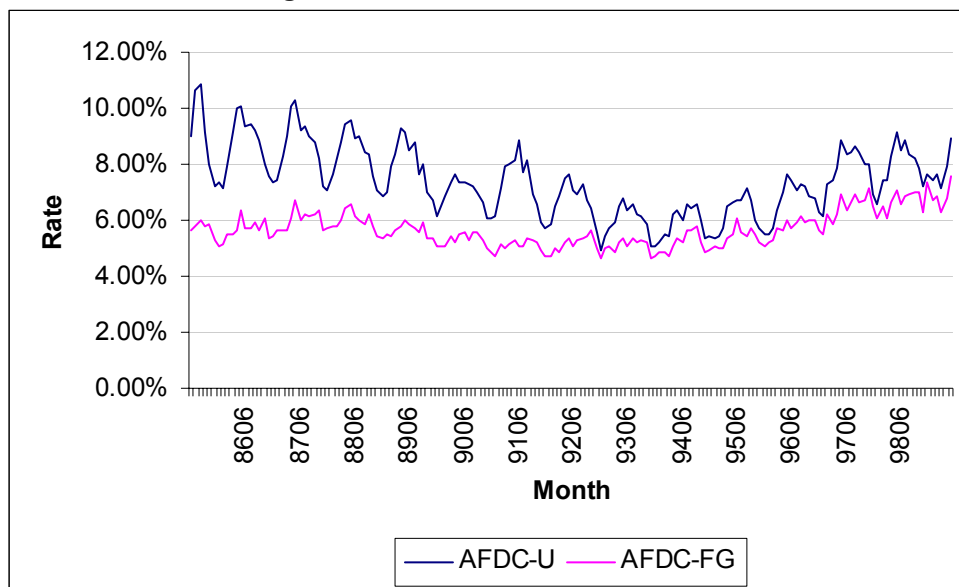


Figure 4.7: California Termination Rate



Both the accession rate and the termination rates for AFDC-FG and AFDC-U show some seasonality. Most obvious is the pattern in termination rates: terminations are high around June of each year and low in December of each year. Accessions have a similar but opposite pattern.

4.1.C. Maximum Benefit Amounts

Also of interest is the amount of benefits available from the AFDC program. Although benefits for families depend on specifics about the household receiving aid, we use AFDC maximum benefit levels as a proxy for the benefit amount available. Benefit levels depend on family size as well as county.⁶ Before 1997 AFDC maximum benefit levels depended only on family size. Beginning in 1997, two maximum benefit schedules were implemented to reflect the State's desire to be more generous in urban counties relative to rural counties. Table 4.1 lists the 17 California counties that the state identifies are urban counties, Region 1. Since AFDC-FG cases represent single parent households on aid, we used the average AFDC-FG case size to determine the maximum AFDC benefit for an AFDC-FG case. Using the monthly caseload counts and person counts we found that the average AFDC-FG case size is three persons. Similarly for two parent households, AFDC-U, we used the average case size of four persons.

Table 4.1: Counties of California in Region 1: Urban Counties

Alameda	Napa	San Mateo	Sonoma
Contra Costa	Orange	Santa Barbara	Ventura
Los Angeles	San Diego	Santa Clara	
Marin	San Francisco	Santa Cruz	
Monterey	San Luis Obispo	Solano	

⁶ Prior to 1997, maximum AFDC Benefit Levels only depended on household sizes. Beginning in January 1997, the State decided to implement two different maximum benefit schedules based on which of two regions the recipient lived in. Counties were placed in two regions based on whether the state felt they were more urban or more rural.

The maximum AFDC benefit level is not adjusted regularly for inflation. Changes to the AFDC benefit level schedule must be legislated and therefore changes often occur at uneven time intervals. Figure 4.8 shows legislated nominal AFDC maximum benefit levels for our average sized AFDC-FG and AFDC-U households. In our analyses, we use deflated benefit amounts to account for inflation. Our deflation factor is based on the Bureau of Labor Statistics Consumer Price Index (CPI).⁷ Figure 4.9 shows the value of AFDC maximum benefit amounts in constant dollars for our two average sized households. Generally, the value of the maximum AFDC benefit in constant dollars for both AFDC-FG and AFDC-U cases has been falling over time.

Figure 4.8 Nominal Maximum AFDC Benefit Amounts

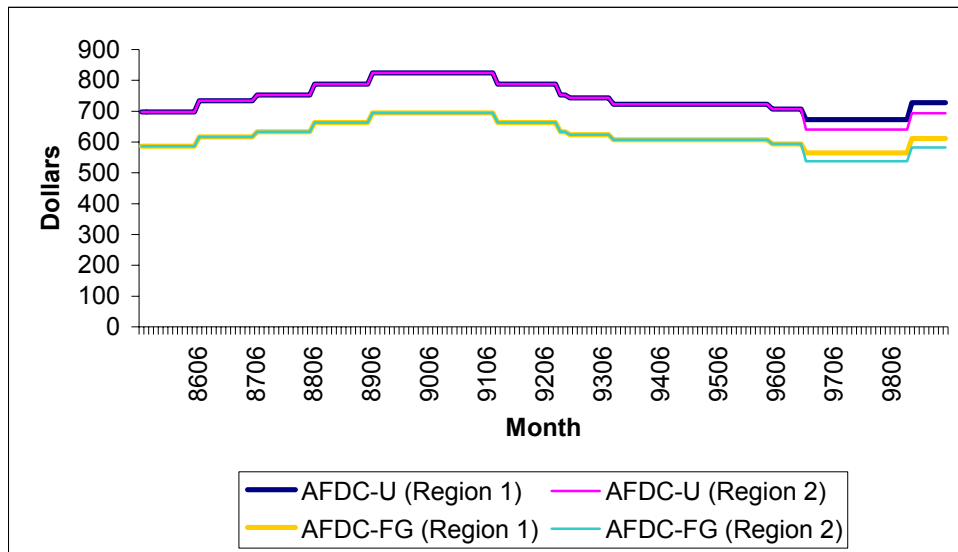
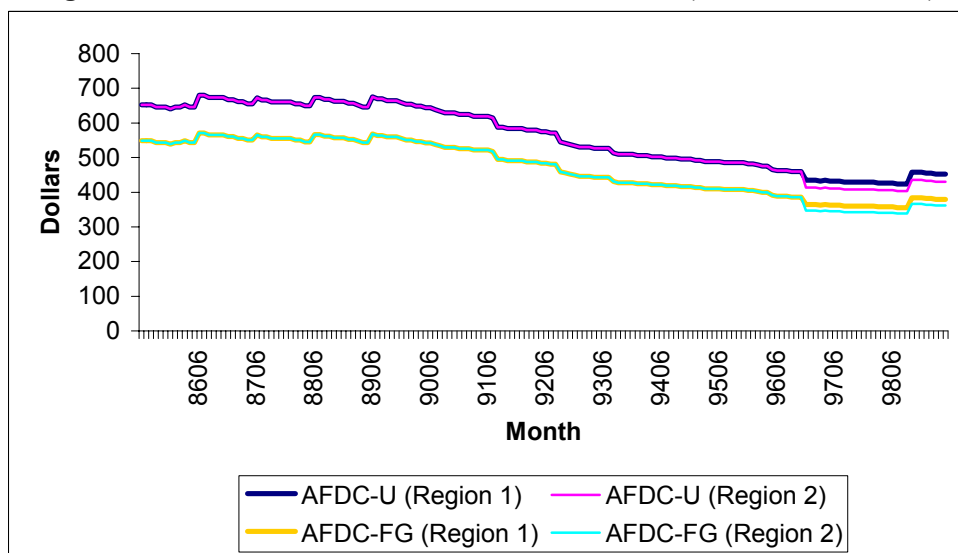


Figure 4.9 Maximum AFDC Benefit in Constant Dollars (Base 1982-84 dollars)

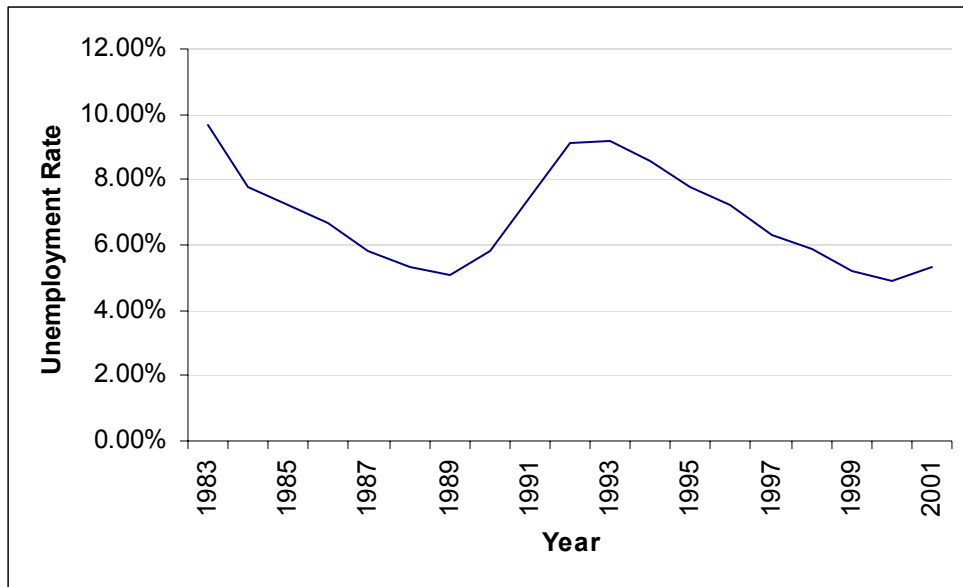


⁷ Specifically, we used the CPI-Urban Excluding Medical Expenditures Not Seasonally Adjusted for the United States to deflate to the current base year of 1982-1984 dollars.

4.2 Employment Variables

The loss or gain of employment is an obvious factor in accessions and terminations to welfare. The time period we are studying is one that encompasses an entire economic cycle. California's economy was quite strong during the late 1980's, followed by a recession in the early 1990's. Recovery began in 1993 as the unemployment rate fell from over 10% to less than 5% in 2000. Figure 4.10 shows the monthly California unemployment rate.

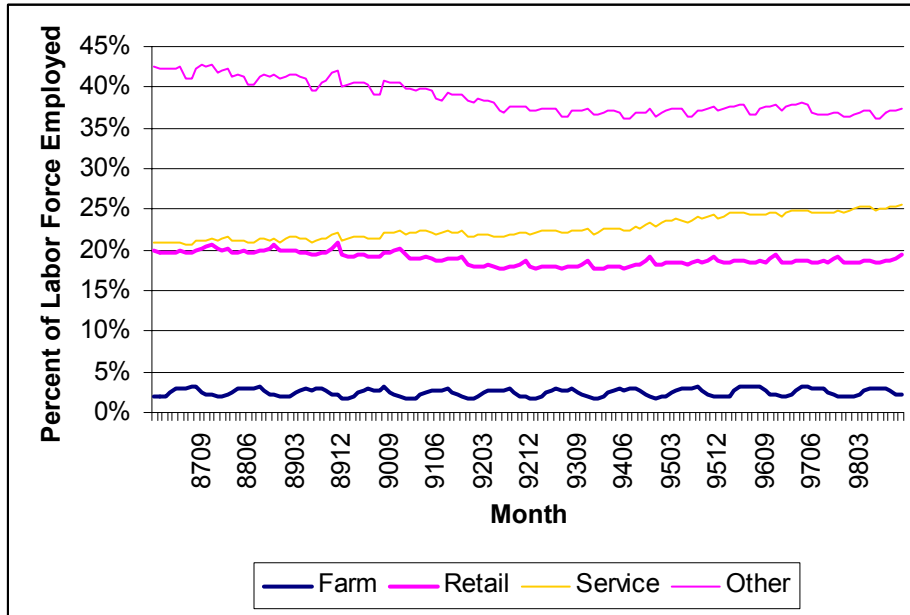
Figure 4.10 California Unemployment Rate



4.2.A. Employment by Industry Groups

Using the overall California unemployment rate however clouds over the diversity of the state's economy. Figure 4.11 shows the share of the labor force employed in four industry groupings over time: farm, services, retail trade, and other. Retail and farm employment rates seem fairly stable over time, but experience yearly cycles. The service industry seems to show growth over the period, especially during the later half of the 1990's when other industries seem to show declines.

Figure 4.11: California Employment by Industry Group



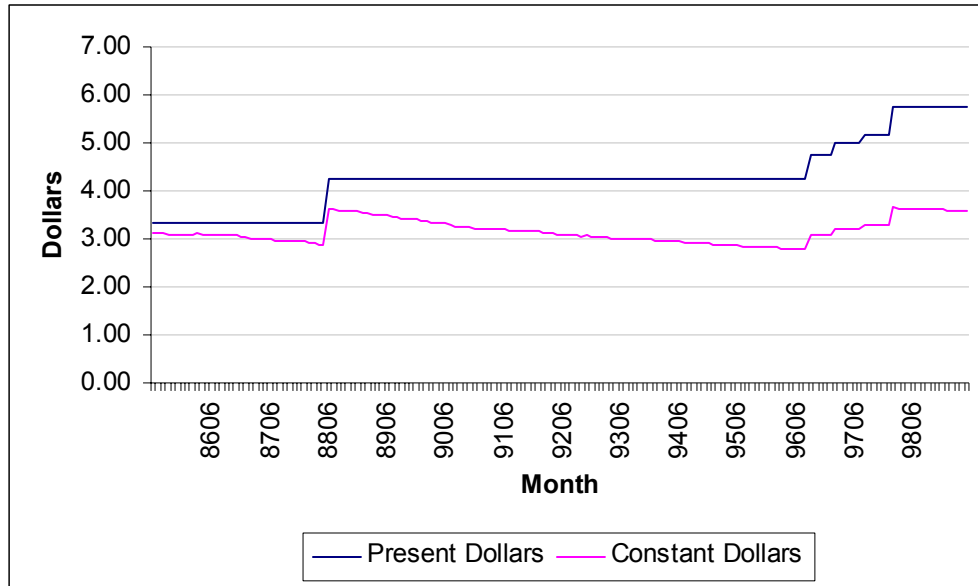
While the state's overall economy shows great diversity by industry, looking at the county level reveals even more diversity in economic activities. In some counties, farming may be the most common industry, while in other more urban counties services may be more relevant. In the analysis of Chapters 5 and 6 we use county level monthly data for four industry categories to consider job opportunities that are available to that particular county's workforce. Specifically we use employment rates that are the proportion of the county's labor force employed in that industry category. By using this form of employment rate we avoid any multicollinearity problems since the rates for all four industry groupings can simultaneously increase or decrease as a result of labor force changes.

4.2.B California Minimum Wage

The minimum wage in California has legislated values both from federal and state government. Usually the California minimum wage is higher than the prevailing federal minimum wage. The California minimum wage has only been adjusted in legislation four times in the 15 years that our data cover. Two of those increases were as a result of the federal minimum wage being set higher than the state minimum wage. In 1985 the minimum wage was \$3.35 and after four

increases was raised to \$5.75 in 1999 (See Figure 4.12). In constant dollars the minimum wage has stayed quite low: \$3.16 on average. We include minimum wage in our analysis as a proxy for wages that the welfare population would most likely earn working.

Figure 4.12: California Minimum Wage



4.3 County Demographics and Characteristic Variables

County demographic data are available from California’s Department of Finance. The data include total county population data as well as county population disaggregated by race, gender, and age categories. The diversity among counties lends itself to find a classification system that groups similar counties together. Counties may show similarity in terms of their land use. Our analysis takes into account county demographics and characteristics by using both demographic data and a county typology.

4.3.A Population Data by County and Race

Total population data are available annually at the county level. Monthly values were calculated by interpolating linearly between annual values. Population data enter into our analysis in several places. First, we aggregate all females aged 15 to 44 to construct a value for females who might already be on assistance or at risk of becoming on aid. Our population at risk of becoming on aid is this count of females subtracted by those already on assistance.⁸

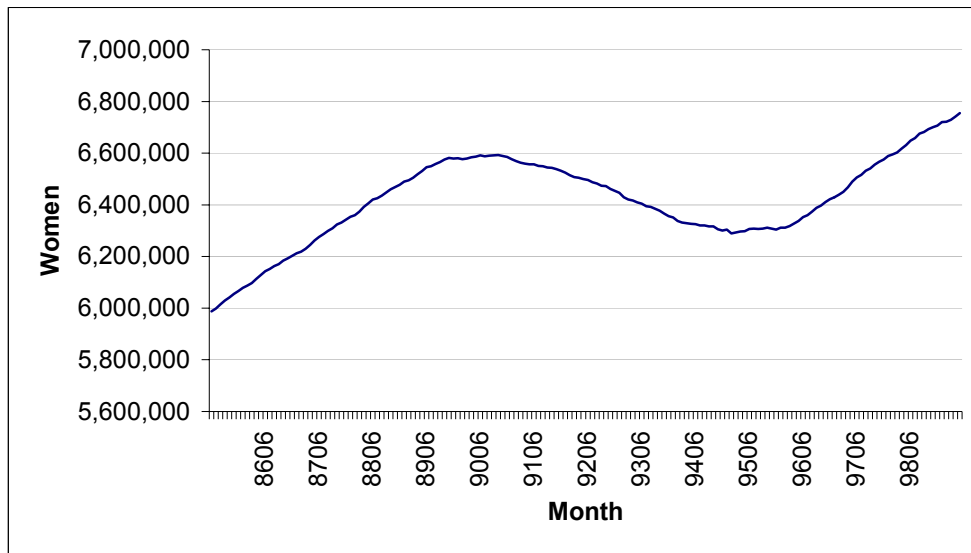
We next calculate the percentages of the female population at risk of giving birth who are black and Hispanic. The female population at risk of giving birth is the count of all females aged 15 to 44 minus the number of aged 0 children (to account for women we have already given birth recently). Counties exhibit very different racial patterns. Nevada County has the lowest percentage of Blacks (0%) whereas Alameda County has the highest percentage (18.4%).

⁸ Population at Risk of Becoming on Aid (PFR) = (Total Females Aged 15 to 44) - (AFDC-U caseload) - (AFDC-FG caseload)

Trinity County in the northern part of the state has the lowest percentage of Hispanics (3.6%). Imperial County, on the border with Mexico in the south, has the highest percentage of Hispanics (71.9%)

Over our study period the California population rose from 27 million people in 1985 to over 34 million people in 1999. Figure 4.13 shows that the shape of the line representing the population at risk of becoming on aid mimics the shape of the caseload over time.

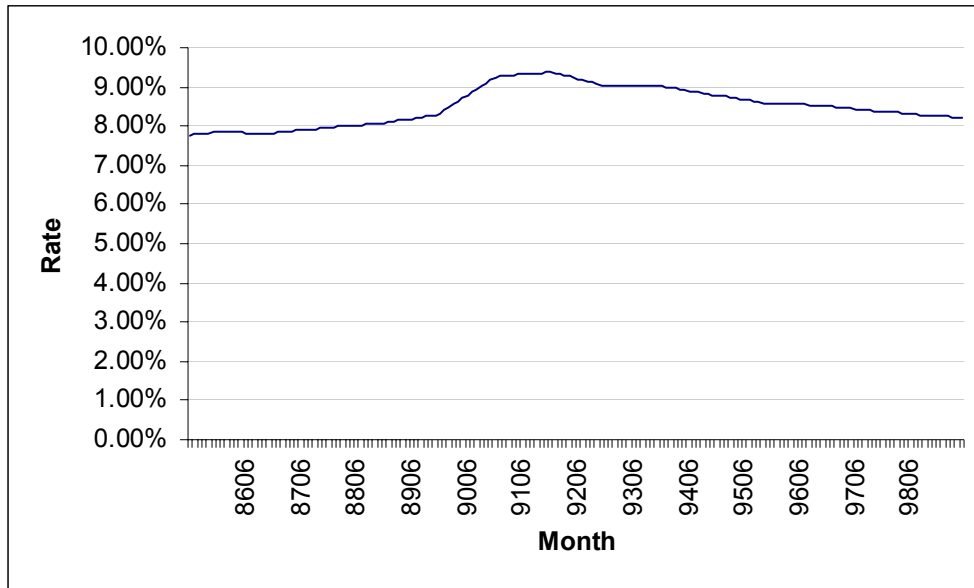
Figure 4.13: Population at Risk Of Becoming on Aid



4.3.B. County Birth Rates

Birth rates are an important factor in AFDC caseloads because AFDC receipt depends on the presence of a child. Births can also be a factor in families choosing to remain on aid. Birth rates are calculated by dividing the number of persons aged zero normalized by the population at risk of becoming on aid. The birth rate over this time period rose from slightly under 8% in 1985 to almost 9.5% in 1992. After that the birth rate fell again to slightly over 8% in 1999.

Figure 4.14: California Birth Rate



4.3.C. County Typology

Variation in California's counties lends itself to developing a county typology. Our typology follows Brady et al. (2000). This categorization is based on the amount of agricultural employment and the amount of the population that lives in rural areas. This definition of rural is the U.S. Census definition⁹. Counties are categorized into one of four groups: agricultural, urban, mixed, and rural. Agricultural counties are counties with more than 11.5 percent of agricultural employment. Urban counties are those with low agricultural employment and also a low percentage of rural population. Rural counties are those with low agricultural employment and more than 50 percent of rural population. The remaining counties are mixed counties and are those counties that a hybrid of rural and urban counties with low agricultural employment. Table 4.2 lists the 58 counties and the classification categories they fall into. It should be noted that not all of the 12 urban counties according to this typology fall into the state's Region 1 urban county classification.¹⁰

⁹ The Census definition of Percent Rural is the percentage of the population that lives in places with less than 2500 population.

¹⁰ Remember that beginning in 1997 the state classified 17 counties as urban counties with higher maximum aid benefit levels.

Table 4.2: Typology of California Counties

AGRICULTURAL	URBAN	MIXED	RURAL
Colusa	Alameda	Butte	Alpine
Fresno	Contra Costa	Humboldt	Amador
Glenn	Los Angeles	Napa	Calaveras
Imperial	Marin	Placer	Del Norte
Kern	Orange	Riverside	El Dorado
Kings	Sacramento	San Joaquin	Inyo
Madera	San Bernardino	San Luis Obispo	Lake
Merced	San Diego	Santa Barbara	Lassen
Modoc	San Francisco	Santa Cruz	Mariposa
Monterey	San Mateo	Shasta	Mendocino
San Benito	Santa Clara	Sonoma	Mono
Sutter	Solano	Stanislaus	Nevada
Tehama		Ventura	Plumas
Tulare		Yolo	Sierra
Yuba			Siskiyou
			Trinity
			Tuolumne

Seasonality may be an important consideration in county typology. This is especially true when we consider counties whose economies are highly tied to agricultural schedules. Our analyses in future chapters will include county type and season interaction terms. Season is specified as a summer dummy for the months May through October.

Chapter 5 Empirical Section: Estimation

Our goal is to develop the best possible model for predicting accession and termination rates. We start by estimating simple correlations. We then develop more sophisticated models by adding many of the elements and techniques described in Chapter 3. We conclude with a model which we believe most accurately describes the processes underlying the accession and termination rates.

The outcome variables of interest in this chapter are county-level accession and termination rates. In particular, we estimate the impact of AFDC benefit level changes on these outcomes. Accession and termination rate models are estimated separately for family grant (AFDC-FG) and unemployed parent (AFDC-U) cases. The basic unit of analysis for the estimation is a county-month observation.

The estimation sample for the models consists of data for 56 California counties from October 1985 to December 1996.¹ Although data are available for accession and termination rates beyond 1996, we limit our estimation sample to the months from October 1985 through December 1996 in order to leave some observations for the unconditional out-of-sample forecasting (see Chapter 6). Two of California's 58 counties, Alpine and Sierra, are excluded from the analysis because industry-specific employment data for these counties were not available for any part of the estimation period. These counties comprise only 0.015% of California's population, 0.016% of California's AFDC-U caseload, and 0.012% of California's AFDC-FG caseload. All other 56 California counties had employment data that became available before or during the estimation period and are included in the analysis.

Chapter 3 described methodologies for determining the impact of a dichotomous (i.e. 0/1) treatment variable. We transform the AFDC benefit levels into a dichotomous variable indicating when there were increases in nominal benefit levels since the last legislatively set level. This dichotomous treatment is equal to 1 for all periods except September 1991 through October 1998 when benefits declined. During the period of declining nominal benefits, the variable is equal to 0. The same analysis can also be used to determine the impact of a continuous treatment variable such as the level of AFDC benefits. In the following sections, we analyze both types of treatments (i.e. dichotomous and continuous). The first treatment is a dichotomous treatment: the indicator variable is equal to 1 whenever there is an increase in benefits in nominal dollars and 0 otherwise. The second treatment is a continuous variable measuring the maximum aid benefit in constant dollars. The maximum aid benefit is set for a family of three for AFDC-FG cases and for a family of four for AFDC-U cases; this reflects the mean observed sizes for these cases. We expect that both the dichotomous and continuous treatment variables should be positively correlated with accessions and negatively correlated with terminations.

While the dichotomous treatment seems very similar to traditional experiments where otherwise similar participants are randomly assigned to either the treatment or control groups, the following

¹ We have data beginning in July 1985. However, because a substantial number of models estimated in this chapter use three lags of employment, we lose July, August and September 1985 in models with lags. For consistency, all models were estimated from October 1985 except where indicated.

sections will describe how the situation differs in important ways which affect the way that we analyze and interpret the data.

5.1 Correlation & Bivariate Analysis

There are two types of simple analyses that can be conducted with only the four outcome variables (AFDC-U and AFDC-FG accession and termination rates) and the two treatment variables (dichotomous, continuous). The first is a simple correlation analysis showing the correlation coefficient between each of the outcome and treatment variables. The second is the bivariate regression designated as Model (1) in Chapter 3. The difference between the two analyses is the inclusion of a constant (intercept) term in the bivariate regression.

$$(1) Y_t^* = \alpha^* + \beta^* D_t^* + \varepsilon_t^*$$

Table 5.1 shows the correlation coefficients for each of the four outcome variables correlated with each of the two treatment variables. These correlations were computed for data pooled across all 56 counties. We expect that benefits should be positively correlated with accession rates and negatively correlated with termination rates. The overall correlation coefficients shown in Table 5.1 do not have the expected sign. Instead, the treatments are negatively correlated with the accession rates for both AFDC-U (ARU) and AFDC-FG (ARFG) cases and positively correlated with the termination rates for AFDC-U (TRU) and AFDC-FG (TRFG) cases.

Table 5.1: Correlation Analysis on Pooled Data

	ARU	ARFG	TRU	TRFG
Dichotomous: Aid Increase	-0.0617	-0.1026	0.2513	0.0406
Continuous: AFDC Benefit Levels	-0.0701	-0.1194	0.2318	0.0123

The same correlation analysis was performed on each of the 56 counties separately. Rather than show each correlation coefficient, we summarize the results in Table 5.2. Most of the individual county correlation coefficients also have the wrong sign for both the dichotomous and continuous treatment variables. The percent of correlation coefficients with the correct² sign ranges from 11% to 23%.

Table 5.2: Correlation Analysis on Individual Counties

	ARU	ARFG	TRU	TRFG
Dichotomous: Aid Increase				
Percent of Counties with Correct Sign	21.4%	23.2%	10.7%	42.9%
Continuous: AFDC Benefit Levels				
Percent of Counties with Correct Sign	23.2%	17.9%	14.3%	44.6%

² We refer to "correct" or "right" signs when we mean that the sign on coefficient is of the expected sign (either positive or negative) as described in this report.

Although the bivariate regressions include a constant in the regressions so that our coefficients are different, the percent of coefficients with the correct sign is the same as in the correlation analysis (see Table 5.3). This result is not unexpected. Recall that the coefficient from the bivariate regression (β) is simply the covariance of the outcome and treatment normalized by the variance of the outcome.

$$\beta = \text{Cov}(D^*, Y^*) / \text{Var}(Y^*)$$

Table 5.3: Bivariate Regression Analysis on Individual Counties

	ARU	ARFG	TRU	TRFG
Dichotomous: Aid Increase				
Percent of Counties with Correct Sign	21.4%	23.2%	10.7%	42.9%
Continuous: AFDC Benefit Levels				
Percent of Counties with Correct Sign	23.2%	17.9%	14.3%	44.6%

The fact that the coefficients on the treatment variables do not have the correct sign in both the correlation and bivariate analyses is not necessarily a concern at this point. The simple correlation and bivariate regressions may be too simplistic to capture the true *causal* relationship. For example, we found that nearly 80% of the counties had the wrong sign in the accession rate regressions for AFDC-U cases. These negative coefficients indicate that the county-level accession rates and maximum aid were moving in opposite directions.

However, this relationship is not necessarily causal. Confounding factors such as those mentioned in Chapter 4 may be simultaneously affecting accession rates. In traditional experiments with random assignment and where only the treatment variables are manipulated, a bivariate regression would be sufficient to determine an impact. However, neither the dichotomous nor the continuous treatment variables are randomly assigned in this case. Nor are the treatments the only variables which change over the sample. Each of the counties experienced changes in employment, minimum wage, and demographics during the period of analysis. Each of these factors could also have affected accession rates. The omission of these variables in the bivariate model is critical if any of these variables is correlated with the aid levels. Omission of correlated variables can bias the estimation of the treatment impact.

We should also note that the adjusted R^2 is generally quite low in all the bivariate regressions. The low adjusted R^2 reinforces the possibility that other factors should be considered. A low R^2 indicates that little of the variation in the outcome variables is explained by variation in benefits. This finding is consistent with the hypothesis that other factors play an important role in determining accession and termination rates.

5.2 Multivariate Analysis

Clearly, the model needs to be enhanced in order to better explain the movements in accession and termination rates. We now specify a multivariate model to include additional factors, which we hypothesize affect the accession and termination rates.

$$(2) Y_t^* = \alpha^* + \gamma^* X_t^* + \beta^* D_t^* + \varepsilon_t^*$$

We include two types of continuous explanatory variables: employment and demographic factors. In addition, we also include additional indicator variables for seasonality (month) and time (year), and other relevant county characteristics.

Employment-related factors affect the decision to enter and exit welfare by providing changing economic opportunities to current and potential participants. Improvements in economic conditions should lower accession rates and improve termination rates. However, general economic indicators such as the state-level unemployment rate may be too aggregate to pick up changes in economic opportunities in particular counties or for persons on welfare or considering welfare. For that reason, it is more relevant to focus on less aggregate indicators such as county-level or industry-specific employment measures. In the following regressions, we include county-level employment-labor force ratios for three industries relevant to the welfare population: farm, retail and services. The remaining industries are aggregated into a single “other” category. The correlation between these employment categories is generally less than 0.5 with the exception of retail and services which are correlated at 0.57. Since changes in the economic conditions can affect accession and termination rates either immediately or with a lag, the current and three months of lags for each of the employment-labor force ratios are included as explanatory variables. While the employment variables are not highly correlated across employment categories, within each category current and lagged employment observations are highly correlated. The correlation coefficients are always above 0.87. Such a high level of collinearity may cause some difficulties in the estimation particularly with respect to the coefficients on the employment variables.

Decisions regarding welfare participation can be affected not only by the availability of employment opportunities, but also by the value or return to those opportunities. Increases in the minimum wage legislated between 1985 and 1996 affect the accession and termination rates by making employment more attractive. The minimum wage in constant dollars measures the return to hourly employment, which is an alternative source of income for welfare recipients. Therefore we include the value of the legislated minimum wage in constant dollars for the state of California in the multivariate regressions.

Demographic factors also affect accession and termination rates. Evidence from previous studies indicates that non-white populations have different patterns of welfare use and face different employment opportunities. While race and ethnicity do not themselves cause these different patterns, they often proxy for other socio-economic variables that are correlated with welfare use. Examples of these variables include education and income. Unfortunately, we often do not have the variables to measure a group’s socio-economic status. As a result, the variation in the racial and ethnic make-up of county populations, our proxy for socio-economic status, may explain some of the variation in accession and termination rates. The multivariate model below includes the percent of the female county population aged 15-44 that is black and the percent Hispanic. Although we do not have a clear notion of the expected sign of these variables, the weak hypothesis used here is that the percent Hispanic and percent black is positively associated with accession rates and negatively associated with termination rates.

Another important demographic factor is the birth rate. Since eligibility for AFDC is determined by the presence of a child in the household, birth rates can affect accession and termination rates. County-level birth rates are defined as the county-level births normalized by the county's female population at risk. Birth rates should have a positive impact on accession rates since a birth can make a household eligible for AFDC. In addition, we expect that births should have a negative impact on termination rates since second births can extend eligibility for a household currently in the caseload.

Finally, as noted in Chapter 4, there are seasonal variation and trends to welfare participation that must be taken into account. Figures 4.4 and 4.5 show that there is monthly variation in accession and termination rates, respectively. The monthly variation is due to the fact that accession levels are generally much higher during the winter months than during other months and termination levels are much higher during the summer months. This is not surprising given the seasonal nature of some employment. To account for this seasonality, the model includes indicator variables for each month of the year; we will denote these month indicator variables as X^{4*} . The model also includes indicator variables for each year (X^{5*}) in the sample to account for other unobservable trends. These can include changes in sentiments toward welfare use or changes in government policies or spending.

The specified model contains current and lagged independent values for the economic factors (X^{1*}), but only current values for the treatment (D^*), demographics (X^{2*}), California minimum wage (X^{3*}), and the month (X^{4*}) and year (X^{5*}) dummies:

$$(3) Y_t^* = \alpha^* + \beta^* D_t^* + \gamma^* X_t^{1*} + \gamma_1^* X_{t-1}^{1*} + \gamma_2^* X_{t-2}^{1*} + \gamma_3^* X_{t-3}^{1*} + \gamma_4^* X^{2*} + \gamma_5^* X^{3*} + \gamma_6^* X^{4*} + \gamma_7^* X^{5*} + \varepsilon_t^*$$

Tables A1.a through A8.a in the Appendix show the detailed summary statistics for the county regression results aggregated; the results in these tables are summarized within each county. The included statistics are (1) the percent of coefficients that have the correct sign, (2) the percent of coefficients that have the correct sign and are significant, (3) the percent of coefficients that have the incorrect sign, (4) the percent of coefficients that have the incorrect sign and are significant, (5) the (out-of-sample) average mean square forecast error³, (6) the adjusted R-squared, and (7) the number of observations available. Tables A1.a through Table A4.a show the summary statistics for each set of county-level dichotomous treatment regressions: accession rate AFDC-U cases, accession rate AFDC-FG cases, termination rate AFDC-U cases, and termination rate AFDC-FG cases. Tables A.5.a through A.8.a show similar summary statistics for the continuous treatment regressions.

Also in the Appendix, Table A1.b through Table A4.b show the summary statistics across counties and within each coefficient. These tables include the number and percent of counties for which each variable has a coefficient with the correct sign and the percent of counties for which each variable has a coefficient with the correct sign and significance. The results confirm that the most problematic coefficients are the employment-labor force ratios. Due to

³ The average mean squared forecast error is a measure of the error in the out-of-sample forecast when compared to the actual values realized. The outcome variables were predicted for the months from January 1997 to December 1998. We will discuss the average mean squared forecast error in Chapter 6 which focuses on forecasting issues.

multicollinearity across employment categories and across lags of each industry employment, the coefficients on these variables often have the wrong sign.

Individual results for the multivariate regressions for individual counties can be found in the Appendix Tables. The median results for the summary statistics for the 56 county regressions are shown in Table 5.4.

This table reports the percentage of coefficients that have the expected sign and the smaller percentage of the coefficients that have the expected sign and are found to be statistically significant across all 56 counties. Similarly it reports the percentage of coefficients that have the opposite sign from what we would expect and the percentage of all coefficients that have the incorrect sign and are also statistically significant. These percentages are taken over all dependent variables and all counties in our regressions. The sign and significance of coefficients is considered not only for the treatment variable (each individually) but also for the included covariates: employment variables, minimum wage, birth rates, percentage black and percentage Hispanic. Table 5.4 provides a general overview of whether our estimations for each county give coefficients of the correct sign and their significance. More information about the estimations is provided in Table 5.5, which discusses the treatment variable only, and in Appendix Tables A1.b through A8.b.

Results for the dichotomous and continuous treatment regressions are strikingly similar. Looking at the median across counties, a majority of the coefficients have the correct sign. However, few are significant. Likewise, hardly any of the coefficients with the incorrect sign are significant. At the median, nearly 60% of the variation in the accession and termination rates for AFDC-U cases is explained by the model. The corresponding percent for AFDC-FG accession and termination rates is slightly less than 50%.

Table 5.4: Median of Summary Statistics Across Coefficients for 56 County Time-series Regressions

	ARU	ARFG	TRU	TRFG
Dichotomous Treatment Regression				
% Coefficients: Right Sign	62%	57%	57%	62%
% Coefficients: Right Sign & Significant	5%	5%	5%	5%
% Coefficients: Wrong Sign	38%	43%	43%	38%
% Coefficients: Wrong Sign & Significant	0%	0%	2%	0%
Adjusted R-squared	0.56	0.47	0.59	0.47
Continuous Treatment Regression				
% Coefficients: Right Sign	62%	57%	57%	62%
% Coefficients: Right Sign & Significant	5%	5%	5%	5%
% Coefficients: Wrong Sign	38%	43%	43%	38%
% Coefficients: Wrong Sign & Significant	2%	0%	0%	0%
Adjusted R-squared	0.56	0.46	0.59	0.47

Table 5.5 summarizes the results of the 56 individual counties for the treatment variables only. The most notable difference from the bivariate regressions is that now both the dichotomous and

continuous treatment variables have the expected sign in each of the 56 county regressions. Both the dichotomous and continuous benefit variables are always positively associated with accession rates and negatively associated with termination rates for AFDC-U and AFDC-FG cases. This consistency indicates the kind of constant conjunction described in Chapters 1 and 2 as the requirement for the neo-Humean theory of causality. However, the coefficients are generally insignificant at the 5% level indicating that in the majority of cases we cannot infer a non-zero impact of benefits on accession and termination rates.

Table 5.5: Summary of Regression Results for Individual Counties: Treatment Coefficients Only

	ARU	ARFG	TRU	TRFG
Dichotomous Treatment Coefficient				
Percent of Counties: Right Sign	100.0%	100.0%	100.0%	100.0%
Percent of Counties: Right Sign & Significant	21.4%	10.7%	21.4%	7.1%
Continuous Treatment Coefficient				
Percent of Counties: Right Sign	100.0%	100.0%	100.0%	100.0%
Percent of Counties: Right Sign & Significant	5.0%	12.5%	12.5%	10.7%

For the accession rates, the dichotomous treatment is significant for only 21% of counties in the AFDC-U regressions and for only 11% of the counties in the AFDC-FG regressions. The corresponding percentages for the continuous treatment are 5% and 13%. The results for the termination rate models are similar with significant coefficients on the dichotomous treatment for only 21% of counties in the AFDC-U regressions and 7% of counties in the AFDC-FG regressions. For the continuous treatment, the percent significant are 13% and 11% for AFDC-U termination rate regressions and AFDC-FG termination rate regressions, respectively.

Although we do not present a separate table here for each independent variable, as we did above with the treatment variables, we can examine each variable in a similar way. Our regressions produce expected and consistent results more often for some variables than others. There is substantial variation in the sign of the employment variables. Generally, each employment-labor force ratio has the correct sign in only slightly more than half of the counties. No one employment variable is simultaneously correct and significant in more than 11% of the counties. See Appendix Tables A1.b through A8.b for details. As noted above, this variation is expected since the current and lagged values of the employment variables are multicollinear. A correlation matrix of the current and lagged variables for each industry shows correlations no lower than 0.87. There is also correlation across industry types: retail and services have a correlation coefficient of 0.57.

In the dichotomous treatment regressions, the coefficient on the minimum wage has the correct sign for 43% of the counties in the AFDC-FG termination rate regressions and 73% in the AFDC-U accession rate regressions. The range for the continuous treatment regressions is 46% for AFDC-FG termination rate regressions and 80% for AFDC-U accession rate regressions. In general, the California minimum wage is significant in only 10% of the regressions.

The coefficient on percent Hispanic has the expected sign in only half of the regressions and is generally insignificant. The percent black always has the expected sign, but is rarely significant.

Like the treatment variables, the birth rate has the expected sign for all counties and outcome variables. It is also the most consistently significant variable in the regressions. For the dichotomous treatment regressions, the birth rate is significant in 13% of counties for the AFDC-U accession rate, 9% for the AFDC-FG accession rate, 18% for the AFDC-U termination rate, and 9% for the AFDC-FG termination rate. For the continuous treatment regressions, the percentage of significance for birth rates is 20% for the AFDC-U accession rate, 13% for the AFDC-FG accession rate, 16% for the AFDC-U termination rate, and 16% for the AFDC-FG termination rate. See Tables A1.b to A8.b in the Appendix for additional details on each of the regressors.

5.3 Panel/Pooled Data Analysis

The multivariate models of the previous section outperform the bivariate models with respect to finding the expected sign on the two treatments. However, the results for the treatments indicate that while increases in benefits are always associated with increases in the accession rates and with decreases in the termination rates, the coefficients on both treatments are generally not significant in the individual county regressions.

Lack of significance may indicate that there is no *causal* relationship or impact of the treatment on the outcome of interest. Or more formally, it may indicate that the impact of the treatments cannot be differentiated from zero. Alternatively, it may simply indicate that the regression lacks sufficient power. Recall that significance is determined from the value of the t-statistic. The t-statistic is the ratio of the point estimate for the coefficient over the standard error of the coefficient. The standard error is a function of the number of observations used. In the individual county multivariate regressions, the median number of observations is 135.⁴ However, we also have a substantial number of coefficients to estimate. Each regression has a total of 44 explanatory variables and hence 44 coefficients to be estimated.⁵ As a result, many of our coefficients may be imprecisely measured and have large standard errors.

5.3.A Fixed Effects

Pooling the data into a panel dataset with variation across both time and counties increases the number of available observations for estimation and so increases the precision of the coefficient estimates. The number of observations increases while the number of coefficients to be estimated does not.

⁴ In some counties, fewer observations were used because employment data were not available until later in the estimation sample. The smallest sample used was comprised of 45 observations - this occurred for only 2 of the 56 included counties.

⁵ The 44 explanatory variables included in Model (3) are 16 variables for current and lagged employment sectors, 2 ethnicity/racial categories, 1 birth rate, 1 minimum wage variable, 1 benefit/treatment, 11 month indicator variables, 11 year indicator variables and 1 constant.

While panel data has the benefit of increasing the number of observations and consequently the precision of the estimates of common variables, it may also complicate the estimation. In the county-level multivariate regressions, we used variation across time within a single county to estimate the impact of benefits on accession and termination rates. With a panel, however, there is variation across both time and counties. This additional variation can be problematic because counties may differ in ways that can undermine the results of the analysis. However, in some cases, such differences between counties may be overcome by using fixed effects.⁶

Counties differ from one another in many ways that are observable and quantifiable. Some of these observable and quantifiable differences were included in our multivariate analysis above. Differences across counties in employment-labor force ratios affect welfare use. Including measures of these differences helps account for across county variation when estimating models using panel data. However, as discussed in Chapter 3, counties may also differ in unobservable characteristics, which cannot be quantified in a variable and yet may affect welfare participation. Such characteristics can include intangibles such as social norms regarding welfare use. For example, there may be greater stigma attached to welfare use in small towns where people are more likely to know each other. If so, then counties dominated by small towns rather than larger urban areas may have lower welfare use despite similar economic situations. Such characteristics are difficult to quantify in a variable which we could include in the regression. However, if the unobservable characteristics are time invariant, a fixed effects regression can eliminate these characteristics without requiring a quantifiable variable. Fixed effects allows the intercept to vary for each of the counties included in the regression. These intercepts capture the effects of all time-invariant characteristics of the counties. The individual county intercepts are estimated by including county-specific indicator variables. This technique increases the number of coefficients by the number of counties which is less than the increase in the number of observations. This process eliminates between county variation so that the panel estimation now uses only variation within each of the counties thus eliminating unobservable county characteristics which may confound the estimation and inference. However, it is important to note that if the unobservable differences between counties are time-varying, fixed effects may not properly estimate the impact of a treatment.

Table 5.6 shows the results of a fixed effects estimation. The panel data regressions are estimated for the continuous treatment only.⁷ The model specification is identical to the county-level multivariate time series regression in the previous section. However, because the model is estimated using panel data, it also includes county-specific intercepts to remove the unobservable “fixed effects”.

⁶ It is important to note that the fixed effects regression does not estimate the impact by using variation across counties. Rather, the fixed effects regression relies solely on within county variation to identify an impact. See Section 3.3.B for further details on fixed effects.

⁷ The continuous treatment is preferred because regression analysis links the variation in the outcome to variation in the explanatory variable. The dichotomous variable, which was constructed from the continuous variable, ignores much of this variation.

Table 5.6: Fixed Effects Regression: Continuous Aid

	ARU	ARFG	TRU	TRFG
Continuous Benefits	4.96E-06***	2.01E-04	-4.90E-05	-2.90E-05
Farm Emp (t)	-3.42E-03***	5.56E-03***	1.78E-01***	2.18E-02*
Farm Emp (t-1)	-4.42E-03***	-7.10E-04	1.00E-01**	6.22E-03
Farm Emp (t-2)	3.61E-04	5.25E-04	-8.92E-03	1.50E-02
Farm Emp (t-3)	5.78E-03***	2.30E-03	-1.45E-01***	-7.26E-03
Service Emp (t)	-3.31E-03*	-1.53E-03	4.07E-02	1.01E-01***
Service Emp (t-1)	-1.34E-03	-2.77E-03	1.04E-01	-8.61E-03
Service Emp (t-2)	1.74E-03	4.34E-03	1.45E-01	-7.00E-02*
Service Emp (t-3)	1.67E-03	-2.30E-03	-1.67E-01***	-2.55E-02
Retail Emp (t)	1.51E-03	-7.41E-03	5.39E-02	9.87E-02**
Retail Emp (t-1)	-5.17E-03	5.17E-03	-4.48E-02	-4.43E-02
Retail Emp (t-2)	-2.10E-03	7.10E-04	-1.74E-01	-3.09E-02*
Retail Emp (t-3)	6.30E-03**	8.59E-03	3.48E-01***	8.13E-02
Other Emp (t)	-4.79E-03***	-2.21E-03	5.33E-02	1.40E-02
Other Emp (t-1)	-1.32E-03	8.57E-04	6.07E-02	1.45E-03
Other Emp (t-2)	8.39E-04	-2.37E-03	2.12E-02	-2.72E-02
Other Emp (t-3)	3.23E-03**	1.75E-03	-2.16E-01***	1.28E-02
CA Minimum Wage	-3.80E-04***	-5.10E-04**	7.35E-03*	3.80E-04
Percent Hispanic	7.04E-03***	-1.01E-02***	7.58E-02	1.11E-01***
Percent Black	8.02E-03***	-3.75E-02***	-4.65E-01***	-1.37E-01***
Birth Rate	3.23E-02***	3.53E-02***	-3.29E-01***	-1.88E-01***
Month Indicators	Yes	Yes	Yes	Yes
Year Indicators	Yes	Yes	Yes	Yes
County Indicators	Yes	Yes	Yes	Yes

Significant at 1% confidence level ***; 5% **, and 10% *.

The coefficient on the continuous treatment indicates that the level of benefits does not significantly affect the accession or termination rates, however we find coefficients of the expected sign. The exception is the accession rate for AFDC-U cases; although the impact for the AFDC-U accession rate is significant at the 1% level, it is extremely small.

As with the individual county regressions, the employment variables are not consistent with respect to the direction of their impact. AFDC-U accession and termination rates respond more to economic indicators than AFDC-FG rates. This finding is consistent with AFDC-U population's greater attachment to the labor force and consistent with their greater flexibility in childcare arrangements because both parents are present in the household.

Increases in California's minimum wage are associated with lower accession rates and higher termination rates. These results are significant except for the AFDC-FG termination rate.

Percent Hispanic is associated with greater accession rates for AFDC-U. However, this same group is associated with lower accession rates for AFDC-FG cases and higher termination rates for both AFDC-U and AFDC-FG termination rates. Percent back is associated with higher

accession rates for AFDC-U cases and lower termination rates for both AFDC-U and AFDC-FG cases. However, it is also associated with lower accession rates for AFDC-FG cases. These results are difficult to interpret, but as we noted earlier, it is difficult to formulate clear hypotheses about how the percent black and Hispanic will affect accession and termination rates.

As expected, increases in the birth rates significantly increase accession rates and lower termination rates for both AFDC-U and AFDC-FG cases.

5.3.B Grouping Counties: County Type

In addition to county-specific intercepts, there may be other associations in panel data that should be taken into account. There may be groups of counties with similar observable characteristics. For California, a natural grouping is county type. In Brady et al. (2000), the 58 California counties are grouped into four types based on percent agricultural employment and Census urban/rural classification: urban, rural, mixed and agricultural.

These county types are associated with different employment and welfare participation patterns. These patterns are particularly important during the summer months. To account for this, we now include interaction variables. County type indicator variables for each county type are interacted with indicator variables for summer months.⁸

⁸ In order to avoid multicollinearity, one of the interaction variables must be excluded. Urban*Summer is excluded in the regression. As a result, the coefficients on Agricultural*Summer, Rural*Summer, and Mixed*Summer must be interpreted with respect to the excluded Urban*Summer group. Recall from Chapter 4 that summer is defined as May through October.

Table 5.7 Fixed Effects Regressions with County Type Summer: Continuous Aid (Homoskedastic)

	ARU	ARFG	TRU	TRFG
Continuous Benefits	5.00E-06***	2.78E-06	-4.90E-05	-2.90E-05
Farm Emp (t)	-2.18E-03***	5.78E-03***	1.68E-01***	2.06E-02*
Farm Emp (t-1)	-3.81E-03***	-5.50E-04**	9.95E-02**	7.41E-03
Farm Emp (t-2)	4.66E-04	5.09E-04	-1.29E-02	1.37E-02
Farm Emp (t-3)	5.11E-03***	2.21E-03***	-1.34E-01***	-4.76E-03
Service Emp (t)	-3.25E-03*	-1.50E-03	3.21E-02	9.75E-02***
Service Emp (t-1)	-1.77E-03	-2.80E-03	1.24E-01	-2.00E-03
Service Emp (t-2)	2.06E-03	4.35E-03	1.34E-01	-7.32E-02**
Service Emp (t-3)	1.47E-03	-2.37E-03***	-1.69E-01***	-2.66E-02
Retail Emp (t)	1.46E-03	-7.36E-03	3.87E-02	9.29E-02**
Retail Emp (t-1)	-4.36E-03	4.96E-03	-1.27E-01	-7.38E-02
Retail Emp (t-2)	-1.29E-03	7.35E-04	-1.96E-01	-3.73E-02
Retail Emp (t-3)	5.75E-03*	8.73E-03***	4.16E-01***	1.06E-01**
Other Emp (t)	-4.40E-03***	-2.26E-03	1.56E-02	4.94E-04
Other Emp (t-1)	-1.60E-03	8.46E-04	6.52E-02	2.47E-03
Other Emp (t-2)	4.79E-04	-2.38E-03	2.57E-02	-2.63E-02
Other Emp (t-3)	3.29E-03**	1.79E-03***	-1.88E-01***	2.39E-02
CA Minimum Wage	-3.80E-04***	-5.40E-04*	7.37E-03*	3.89E-04
Agriculture*Summer	-6.10E-04***	-1.10E-04***	1.25E-02***	3.40E-03***
Mix*Summer	-1.30E-04*	-7.00E-05***	1.16E-02***	4.28E-03***
Rural*Summer	-3.90E-04***	-1.90E-05***	2.15E-02***	7.34E-03***
Percent Hispanic	7.03E-03***	-1.02E-02	7.69E-02	1.12E-01***
Percent Black	8.15E-03**	-3.76E-02***	-4.69E-01***	-1.38E-01***
Birth Rate	3.30E-02***	3.54E-02***	-3.33E-01***	-1.88E-01***
Month Indicators	Yes	Yes	Yes	Yes
Year Indicators	Yes	Yes	Yes	Yes
County Indicators	Yes	Yes	Yes	Yes

Significant at 1% confidence level ***; 5% **; and 10% *.

The county-type-summer interaction effects shift the intercept for each county type relative to the excluded group. The excluded group is urban-summer. We expect that accession rates would be lower and termination rates higher in the summer months. This is particularly true in the non-urban counties (the included groups) where seasonal employment is more prevalent. The interaction variables always have the expected sign and are significant. According to the results in Table 5.7, all three included county types experience lower accession rates and higher termination rates relative to urban counties during the summer months.

Again, the treatment variable has the right sign, but is generally not significant. The exception again is the AFDC-U accession rate where it is significant, but extremely small.

The results for the employment variables are similar to previous findings: the coefficients vary in sign. The minimum wage variable decreases accessions and increases terminations for both

AFDC-U and AFDC-FG cases. Not surprisingly, the results are stronger for AFDC-U cases where a stronger attachment to the labor force is expected.

The results for percent Hispanic are also similar to previous findings. Larger Hispanic populations are associated with higher AFDC-U accession rates and higher AFDC-FG termination rates. Percent black is associated with higher AFDC-U accession rates and lower termination rates for both AFDC-U and AFDC-FG cases. As in the previous regressions, birth rates are strongly and significantly associated with increased accession rates and decreased termination rates for both groups.

5.3.C Heteroskedasticity

In plotting the residuals from the homoskedastic regression, we find that the residuals are not constant across counties. This result indicates that the constant variance assumption has been violated. However, the pattern of heteroskedasticity indicates that the relationship between the variance and independent variables is not clear. It is a combination of several factors including county size and county type. Given the non-linear relationship, we will not specify a form for the heteroskedasticity. Rather we will simply correct for heteroskedasticity of the unknown form as discussed in Section 3.2.D.i.

The results in Table 5.8 have been adjusted for a pattern of heteroskedasticity. Recall that standard regression analysis assumes that there is a constant variance (i.e., homoskedasticity). However, with panel data which consists of multiple observations for individual counties over time, we may find that variance is not constant. Instead it may vary across counties. Accession and termination rates are likely to have more variance in smaller counties. This variance is due to small numbers of accessions and terminations in any given month. See Figures A1.a through A2.d in the Appendix for the distribution of variances for accession and termination rates by population.⁹

⁹ In Figures A1.a through A2.d, individual counties can be identified by the county number. The variances are graphed both with (Figures A1.a-A1.d) and without (Figures A2.a-A2.d) Los Angeles County (county #19) because of its extremely small variance and large population.

Table 5.8: Fixed Effects Regressions with County Type Summer: Continuous Aid (Heteroskedastic)

	ARU	ARFG	TRU	TRFG
Continuous Benefits	7.93E-08	6.67E-07	-8.50E-05***	-6.30E-05***
Farm Emp (t)	-2.57E-03***	4.60E-03***	1.24E-01***	1.90E-02**
Farm Emp (t-1)	-2.64E-03***	-1.94E-03	5.32E-02*	8.64E-03
Farm Emp (t-2)	-8.90E-04	2.48E-04	2.11E-02	2.04E-02*
Farm Emp (t-3)	5.15E-03***	3.61E-03***	-1.06E-01***	-1.01E-03
Service Emp (t)	-2.56E-03**	-3.45E-03	-6.67E-02	-1.24E-02
Service Emp (t-1)	-1.40E-05	-1.43E-03	1.81E-01*	5.11E-04
Service Emp (t-2)	8.13E-04	2.31E-03	-8.30E-02	-2.56E-02
Service Emp (t-3)	-8.40E-04	-4.55E-03**	1.50E-01	-8.72E-03
Retail Emp (t)	4.57E-03**	-3.64E-03	1.35E-01	7.65E-02*
Retail Emp (t-1)	-6.24E-03**	-8.00E-04	-1.10E-01	-8.07E-02
Retail Emp (t-2)	1.90E-04	4.64E-03	-2.71E-02	1.55E-02
Retail Emp (t-3)	2.82E-03	7.54E-04	6.16E-02	4.47E-03
Other Emp (t)	-1.57E-03*	1.22E-03	2.04E-02	-1.16E-02
Other Emp (t-1)	-1.90E-04	4.94E-04	9.88E-03	1.90E-02
Other Emp (t-2)	5.18E-04	-2.66E-03	-3.44E-02	-1.65E-02
Other Emp (t-3)	1.07E-03	3.03E-03	-9.34E-02*	1.96E-02
CA Minimum Wage	-1.50E-04***	-2.70E-04**	5.53E-03**	3.87E-04
Agriculture*Summer	-4.30E-04***	-8.10E-05	1.11E-02***	3.27E-03***
Mix*Summer	-7.80E-05***	4.41E-05	1.10E-02***	4.20E-03***
Rural*Summer	-3.10E-04***	-1.30E-05	2.25E-02***	7.48E-03***
Percent Hispanic	3.07E-03***	-9.34E-03***	-1.23E-02	-1.94E-02
Percent Black	5.33E-03**	-1.64E-02***	-5.71E-01***	-1.13E-01***
Birth Rate	3.31E-02***	4.65E-02***	-2.17E-01***	7.28E-03
Month Indicators	Yes	Yes	Yes	Yes
Year Indicators	Yes	Yes	Yes	Yes
County Indicators	Yes	Yes	Yes	Yes

Significant at 1% confidence level ***; 5% **; and 10% *.

Benefit levels are not significant in the accession rate regressions, but do have the expected sign. For both AFDC-U and AFDC-FG termination rates, the coefficients for the benefit level are significant and negative, but very small.

The heteroskedasticity adjustment reduces the number of significant employment variables. The county type and summer month interaction variables are still highly significant in the AFDC-U accession rate regressions and in both termination rate regressions.

The minimum wage variable significantly reduces accession rates and increases termination rates with the exception of AFDC-FG termination rates. Although the coefficient on minimum wage has the expected sign in the AFDC-FG termination rate regression, it is not significant.

Percent Hispanic and percent black are associated with significantly higher AFDC-U accession rates, but significantly lower AFDC-FG accession rates. Percent black is also associated with significantly lower termination rates for both AFDC-U and AFDC-FG cases. Although percent Hispanic is also associated with lower termination rates for both AFDC-U and AFDC-FG cases, these coefficients are not significant.

The results for birth rates continue to be strong with expected and significant impacts for three of the outcome categories. The exception is the termination rate for AFDC-FG cases where we cannot reject the null hypothesis that the impact of birth rates is zero.

5.4 Notes on Employment Variables and Coefficients

Throughout Chapter 5, we alluded to concerns regarding the inclusion of employment variables. In this section, we address some of these issues further. In particular, we investigate the use of employment versus month dummies, the multicollinearity of employment variables and aggregation of employment coefficients, and the choice of the number of lags to include in the model.

5.4.A Including Both Employment Variables and Month Dummies

Both the employment variables and the month dummies capture some of the same aspects of seasonality – particularly the seasonality of employment. It is worth investigating whether including both measures, rather than only one, significantly improves the estimation.

To investigate this issue, we re-estimate the homoskedastic models three times: once without the employment variables (Model 2), once without the month dummies (Model 3), and once without both employment variables and month dummies (Model 4).¹⁰ We then compare these models to the preferred model¹¹ from the previous section which contains both the employment variables and the month dummies (Model 1). All four models were estimated using both OLS and maximum likelihood estimation.¹²

Model 1: Both Employment Variables and Month Dummies (Preferred Model)

Model 2: Month Dummies Only

Model 3: Employment Variables Only

Model 4: Neither Employment Variables nor Month Dummies

Recall from Chapter 3 that the fixed effects estimator is a *within* estimator, so we will focus on the within R-squared as our goodness-of-fit statistic when comparing these models. Estimating

¹⁰ Note that we have returned to the use of the homoskedastic model for these comparisons. Some of the comparisons of models with a varying number of explanatory variables performed in this section are not valid for models which assume heteroskedasticity of the type used in Section 5.3.C above. The inclusion of additional explanatory variables changes the weights and hence the variance-covariance matrix in the heteroskedastic model such that the log-likelihoods used here are not comparable across models.

¹¹ The preferred model is the specification presented in Tables 5.7 (homoskedastic) and 5.8 (heteroskedastic).

¹² OLS and Maximum Likelihood (ML) give the same regression results but report different goodness-of-fit measures. OLS reports R-squared and ML reports log-likelihood. Specifically, the maximum likelihood estimation was done with the Stata procedure `xtgls` with the variance-covariance matrix set to the identity matrix.

the homoskedastic model using fixed effects, we find that Model 1, our preferred model, with both the employment variables and the month dummies has the highest within R-squared for all four outcome variables (see Table 5.9.a). The employment variables and the month dummies seem to be complementary in that their simultaneous inclusion improves Model 1's performance in terms of the within R-squared relative to both Models 2 and 3.

Table 5.9.a: Comparison of Models Using the Within R-squareds

	ARU	ARFG	TRU	TRFG
Model 1: Employment & Month	0.1798	0.1143	0.2526	0.1270
Model 2: Month Only	0.1591	0.1087	0.2302	0.1173
Model 3: Employment Only	0.1593	0.0860	0.2387	0.1183
Model 4: Neither	0.1276	0.0716	0.2027	0.1070

We also re-estimated each of the four models using maximum-likelihood estimation under the assumption of normal errors and homoskedasticity.¹³ Maximization of the likelihood function is akin to minimization of the sum of squared residuals in OLS. In maximizing the log-likelihood, we find the parameters which are most likely to produce the actual outcomes (i.e. our data). Rather than producing an R-squared as its goodness-of-fit statistic, this maximum-likelihood procedure generates a log-likelihood.

Model 1 has the highest log-likelihood for all four outcomes indicating that it produces the best fit (see Table 5.9.b).¹⁴ In order to ascertain whether the fit is *significantly* better for Model 1 relative to the other three models, we perform a likelihood-ratio test. A likelihood-ratio test is a test-statistic calculated from the log-likelihood for the unconstrained Model 1 ($\ln L_u$), the log-likelihood for the constrained comparison model ($\ln L_c$), and the difference in the degrees of freedom (j).¹⁵ Model 1 is the unconstrained model because the full set of coefficients is estimated by the model. Models 2, 3 and 4 are nested or constrained versions of Model 1 because the coefficients estimated are subsets of the coefficients in Model 1. For example, we can think of Model 2 as equivalent to a Model 1 where the coefficients on the employment variables are constrained to be zero. Likewise, Model 3 is a version of Model 1 where the coefficients on the month dummies are constrained to be zero. For Model 4, the coefficients on both the employment variables and the month dummies are constrained to be zero. It is important to note that valid comparisons can only be made between the unconstrained model and a nested model.

The likelihood-ratio test statistic is calculated as follows and is distributed as a Chi-Squared (χ^2).

$$\text{L-R Test Statistic} = -2[\ln L_c - \ln L_u] \sim \chi_j^2$$

¹³ The assumption of homoskedasticity is equivalent to estimation of the generalized least squares model with the omega matrix equal to the identity matrix.

¹⁴ Comparisons of log-likelihoods are only valid across nested models.

¹⁵ The difference in the degrees of freedom is simply the difference in the number of coefficients estimated by the two models being compared.

The likelihood-ratio statistics comparing Model 1 to each of the three models can be found in the first three panels of Table 5.9.b. Comparison of the likelihood-ratio test statistics with the Chi-Squared (χ^2) Distribution table shows that Model 1 is significantly better than each of the other models at the .005 confidence level. Like the within R-squareds shown in Table 5.9.a, the log-likelihoods and the likelihood-ratio test statistics in Table 5.9.b support the simultaneous inclusion of both the employment variables and the month dummies. This finding supports the hypothesis that these variables are complementary.

Although the likelihood-ratio tests again support Model 1 over the more parsimonious Models 2, 3 and 4, they also indicate that Model 2 and Model 3 are each superior to Model 4. Model 4, which includes neither month dummies nor employment variables, is a nested version of not only Model 1, but also of Model 2 which contains the month dummies and of Model 3 which contains the employment variables. The likelihood-ratio tests show the inclusion of either month dummies or employment variables significantly improves the fit of the model (see Table 5.9.b Panels 4 and 5) relative to Model 4. However, again, Model 1 is still preferred overall.

Table 5.9.b: Comparison of Models – Log-Likelihoods and Likelihood-Ratio Tests

	ARU	ARFG	TRU	TRFG
<i>Model 1-Model 2</i>				
Log-Likelihood (Model 1)	38724.45	33936.64	13161.16	19523.72
Log-Likelihood (Model 2)	38636.09	33914.08	13056.41	19484.76
Likelihood-Ratio Test Statistic	176.72	45.12	209.5	77.92
Difference in Degrees of Freedom	16	16	16	16
Significance Level	0.005	0.005	0.005	0.005
<i>Model 1-Model 3</i>				
Log-Likelihood (Model 1)	38724.45	33936.64	13161.16	19523.72
Log-Likelihood (Model 3)	38636.78	33825.10	13095.80	19488.76
Likelihood-Ratio Test Statistic	175.34	223.08	130.72	69.92
Difference in Degrees of Freedom	11	11	11	11
Significance Level	0.005	0.005	0.005	0.005
<i>Model 1-Model 4</i>				
Log-Likelihood (Model 1)	38724.45	33936.64	13161.16	19523.72
Log-Likelihood (Model 4)	38505.38	33769.71	12931.68	19443.59
Likelihood-Ratio Test Statistic	438.14	333.86	458.96	160.26
Difference in Degrees of Freedom	27	27	27	27
Significance Level	0.005	0.005	0.005	0.005
<i>Model 2-Model 4</i>				
Log-Likelihood (Model 2)	38636.1	33914.1	13056.4	19484.8
Log-Likelihood (Model 4)	38505.4	33769.7	12931.7	19443.6
Likelihood-Ratio Test Statistic	261.42	288.74	249.46	82.34
Difference in Degrees of Freedom	11	11	11	11
Significance Level	0.005	0.005	0.005	0.005
<i>Model 3-Model 4</i>				
Log-Likelihood (Model 3)	38636.8	33825.1	13095.8	19488.8
Log-Likelihood (Model 4)	38505.4	33769.7	12931.7	19443.6
Likelihood-Ratio Test Statistic	262.8	110.78	328.24	90.34
Difference in Degrees of Freedom	16	16	16	16
Significance Level	0.005	0.005	0.005	0.005

In the above comparisons, the log-likelihoods were consistently higher for Model 1, the model with the greatest number of explanatory variables. This finding is not surprising given that including additional explanatory variables can *never* reduce the explanatory power of the model. It is possible, of course, that the likelihoods might not be sufficiently larger to be statistically significant, but in this case all the likelihood ratios turned out to be statistically significant. Still, with a large enough sample size most additional explanatory variables will be statistically significant even if their substantive impact is small. The real question is whether adding more variables improves the fit enough to justify the added complexity of the models. For this reason, modellers who seek parsimony often propose stricter tests that provide penalties for additional variables.

Like the R-squared, the log-likelihood does not penalize for overfitting the model by including too many explanatory variables. As a result, relying on the log-likelihood alone in model selection may encourage the use of less parsimonious models. However, there are alternative measures of goodness-of-fit that do penalize for overfitting. For OLS, such a measure was the adjusted R-squared. For maximum likelihood, one alternative is the Akaike Information Criterion (AIC), which adjusts (i.e. penalizes) the log-likelihood by the number of coefficients estimated. The AIC is calculated as:

$$AIC_m = (-2 * L_m) + (2 * k_m)$$

where L is the log-likelihood value produced by the estimation, k is the number of coefficients, and the subscript m indicates the model number. Model selection is based on *minimization* of the AIC.

The log-likelihood, number of parameters, and AIC for each model can be found in Table 5.9.c. Our preferred model, Model 1, minimizes the Akaike Information Criterion despite the penalty for the additional coefficients estimated. Like the R-squared, log-likelihood, and likelihood-ratio test, the AIC supports the inclusion of both the month dummies and the employment variables. Model 2 and 3 are each preferred relative to Model 4, while Model 1 is preferred relative to Models 2, 3 and 4.

Our preferred model performs best regardless of the goodness-of-fit measure used to select among models. In addition, for all four outcomes, including the employment variables in Model 1 does not reduce the significance of the coefficients on the month dummies relative to Model 2 where only month dummies were included. Likewise, including the month dummies does not reduce the significance of the coefficients on the employment variables relative to Model 2 where only employment variables were included.

As a result, the preferred model chosen for Chapters 5 and 6 is Model 1, which contains both the employment variables and the month dummies.

Table 5.9.c: Comparison of Models – Akaike Information Criterion

	ARU	ARFG	TRU	TRFG
Model 1: Employment & Month				
Log-Likelihood	38724.45	33936.64	13161.16	19523.72
Number of parameters	102	102	102	102
AIC	-77244.90	-67669.28	-26118.32	-38843.44
Model 2: Month Dummies Only				
Log-Likelihood	38636.09	33914.08	13056.41	19484.76
Number of parameters	86	86	86	86
AIC	-77100.18	-67656.16	-25940.82	-38797.52
Model 3: Employment Vars Only				
Log-Likelihood	38636.78	33825.10	13095.80	19488.76
Number of parameters	91	91	91	91
AIC	-77091.56	-67468.20	-26009.60	-38795.52
Model 4: Neither				
Log-Likelihood	38505.38	33769.71	12931.68	19443.59
Number of parameters	75	75	75	75
AIC	-76860.76	-67389.42	-25713.36	-38737.18

5.4.B Employment Variables – Multicollinearity and Aggregation

In the above time-series and panel data models, multicollinearity among the employment variables was a serious concern. In order to address the issue of a cumulative impact of employment on the dependent variables, we now look at measures of the aggregate impacts of the employment variables.

It is interesting to note that when the employment coefficients (γ) are aggregated, the signs are more often correct. There are three aggregations that can be performed. The first method is to aggregate the results across both sectors and lags (time) resulting in one aggregate measure. The second method is to aggregate within sectors but across time resulting in four aggregate measures – one for each sector. The third method is to aggregate within lags but across sectors resulting again in four aggregate measures – one for each lag.

Method 1: Total Employment Effect: $\sum_{i=1}^4 \sum_{l=0}^3 \gamma_{i,t-l}$

Method 2: Sector Employment Effects: $\sum_{l=0}^3 \gamma_{i,t-l}$ for each sector $i = 1$ through 4

Method 3: Lag Employment Effects: $\sum_{i=1}^4 \gamma_{i,t-l}$ for each lag $l = 0$ through 3

For the multivariate time-series models described in Section 5.2, Table 5.10.a shows the results for the three aggregation methods. Recall that the multivariate time-series models estimate the accession and termination rates with current and lagged employment, a treatment variable, demographics, minimum wage and month and year dummies. The table shows the percent of California counties for which the aggregated coefficient had the expected sign.

For Method 1, which aggregates across all sixteen employment variables, more than half of the counties had the expected sign on the total employment coefficient. Using the second method,

we find that generally a majority of the counties had the proper sign for the aggregated coefficients for each of the employment sectors. The exceptions are the farm sector for three of the dependent variables and the retail sector for one dependent variable. Looking at the aggregated coefficients on the lags of employment, the aggregated coefficient for the current employment variable had the proper sign for all counties in the AFDC-U and AFDC-FG accession rates and for the AFDC-FG termination rates. The aggregated coefficient for the first lag likewise always had the expected sign for three of the four dependent variables. However, the aggregated coefficient on the second lag never had the expected sign for any counties, while the aggregated coefficient on the third lag always had the proper sign.

**Table 5.10.a: Time-Series Models – Aggregated Employment Coefficients
(Percent of Counties with Correct Sign)**

	ARU	ARFG	TRU	TRFG
Method 1				
Total Employment	58.93	50.00	53.57	51.79
Method 2				
Farm	51.79	48.21	48.21	48.21
Retail	46.43	51.79	51.79	53.57
Services	58.93	50.00	66.07	48.21
Other	57.14	64.29	58.93	60.71
Method 3				
Current (t)	100	100	0	100
Lag 1 (t-1)	100	0	100	100
Lag 2 (t-2)	0	0	0	0
Lag 3 (t-3)	100	100	100	100

The same aggregations were also performed for both the homoskedastic and heteroskedastic panel data models. Recall the panel data models are the ones from Section 5.3.B (Homoskedastic) and Section 5.3.C (Heteroskedastic) whose estimations include fixed effects and the county type/summer months interaction terms. The results are found in Table 5.10.b. Since each panel data model stacks all the county data, there is only one coefficient to be signed after aggregation. As a result, we report below whether the sign of the resulting aggregated coefficient was correct.

When the coefficients are aggregated across all 16 employment variables (i.e., across both sectors and lags), the aggregated coefficient has the proper sign for all models except the AFDC-FG accession rate in both the homoskedastic and heteroskedastic specifications. This finding is not surprising given this group's low attachment to the labor force.

Using the second method, we have four aggregated employment coefficients – one for each sector. In this case, aggregated coefficients for three of the four sectors have the expected sign for three of the outcome variables; the exception again is the AFDC-FG accession rate model where the aggregated coefficients for only two sectors have the expected sign in the homoskedastic specification and for only one sector in the heteroskedastic specification.

With the third method, we again have four aggregated employment coefficients – one for the current employment variables and one for each of the three lags. The aggregate coefficients for the current employment variables summed across sectors have the expected sign for all models. All models also have at least one lag with the expected sign on the aggregated coefficient except for the homoskedastic specification of the AFDC-FG accession rate model. For models with correctly signed lags, the correct sign occurred only for the first and/or third lags.

Overall, the signs on the employment coefficients do improve when the coefficients on employment variables are aggregated using any of the three methods discussed above.¹⁶

Table 5.10.b: Panel Data Models - Aggregated Employment Coefficients

	ARU	ARFG	TRU	TRFG
Aggregated Coefficients	Homoskedastic			
<i>Method 1</i>				
Total Employment	Correct		Correct	Correct
<i>Method 2</i>				
Farm	Correct		Correct	Correct
Retail			Correct	Correct
Services	Correct	Correct	Correct	
Other	Correct	Correct		Correct
<i>Method 3</i>				
Current (t)	Correct	Correct	Correct	Correct
Lag 1 (t-1)	Correct		Correct	
Lag 2 (t-2)				
Lag 3 (t-3)				Correct
Aggregated Coefficients	Heteroskedastic			
<i>Method 1</i>				
Total Employment	Correct		Correct	Correct
<i>Method 2</i>				
Farm	Correct		Correct	Correct
Retail			Correct	Correct
Services	Correct	Correct	Correct	
Other	Correct			Correct
<i>Method 3</i>				
Current (t)	Correct	Correct	Correct	Correct
Lag 1 (t-1)	Correct	Correct	Correct	
Lag 2 (t-2)				
Lag 3 (t-3)			Correct	Correct

¹⁶ The aggregations are also calculated for both the homoskedastic and heteroskedastic models estimated using varying estimation lengths (which are estimated for forecasting in Chapter 6). The results of the aggregation can be found in Appendix Table A9.a and A9.b for the homoskedastic and heteroskedastic models, respectively.

5.4.C Employment Variables – Choosing the Number of Lags of Employment

In Chapter 5, all models included three lags of each employment variable. It is, of course, possible to include additional lags or fewer lags.

The argument for including any lags of employment in modeling accession and termination rates is that changes in employment may not have an immediate impact. For example, we may find that increases in employment indicate an improvement in the economy, but that welfare participants are not among the first group to return to employment. Likewise, we may find that decreases in employment indicate a worsening of the economy, but that recently displaced workers do not immediately enter welfare. Rather, they may first be required to draw down their past savings before become eligible or they may first search for alternate employment opportunities.

For these reasons, we would like to include lags of the employment variables in the models. However, there is still the issue of how many lags to include. To address this issue, we can again employ the Akaike Information Criterion (AIC) used in Section 5.4.A to compare the performance of models which include a different number of lags.

Panel data models for each of the four outcome variables were estimated with the models from Section 5.3.B that include fixed effects and county type summer month interaction terms but also with varying numbers of lags of employment variables. It is important to note that with our sample, increasing the number of lags also caused the sample size to decrease. Our full sample model with three lags started in October 1985 because three months of data (July, August and September) were lost to lags.

We now estimated the model two ways. First, we estimate the model with zero to six lags and allow the lags to determine the sample size. This method means that the model with fewer lags will have more observations than models with a greater number of lags. Then, we re-estimate the models with two, three, four, five and six lags determining the sample size, respectively, and then compare each iteration to models with fewer lags, but with a constant sample.

Table 5.11.a shows the results from the model where the sample is left unrestricted. From this table, it is clear that models with no lags (i.e. with only the current level of employment) fit best based on the Akaike Information Criterion, followed by models with one lag, two lags, etc. However, this result is a function not only of the additional variables, but also of the number of observations. With an unrestricted sample, the model with no lags has the greatest number of observations. In the panel analysis, addition of one lag of employment reduces the number of observations by 56 – the number of included counties.¹⁷

¹⁷ Actually the reduction will be less than or equal to 56 depending on how many counties have data for that period.

Table 5.11.a: Comparing Lags in an Unrestricted Sample

	ARU	ARFG	TRU	TRFG
Akaike Information Criterion				
Model: 0 Lags	-79057.1	-69346.3	-26573.3	-39671.8
Model: 1 Lags	-78411	-68784.6	-26358.6	-39419.5
Model: 2 Lags	-77845.4	-68217.8	-26270.5	-39138.2
Model: 3 Lags	-77244.9	-67669.3	-26118.3	-38843.4
Model: 4 Lags	-76622.5	-67111.4	-25998	-38571.2
Model: 5 Lags	-76011.7	-66572.8	-25891.7	-38326.1
Model: 6 Lags	-75437.6	-66022.2	-25784	-38016.4
Rank				
Model: 0 Lags	1	1	1	1
Model: 1 Lags	2	2	2	2
Model: 2 Lags	3	3	3	3
Model: 3 Lags	4	4	4	4
Model: 4 Lags	5	5	5	5
Model: 5 Lags	6	6	6	6
Model: 6 Lags	7	7	7	7

But what happens if we restrict the sample and compare the AIC of models with different numbers of lags, but using the same number of observations? For the results in Table 5.11.b, the sample for each regression was restricted so that the model with six lags was estimated with exactly the same number of observations as the model with no lags. The results in Table 5.11.b indicate that for three of the four outcome variables, a greater number of lags improves the fit of the model given a fixed (constant) sample. For the AFDC-U accession rate and both the AFDC-U and AFDC-FG termination rates, the models with the greatest number of lags now rank first among models in terms of fit. Models with five lags generally rank second and so on. The model with no lags now ranks among the worst performers for three of the four outcome variables.

The notable exception to this rule is the AFDC-FG accession rate where the model with no lags performs best. This result is not surprising given the results in the previous section. Recall that in Section 5.4.B, we aggregated the employment coefficients across sectors, but within time period. For the AFDC-FG accession rate, only the current employment level had the proper sign on the aggregated employment coefficient when long estimation periods were used (see Table 5.11.b and Appendix Table A9.a)¹⁸. The other outcome variables had at least one lag with the properly signed aggregated coefficient.

¹⁸ When shorter estimation periods are used, this result does not persist.

**Table 5.11.b: Comparing Lags in a Restricted Sample – Six Lags
(Sample Determined by Model with Six Lags)**

	ARU	ARFG	TRU	TRFG
Akaike Information Criterion				
Model: 0 Lags	-75319.7	-66036.2	-25685.1	-37954.3
Model: 1 Lags	-75314	-66032.6	-25679.5	-37963.8
Model: 2 Lags	-75354.8	-66030	-25708.3	-37977.5
Model: 3 Lags	-75393.6	-66026.7	-25739.9	-37978.5
Model: 4 Lags	-75412.2	-66024.2	-25782.6	-38014.2
Model: 5 Lags	-75428.7	-66022.9	-25777.1	-38015.4
Model: 6 Lags	-75437.6	-66022.2	-25784	-38016.4
Rank				
Model: 0 Lags	6	1	7	7
Model: 1 Lags	7	2	6	6
Model: 2 Lags	5	3	5	5
Model: 3 Lags	4	4	4	4
Model: 4 Lags	3	5	3	3
Model: 5 Lags	2	6	2	2
Model: 6 Lags	1	7	1	1

The same comparisons were performed with three, four, and five lags defining the sample instead of six lags. The results can be found in Table 5.11.c, 5.11.d and 5.11.e, respectively. The patterns in these three tables are similar to that found in Table 5.11.b. The greater number of lags produces a better fit if the sample is restricted.

**Table 5.11.c: Comparing Lags in a Restricted Sample – Five Lags
(Sample Determined by Model with Five Lags)**

	ARU	ARFG	TRU	TRFG
Akaike Information Criterion				
Model: 0 Lags	-75319.7	-66036.2	-25685.1	-37954.3
Model: 1 Lags	-75314	-66032.6	-25679.5	-37963.8
Model: 2 Lags	-75354.8	-66030	-25708.3	-37977.5
Model: 3 Lags	-75393.6	-66026.7	-25739.9	-37978.5
Model: 4 Lags	-75412.2	-66024.2	-25782.6	-38014.2
Model: 5 Lags	-75428.7	-66022.9	-25777.1	-38015.4
Rank				
Model: 0 Lags	5	1	5	6
Model: 1 Lags	6	2	6	5
Model: 2 Lags	4	3	4	3
Model: 3 Lags	3	4	3	4
Model: 4 Lags	2	5	1	2
Model: 5 Lags	1	6	2	1

**Table 5.11.d: Comparing Lags in a Restricted Sample – Four Lags
(Sample Determined by Model with Four Lags)**

	ARU	ARFG	TRU	TRFG
Akaike Information Criterion				
Model: 0 Lags	-75898.6	-66585.6	-25790.1	-38265
Model: 1 Lags	-75892.6	-66582.1	-25785.2	-38274.7
Model: 2 Lags	-75937	-66579.6	-25817.4	-38288.8
Model: 3 Lags	-75976	-66576.2	-25854.1	-38288.3
Model: 4 Lags	-75995.2	-66573.9	-25897.2	-38324.2
Rank				
Model: 0 Lags	4	1	4	5
Model: 1 Lags	5	2	5	4
Model: 2 Lags	3	3	3	3
Model: 3 Lags	2	4	2	2
Model: 4 Lags	1	5	1	1

**Table 5.11.e: Comparing Lags in a Restricted Sample - Three Lags
(Sample Determined by Model with Three Lags)**

	ARU	ARFG	TRU	TRFG
Akaike Information Criterion				
Model: 0 Lags	-76525.4	-67121.9	-25890.1	-38511.5
Model: 1 Lags	-76519.3	-67118.7	-25885.2	-38523.6
Model: 2 Lags	-76564.4	-67116.1	-25909	-38533.6
Model: 3 Lags	-76604.3	-67113	-25946.1	-38534
Rank				
Model: 0 Lags	3	1	4	4
Model: 1 Lags	4	2	3	3
Model: 2 Lags	2	3	2	2
Model: 3 Lags	1	4	1	1

The results in this section indicate that there is a trade-off between including more lags and decreasing the sample size. For our preferred model, we considered both factors and selected the model with three lags. This choice was based partly on theoretical concerns. For accessions, we hypothesized that those who were recently unemployed might not enter welfare directly as they might draw first upon savings or spend some time looking for alternate employment. Likewise, for terminations, we also hypothesized that those on welfare might not be the first beneficiaries of an upswing in economic activity. As a result, we felt that some lags were necessary. However, recognizing that there is a trade-off between lags and sample size, we chose not to include a large number of lags.

5.5 Summary

We began this chapter with correlational and bivariate regression analyses performed within and across counties. In the results within and across counties, both the correlational and bivariate analyses produced coefficients that generally had the wrong sign for the treatment variables. Since these data are not from a random experiment, it is not surprising that a bivariate regression yielded strange results. The treatment was not randomly assigned nor was the treatment the only potential contributing factor which changed during this period. Other factors such as economic conditions also played a role in determining accession and termination rates.

Using the methodologies and techniques outlined in Chapter 3, we then built more sophisticated multivariate models which included employment and demographic variables as well as county and time-specific characteristics. The use of multivariate and panel data techniques allow us to estimate the impact of changing benefit levels in non-experimental situations where other factors may be changing simultaneously.

However, not all model-building involves increasing the number of variables or observations in the estimation. It is important to note that the usefulness of OLS and other regression techniques, such as panel data techniques, rely on certain assumptions. Violation of these assumptions can adversely affect inference. An example of such a violation is heteroskedasticity. In looking at the variance across counties in the panel data, we found that the assumption of constant variance was violated. As a result, it was necessary to adapt the model to avoid incorrect inference. A heteroskedasticity correction such as the one performed in Section 5.3.C improved the model's accuracy by correcting the standard errors for more complicated variance structures.

Further issues of model selection regarding choice of employment variables versus month dummies, multicollinearity of employment variables, and determination of the number of lags were addressed in Section 5.4. This section confirmed the inclusion of both employment variables and month dummies and supported the inclusion of a limited number of lags of employment variables.

With these extended multivariate models we find that both within a single county (i.e., time-series) and across multiple counties (i.e., panel), we get much more reliable results. In the majority of cases, the treatment variables had the expected sign. Most of the additional explanatory variables also had the expected sign except in cases where multicollinearity was severe.

To summarize, our best model in terms of what we believe to be reasonable is the panel heteroskedastic model that estimates accession and termination rates with county fixed effects on continuous benefits, current plus three lags of employment, minimum wage, county type summer month interaction terms, percent Hispanic, percent black, birth rate, and month and year dummies. The results of this estimation are shown in Table 5.8. Our preference for this model is based on our belief that it most accurately addresses all our estimation concerns. However, we also find that generally the resulting coefficients are not too far from what would be expected.

Chapter 6 Unconditional Out-of-Sample Forecasting

In Chapter 3, we discussed using models to make predictions or forecasts about future values of the outcome variables. In Chapter 5, we developed several models to predict values of the outcome variables. We will now use the multivariate time-series and multivariate panel data models we developed in the previous chapter to forecast values for accession and termination rates.

In this chapter, we will focus on unconditional out-of-sample forecasts. Out-of-sample forecasting indicates that we are predicting values for observations which are not included in the estimation sample. Unconditional refers to the fact that we are not estimating the regressors of the model. That is, actual observations for our explanatory variables are available for the forecasted periods.

Forecasts are useful for planning. We may want to know the future value of accession or termination rates in order to plan for the necessary budget levels or case manager manpower. Forecasts also allow us to evaluate the model on an additional criterion. We used adjusted R-squared to evaluate the model fit during the estimation sample. But unconditional out-of-sample forecasts allow us to evaluate how well the model performs in predicting after the end of the estimation period. Since we have the actual values for the outcomes as well, we can evaluate the forecast accuracy.

One measure of forecast accuracy is the average mean squared forecasting error (AMSFE).

$$\text{AMSFE} = \frac{\sum_{t=1}^T (Y_t^* - Y_t)^2}{T} = \frac{\sum_{t=1}^T (\varepsilon_t)^2}{T}$$

Thus if the forecast Y_t always equals the observed value Y_t^* , then ε_t for every observation is zero, and AMSFE is zero. The AMSFE is calculated as the average of the sum of the squared errors (ε_t) over the number of forecast periods (T).

6.1 Time Series Forecasting

Using the multivariate time-series regressions for individual counties, we forecasted two years of accession and termination rates. The models originally shown in Section 5.2 will now be estimated from October 1985 to December 1996 and used to forecast January 1997 to December 1998.

The AMSFE is calculated over the 24 months in the forecast period. Table 6.1 shows summary statistics for the AMSFE for each of the four outcome variables. These statistics are the minimum, maximum, mean and median of the AMSFEs across the 56 county-specific forecasts. The individual county AMSFEs can be found in Tables A1.a through A8.a.

Table 6.1: Summary Statistics for AMSFE Level across Counties

	ARU	ARFG	TRU	TRFG
Dichotomous Treatment				
Minimum	7.66E-10	8.14E-09	9.66E-06	1.19E-06
Mean	2.29E-05	1.24E-04	8.89E-03	4.86E-03
Median	9.64E-07	6.18E-06	2.45E-03	4.78E-04
Maximum	7.37E-04	2.23E-03	6.60E-02	1.58E-01
Continuous Treatment				
Minimum	1.81E-09	1.68E-08	5.07E-06	1.54E-06
Mean	2.47E-05	9.86E-05	1.13E-02	5.38E-03
Median	1.71E-06	6.49E-06	4.25E-03	4.89E-04
Maximum	7.25E-04	2.11E-03	9.88E-02	1.49E-01

Note that while the AMSFE seem much smaller for accession rates than for termination rates, this is simply an artifact of the data. The AMSFE are smaller for accession rates, but so are the actual levels of accession rates. The mean (median) accession rate for the forecast period is 0.14% (0.13%) for AFDC-U and 0.6% (0.6%) for AFDC-FG, while the mean (median) termination rate is 7.8% (8.0%) for AFDC-U and 6.5% (6.6%) for AFDC-FG.

We can translate these AMSFE statistics into a percent error from the actual level of outcomes by dividing the mean (median) AMSFE over all counties by the mean (median) actual accession or termination rate¹. This measure indicates that the model prediction misses the true accession rate at the mean (median) by: 9.6% (0.8%) for the AFDC-U accession rate, 12.3% (0.9%) for the AFDC-FG accession rate, 97.3% (39.7%) for the AFDC-U termination rate, and 71.1% (7.4%) for the AFDC-FG termination rate.

**Table 6.2: Forecast Percent Deviation from Actual Value
(Mean/Median AMSFE as a Percent of Mean/Median Actual Value)**

	ARU	ARFG	TRU	TRFG
Continuous Treatment Regressions				
<i>AMSFE as a Percent of</i>				
Mean Actual Value	9.58%	12.28%	97.28%	71.06%
Median Actual Value	0.81%	0.91%	39.71%	7.38%

6.2 Panel Data Forecasting

Using the full panel data model both with and without the heteroskedasticity correction, we now forecast the accession and termination rates. As with the time-series, the estimation period begins in October 1985 and ends in December 1996, while the forecasting period begins in

¹ While we refer to this as a percent error, it should be noted that the numerator of this value is actually a squared term while the denominator is not. Section 6.6 in the calculation of minimum required change considers another normalization of the AMSFE that does not have this characteristic.

January 1997 and continues through December 1998. The forecasts for panel data are constructed for the continuous treatment only.

Using the multivariate time-series regressions for individual counties, we forecasted two years of accession and termination rates. Recall from Section 5.2 the model that we estimate includes current and lagged independent values for the economic factors (X^{1*}), but only current values for the treatment (D^*), demographics (X^{2*}), California minimum wage (X^{3*}), and the month (X^{4*}) and year (X^{5*}) dummies.

The inclusion of year dummies presents an interesting dilemma. Coefficients on the year dummies are estimated with year indicator variables only for data in the estimation period. Yet these dummy variables clearly matter for the fit. Should we let their effects be zero for the forecast period or should we try to include them in some way?

There are several approaches we could follow to deal with this situation. The first would be to exclude the dummy variables from the estimation and then forecast this new version of the model. Recall however that the year dummies were included in the original estimation because they represented unobserved trends. If the coefficients on the year dummy variables are large and significant, as we find in our analysis of Chapter 5, then this would lead to a much poorer fit in the original estimation but perhaps a smaller forecast error.

A second approach would be to maintain the same model estimated in Chapter 5 that includes the year dummies and ignore any contribution from the coefficients of the year dummy variables to the forecast. This leads to be a better fit during the estimation period but larger forecast errors.

A third approach would be to include a proxy for the contribution of the year dummy variable in the forecast. The exact specification of this approach can take many forms. One possible specification would be to proxy the contribution by including in the forecast the effect of a polynomial of the effects of the year dummy variables included in the estimation period. In many ways this is similar to the polynomial distributed lag or geometric lag methodology described in Section 3.2. Using a lag structure of either form however requires a functional form for weighting the lags. Since our yearly dummies are meant to represent unobserved trends it is unclear what weights would be appropriate. As such, we have decided to include a simpler functional form. We include in the forecast the contribution of the coefficient of the most recent year dummy variable in the estimation as the contribution of the missing year dummy variables. This is exactly identical to estimating with a lag structure where we place all the weight on the most recent lag. Since we have no good assumptions about what weighting scheme would be best, this straight-forward approach seems reasonable and is computationally easy.²

Table 6.3 shows the AMSFE for the homoskedastic and heteroskedastic forecasts for each of the three alternative approaches for the year dummy variable issue described earlier. The AMSFEs are again quite small. Even in the case of the original model, where the AMSFEs are larger than the other two alternatives, the numbers below translate to errors of only 1.2 to 2.7% of the actual

² We could also have taken each of these three approaches in the time-series regressions described in Section 6.1. However, we generally do not find significant coefficients on the year dummy variables in the individual county regressions. As a result, we ignored the impact of the year dummy variables.

values for the two-year forecasts. The heteroskedastic model performs better in terms of lower AMSFEs for the accession rates while the homoskedastic model performs better for the termination rates.

Table 6.3: Homoskedastic and Heteroskedastic Panel Forecasts - Three Alternatives

Panel Forecasts - No Year Dummies				
	ARU	ARFG	TRU	TRFG
Homoskedastic				
AMSFE 1997-1998	2.25E-06	1.73E-05	7.86E-03	1.51E-03
AMSFE 1997 Only	2.10E-06	1.41E-05	5.59E-03	1.10E-03
AMSFE 1998 Only	2.41E-06	2.06E-05	1.02E-02	1.93E-03
Heteroskedastic				
AMSFE 1997-1998	2.17E-06	1.66E-05	7.71E-03	1.52E-03
AMSFE 1997 Only	2.13E-06	1.37E-05	5.36E-03	1.10E-03
AMSFE 1998 Only	2.21E-06	1.96E-05	1.01E-02	1.94E-03
Panel Forecasts - Original Model				
	ARU	ARFG	TRU	TRFG
Homoskedastic				
AMSFE 1997-1998	1.33E-05	9.43E-05	2.55E-02	6.36E-03
AMSFE 1997 Only	1.59E-05	1.03E-04	2.66E-02	6.46E-03
AMSFE 1998 Only	1.07E-05	8.61E-05	2.44E-02	6.26E-03
Heteroskedastic				
AMSFE 1997-1998	1.29E-05	9.40E-05	2.63E-02	6.52E-03
AMSFE 1997 Only	1.59E-05	1.03E-04	2.74E-02	6.61E-03
AMSFE 1998 Only	9.95E-06	8.51E-05	2.51E-02	6.43E-03
Panel Forecasts - Estimated Year Dummies				
	ARU	ARFG	TRU	TRFG
Homoskedastic				
AMSFE 1997-1998	2.26E-06	1.68E-05	5.05E-03	1.21E-03
AMSFE 1997 Only	2.11E-06	1.38E-05	3.54E-03	8.85E-04
AMSFE 1998 Only	2.41E-06	1.98E-05	6.58E-03	1.54E-03
Heteroskedastic				
AMSFE 1997-1998	2.17E-06	1.67E-05	5.20E-03	1.24E-03
AMSFE 1997 Only	2.11E-06	1.39E-05	3.64E-03	9.05E-04
AMSFE 1998 Only	2.23E-06	1.95E-05	6.77E-03	1.58E-03

The forecast period is comprised of 24 months. Given the number of months over which the forecast is constructed, it is possible to determine whether there is any decay or improvement in the accuracy of the forecast over time. Table 6.3 shows the AMSFE for the split forecast sample so that the AMSFE is calculated for the first and second year of the forecast separately. The AMSFE for 1997, the first year of the forecast, is larger than the AMSFE for 1998. This result seems strange given that we would expect the first year of the forecast to be more similar to the estimation period and hence have a more accurate forecast. However, recall that regression

explains the variation in the outcome variable. This variation is not constant across time periods. For all four outcome variables, the percent change during 1997 was greater than during 1998. The fact that 1997 contained more variation than 1998 is true whether we look at variation across counties or at the statewide variation. Also, looking at a standard month, November of each year, graphs of the variance of each outcome show that 1997 had a higher variance for the AFDC-U accession rate, the AFDC-FG accession rate, and the AFDC-FG termination rate (see Figures A3.a through A3.d in the Appendix). As a result, it is not unexpected that the forecasts for 1998, a more stable year, perform better.

The remainder of this chapter considers forecasting the panel model only and not the time-series model described in the earlier section. While it would be possible to replicate all the results in the following sections for all three panel model alternatives, we will only proceed with the last alternative where we include an estimated effect of the year dummy variables during the forecast period. We find that regressions that exclude the year dummies entirely produce strange in-sample estimations and therefore cannot be counted on to produce good out-of-sample forecasts. Running the original model from Chapter 5 eliminates some of the perverse findings that we get when we exclude the year dummies but also has much larger AMSFEs because there are no estimated coefficients on the year dummies in the forecast calculations. We find that the third alternative where we estimate the value of the coefficient(s) on the year dummy variable(s) to be the coefficient on the last in-sample year dummy to be the best solution. While many could argue that the very best solution would be to model exactly the excluded explanatory variable the year dummy variables are representing, we cannot identify what is missing from our model. Others would argue that a more complex weighting scheme of prior dummy variables would be more appropriate, but again without knowing what is being represented by the dummy variables it is hard to know exactly what form the weights should take.

We find that using this third approach that estimates the effect of the year dummy variables in the forecast leads to much smaller AMSFEs as well as estimated coefficients of the expected sign. Generally when we add our estimate of the effect of the yearly dummy variables we find the AMSFE is 5 to 6 times smaller than if we just ran our original model that left out any contribution. This difference is extremely important especially later in this chapter when we consider how large a change would have to be before analysis could measure any impact. The remainder of this chapter will follow this third approach without further discussion of the other alternatives.

6.3 Varying the Estimation Sample: The Effect on Forecasts

Changing the estimation sample can have a critical impact on the accuracy of the forecasts. To illustrate this, we re-estimated the panel model with varying start dates for the estimation period. The first estimation period begins in January 1986 and is estimated for eleven years through December 1996. We then re-estimate the model starting in January 1987 for ten years. We then continue to re-estimate the model while shortening the estimation period by one year. The truncation of sample always occurs at the beginning of the estimation sample period. For each iteration the model always forecasts the same 24 months from January 1997 through December 1998. Table 6.4.a shows the AMSFE for each of the forecasts.

Homoskedastic results vary across the four outcome variables. Forecasts of accession rates for 1997 through 1998 for AFDC-U cases are best predicted by very short and very long estimation periods. The one-year (1996) estimation period works well because 1996 has decreasing accessions similar to the forecast period 1997-1998. Estimation lengths of three through seven years fail to forecast well because the accession rate is mostly flat or increasing during the estimation period. However, eight to eleven year estimation periods perform well because they include enough data and sufficient variation in accession rate movements. There is not much variation in forecast accuracy for accession rates for AFDC-FG because movements in the overall accession rates were less dramatic. The termination rates for AFDC-U are best forecasted by long and short estimation periods. The worst performers were the five through seven year estimation periods, which were comprised mostly of flat termination rates. As a result, these three forecasts were unable to forecast the upward trend of 1997-98. The results for the AFDC-FG termination rate are similar.

There is less variation in forecast accuracy across the estimation periods when the model corrects for county-level heteroskedasticity. The overall pattern, however, is similar to the homoskedastic results.

Table 6.4.a: AMSFE by Beginning of Estimation Period: Forecast 1997-98 (AMSFE Level)

		ARU	ARFG	TRU	TRFG
Year	Length	Homoskedastic			
1986	11	2.41E-05	1.29E-04	2.15E-02	5.60E-03
1987	10	2.74E-05	1.08E-04	2.10E-02	4.84E-03
1988	9	1.37E-05	9.45E-05	2.62E-02	4.75E-03
1989	8	2.82E-05	9.44E-05	2.78E-02	4.93E-03
1990	7	8.01E-05	9.57E-05	6.21E-02	5.05E-03
1991	6	5.94E-05	9.83E-05	6.10E-02	5.21E-03
1992	5	4.96E-05	9.17E-05	3.83E-02	5.19E-03
1993	4	8.96E-05	1.02E-04	1.96E-02	8.76E-03
1994	3	7.66E-05	1.01E-04	2.48E-02	7.32E-03
1995	2	6.14E-05	1.06E-04	1.94E-02	9.79E-03
1996	1	1.31E-05	9.01E-05	3.11E-02	8.09E-03
Year	Length	Heteroskedastic			
1986	11	1.34E-05	1.14E-04	2.28E-02	5.43E-03
1987	10	1.32E-05	9.95E-05	2.25E-02	4.77E-03
1988	9	1.28E-05	9.40E-05	2.66E-02	4.67E-03
1989	8	1.38E-05	9.35E-05	2.24E-02	5.13E-03
1990	7	1.73E-05	9.20E-05	4.39E-02	4.94E-03
1991	6	1.36E-05	9.23E-05	4.23E-02	5.08E-03
1992	5	1.23E-05	9.22E-05	2.70E-02	4.41E-03
1993	4	1.39E-05	9.19E-05	2.00E-02	5.03E-03
1994	3	1.32E-05	9.40E-05	2.08E-02	4.59E-03
1995	2	1.34E-05	9.22E-05	1.81E-02	5.64E-03
1996	1	1.20E-05	8.37E-05	2.08E-02	5.37E-03

Table 6.4.a above shows the level of the AMSFE for each of the models. However, it is difficult to compare model performance when looking only at the level of the AMSFE. Tables 6.4.b, 6.4.c, and 6.4.d show the size of the AMSFE relative to the mean, median and standard deviation, respectively, of the dependent variables.

The first two tables below, Tables 6.4.b and 6.4.c, show the AMSFE as a percent of the level measured as the mean and median, respectively, of the dependent variable in order to normalize (or standardize) the forecast error by the dependent variable. These measures allow comparison of performance across the models. With the exception of the AFDC-U termination rate models, both the homoskedastic and heteroskedastic specifications have AMSFE averaging less than 2% of the mean and median of the dependent variable (see Tables 6.4.b and 6.4.c). For both the AFDC-U and AFDC-FG accession rates the AMSFE is less than 1% of the mean or median. The AMSFE for the AFDC-U termination rates are much higher with a range of 2 to 4% (2 to 5%) of the mean (median) in both the homoskedastic and heteroskedastic specifications.

A third normalized measure is the AMSFE as a percent of the standard deviation in the dependent variable (see Table 6.4.d). This measure normalizes the forecast error by the amount of variation in the dependent variable to give an indication of the relative volatility of the forecast versus the actual level. For both the AFDC-U and AFDC-FG accession rates, the forecast error is quite low – never reaching 1% of the standard deviation. However, the termination rate models perform less satisfactorily with the measure ranging from 2 to 6% for the AFDC-FG termination rate and 4 to 7% for the AFDC-U termination rate.

All three normalizations show that when the AMSFE is normalized by the mean, median or standard deviation of the dependent variables it is actually quite low.

**Table 6.4.b: AMSFE by Beginning of Estimation Period: Forecast 1997-98
(As a Percent of the Mean of Dependent Variable)**

		ARU	ARFG	TRU	TRFG
Year	Length	Homoskedastic			
1986	11	0.09 %	0.20 %	3.53 %	1.23 %
1987	10	0.09	0.20	3.66	1.15
1988	9	0.09	0.20	3.64	1.11
1989	8	0.11	0.20	2.73	0.93
1990	7	0.16	0.21	3.23	0.95
1991	6	0.15	0.22	3.53	0.99
1992	5	0.15	0.20	3.29	0.97
1993	4	0.23	0.22	2.71	1.14
1994	3	0.25	0.22	2.79	1.11
1995	2	0.43	0.23	2.65	1.76
1996	1	0.09	0.20	3.77	1.34
Year	Length	Heteroskedastic			
1986	11	0.08 %	0.20 %	3.63 %	1.26 %
1987	10	0.08	0.20	3.63	1.17
1988	9	0.08	0.20	3.69	1.13
1989	8	0.09	0.20	2.85	0.97
1990	7	0.09	0.20	3.33	0.93
1991	6	0.09	0.20	3.66	0.97
1992	5	0.08	0.20	3.37	0.93
1993	4	0.09	0.20	2.89	0.99
1994	3	0.09	0.21	2.57	0.92
1995	2	0.09	0.21	2.51	1.07
1996	1	0.08	0.19	2.70	0.94

**Table 6.4.c: AMSFE by Beginning of Estimation Period: Forecast 1997-98
(As a Percent of the Median of Dependent Variable)**

		ARU	ARFG	TRU	TRFG
Year	Length	Homoskedastic			
1986	11	0.11 %	0.23 %	4.04 %	1.25 %
1987	10	0.11	0.23	4.03	1.24
1988	9	0.12	0.23	4.18	1.16
1989	8	0.12	0.23	4.16	1.12
1990	7	0.15	0.24	3.12	0.94
1991	6	0.21	0.24	3.69	0.96
1992	5	0.19	0.25	4.04	1.00
1993	4	0.19	0.24	3.76	0.98
1994	3	0.30	0.25	3.09	1.16
1995	2	0.33	0.25	3.19	1.13
1996	1	0.56	0.27	3.02	1.78
Year	Length	Heteroskedastic			
1986	11	0.11 %	0.23 %	4.15 %	1.27 %
1987	10	0.11	0.23	4.15	1.18
1988	9	0.11	0.23	4.22	1.14
1989	8	0.12	0.23	3.25	0.98
1990	7	0.12	0.23	3.80	0.94
1991	6	0.12	0.23	4.18	0.98
1992	5	0.11	0.23	3.85	0.94
1993	4	0.12	0.24	3.30	1.00
1994	3	0.12	0.24	2.94	0.92
1995	2	0.12	0.24	2.86	1.08
1996	1	0.11	0.22	3.08	0.95

**Table 6.4.d: AMSFE by Beginning of Estimation Period: Forecast 1997-98
(As a Percent of the Standard Deviation of Dependent Variable)**

		ARU	ARFG	TRU	TRFG
Year	Length	Homoskedastic			
1986	11	0.10 %	0.31 %	6.39 %	3.75 %
1987	10	0.10	0.31	6.38	3.73
1988	9	0.11	0.30	6.61	3.49
1989	8	0.11	0.30	6.59	3.38
1990	7	0.13	0.31	4.94	2.83
1991	6	0.19	0.32	5.84	2.90
1992	5	0.18	0.33	6.39	3.02
1993	4	0.18	0.31	5.95	2.96
1994	3	0.28	0.33	4.89	3.48
1995	2	0.30	0.33	5.04	3.39
1996	1	0.51	0.35	4.78	5.36
Year	Length	Heteroskedastic			
1986	11	0.10 %	0.30 %	6.56 %	3.83 %
1987	10	0.10	0.30	6.57	3.55
1988	9	0.10	0.30	6.67	3.43
1989	8	0.11	0.31	5.14	2.96
1990	7	0.11	0.30	6.02	2.84
1991	6	0.11	0.30	6.61	2.94
1992	5	0.10	0.30	6.10	2.82
1993	4	0.11	0.31	5.22	3.00
1994	3	0.11	0.32	4.65	2.78
1995	2	0.11	0.31	4.53	3.26
1996	1	0.10	0.29	4.88	2.86

6.4 Moving Forecasts and Estimation Sample

In the previous section, we found that short estimation periods sometimes work well in predicting 1997 through 1998. However, this result seems dependent on the stability of the rates during the estimation and forecast period. More clearly, if the two years prior to the forecast period are similar to the forecast period, then the two-year estimation period will produce an accurate forecast.

To test the usefulness of the two-year estimation period, we use a rolling two-year estimation period to predict the following two years. For example, we use data from January 1986 through December 1987 to forecast January 1988 through December 1989. Then, we use January 1987 through December 1988 to forecast January 1989 through December 1990, and so on.

The results in Table 6.5 and Figures 6.1.a through Figure 6.1.d show that the accuracy of these forecasts depends on the shape of the forecasted periods relative to the estimation periods. Two-year estimation periods fail to properly predict inflection or turning points in the rates.

Table 6.5: AMSFE for Rolling Two-Year Estimation and Forecast Periods

Estimation Begin	Estimation End	Forecast Begin	Forecast End	AMSFE ARU	AMSFE ARFG	AMSFE TRU	AMSFE TRFG
Homoskedastic							
8601	8712	8801	8912	4.27E-06	5.63E-06	1.00E-02	8.39E-04
8701	8812	8901	9012	1.80E-06	1.09E-05	1.18E-02	3.02E-04
8801	8912	9001	9112	1.23E-06	6.94E-06	1.97E-03	2.53E-04
8901	9012	9101	9212	1.71E-06	1.97E-05	3.84E-03	3.27E-04
9001	9112	9201	9312	1.42E-06	3.10E-06	1.65E-03	4.01E-04
9101	9212	9301	9412	1.32E-06	1.29E-05	4.82E-03	4.05E-04
9201	9312	9401	9512	1.62E-06	7.41E-06	4.14E-03	3.80E-04
9301	9412	9501	9612	2.89E-06	1.05E-05	4.45E-03	7.23E-04
9401	9512	9601	9712	2.26E-06	5.26E-05	2.87E-03	1.56E-03
9501	9612	9701	9812	1.10E-05	1.94E-05	3.78E-03	1.73E-03
Heteroskedastic							
8601	8712	8801	8912	1.45E-06	5.79E-06	4.20E-03	3.89E-04
8701	8812	8901	9012	9.04E-07	5.96E-06	1.97E-03	2.91E-04
8801	8912	9001	9112	1.01E-06	4.08E-06	1.80E-03	2.28E-04
8901	9012	9101	9212	1.18E-06	4.28E-06	4.15E-03	4.24E-04
9001	9112	9201	9312	1.22E-06	3.58E-06	2.61E-03	2.67E-04
9101	9212	9301	9412	1.14E-06	5.46E-06	4.62E-03	6.07E-04
9201	9312	9401	9512	1.53E-06	7.18E-06	1.61E-03	3.92E-04
9301	9412	9501	9612	2.06E-06	8.67E-06	2.16E-03	4.71E-04
9401	9512	9601	9712	2.34E-06	2.04E-05	2.10E-03	5.59E-04
9501	9612	9701	9812	2.44E-06	1.72E-05	3.58E-03	1.05E-03

These failures can be seen most clearly in the Figures below. Generally forecast errors are higher following fluctuation points in the pattern of accessions or terminations. When the outcome variable has a highly volatile pattern with spikes, for example AFDC-U termination rates, the forecast performs very poorly.

Figure 6.1.a: AMSFE ARU (Rolling 2 Years)

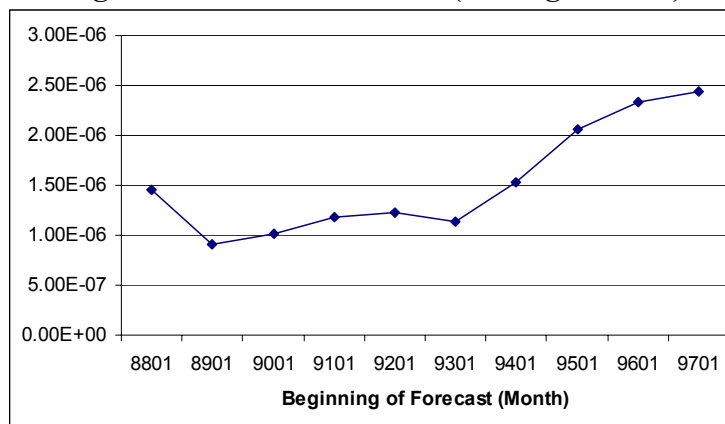


Figure 6.1.b: AMSFE ARFG (Rolling 2 Years)

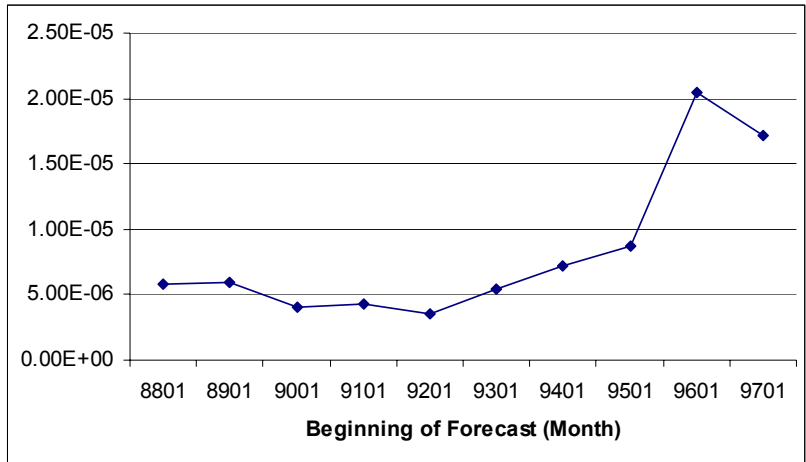


Figure 6.1.c: AMSFE TRU (Rolling 2 Years)

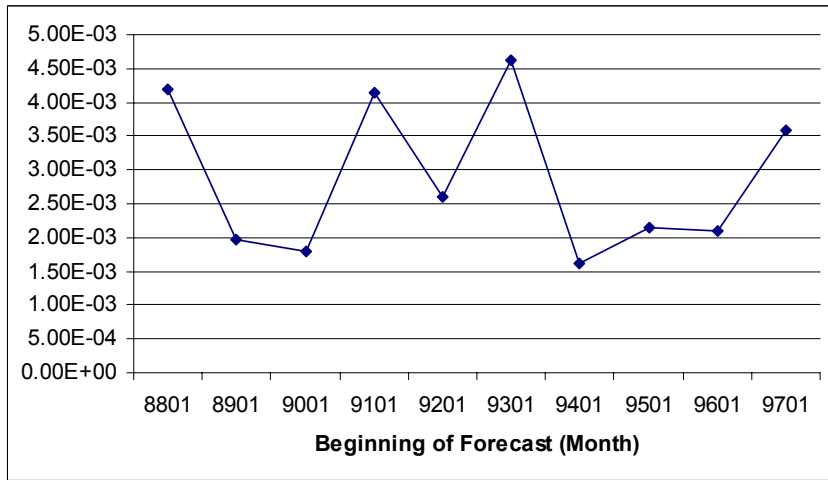
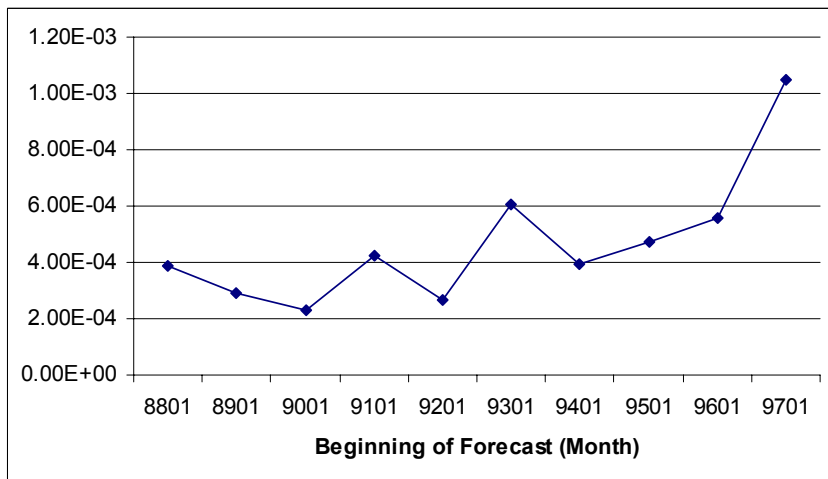


Figure 6.1.d: AMSFE TRFG (Rolling 2 Years)



6.5 Optimal Estimation Lengths for Varying Forecast Years

In Section 6.2, we estimated the same model multiple times with varying start dates for the estimation period. Restricting the estimation period had an impact on forecast accuracy. In that case, however, the forecast period was fixed. In Section 6.3, we varied both the estimation and forecast periods, but kept the length of the estimation period constant. We found that short estimation periods may work better for stable periods, but less well for unstable periods. Longer estimation periods tend to forecast unstable periods more accurately because they include a memory of previous changes.

In this section, we now vary the forecast and estimation periods as well as the length of the estimation period. Only the length of the forecast is fixed at two years. Tables 6.6.a through 6.6.d rank the accuracy of forecasts (AMSFE) by the length of the estimation period for each of the forecast periods. The results confirm the findings from Section 6.2 and 6.3. Shorter time periods have the smallest AMSFE except during inflection points. During changes in the trends of each of the outcome variables, longer time periods tend to forecast more accurately. Comparison of each table with the corresponding graph illustrates this result. The actual AMSFE for each of the forecasts can be found in Tables A10.a through A10.d in the Appendix.

Shorter time periods are more useful counterfactuals only in stable time periods. Longer estimation periods perform well in both stable and unstable periods. Given that there is a probability of an inflection point (i.e., instability) in any given period, the longer estimation period seems more useful for forecasting than the shorter estimation period. The longer estimation period provides a more useful counterfactual because it includes more information about the movements of variables.

Table 6.6.a: Optimal Estimation Period Length by First Year of Forecast: AFDC-U Accession Rate

Rank	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
1	2	3	2	3	2	5	2	3	5	5
2		2	3	4	4	2	7	4	3	7
3			4	5	5	6	8	7	7	10
4				2	6	7	6	5	8	4
5					3	3	5	2	4	6
6						4	3	8	6	9
7							4	9	9	8
8								6	10	2
9									2	3
10										11

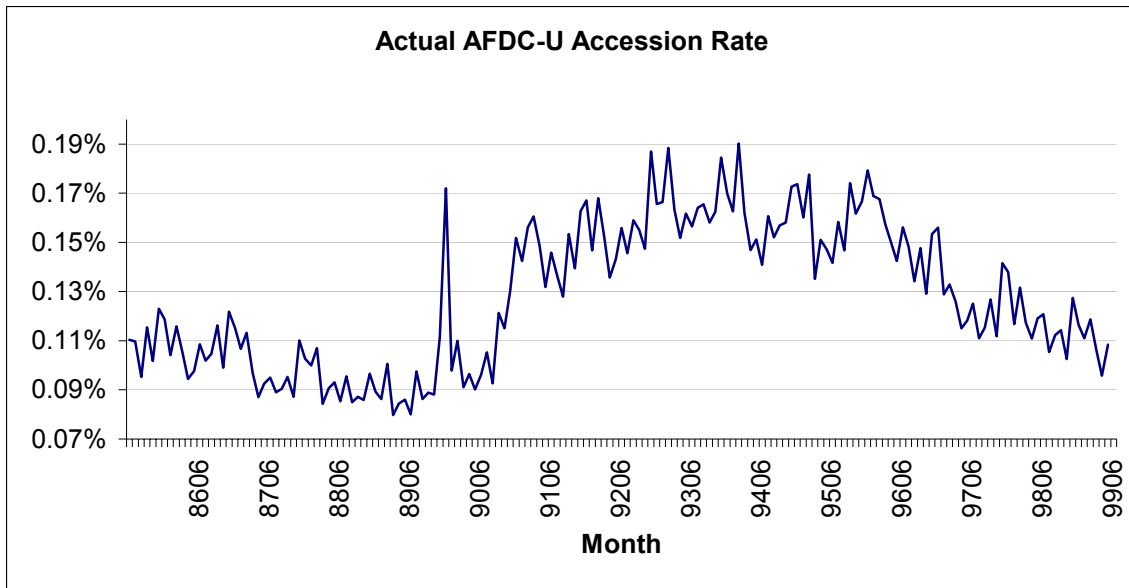


Table 6.6.b: Optimal Estimation Period Length by First Year of Forecast: AFDC-FG Accession Rate

Rank	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
1	2	2	2	4	6	2	5	3	3	9
2		3	4	3	5	4	6	6	4	5
3			3	5	2	3	7	2	7	11
4				2	3	7	8	7	9	10
5					4	5	2	5	5	8
6						6	4	9	10	3
7							3	8	8	6
8								4	2	2
9									6	4
10										7

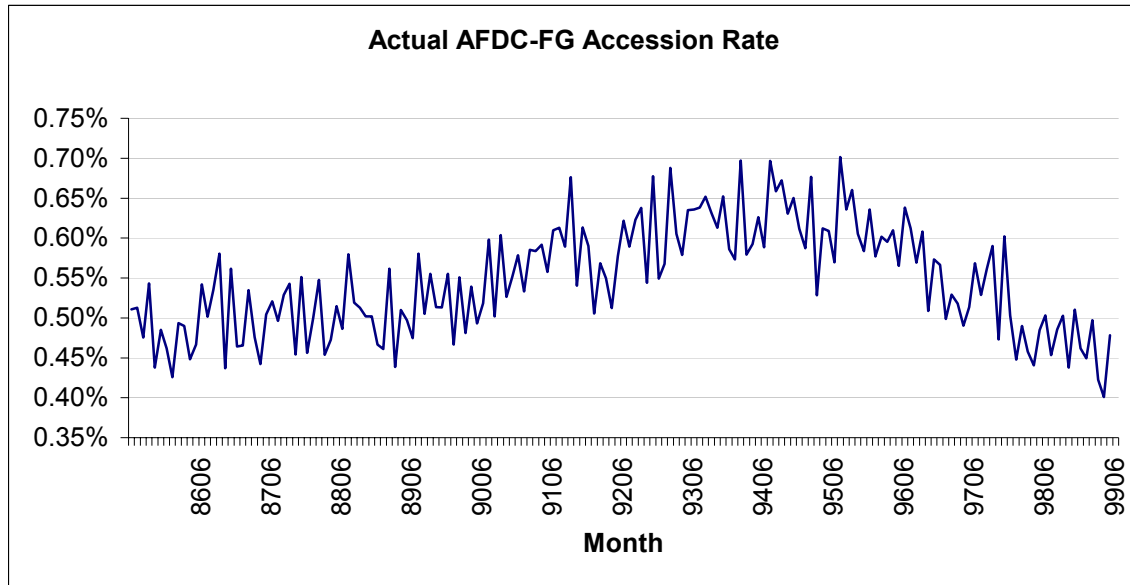


Table 6.6.c: Optimal Estimation Period Length by First Year of Forecast: AFDC-U Termination Rate

Rank	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
1	2	2	2	5	6	7	2	3	2	5
2		3	4	4	2	6	5	6	6	9
3			3	3	5	3	7	9	7	3
4				2	4	5	8	2	4	7
5					3	4	4	8	9	6
6						2	6	5	8	8
7							3	7	3	4
8								4	10	2
9									5	11
10										10

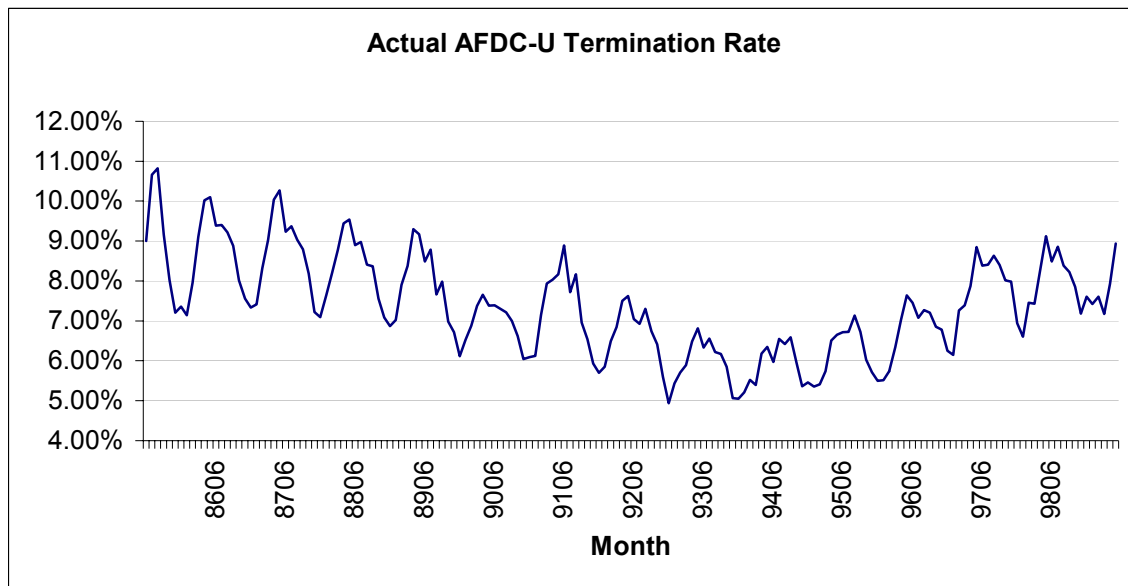
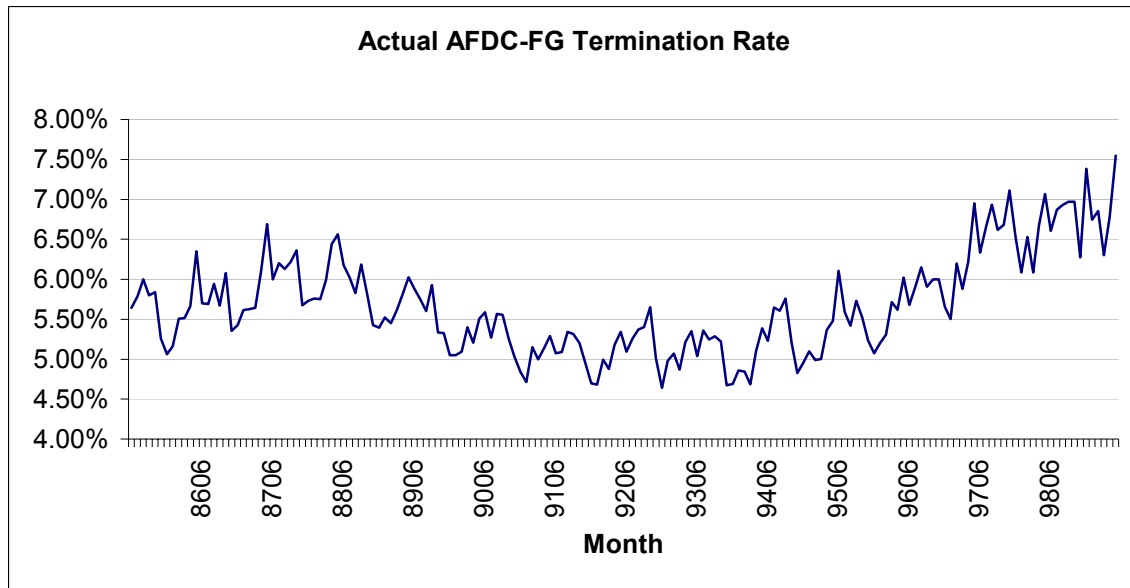


Table 6.6.d: Optimal Estimation Period Length by First Year of Forecast: AFDC-FG Termination Rate

Rank	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
1	2	2	2	3	2	7	2	7	2	2
2		3	4	5	6	3	6	3	6	3
3			3	4	5	4	7	2	8	4
4				2	4	6	5	6	9	8
5					3	2	3	8	10	10
6						5	8	9	3	11
7							4	4	4	5
8								5	7	9
9									5	6
10										7



6.6 Calculating the Minimum Required Impact

What would it take to identify the impact of a change in policy using these forecasting models? Specifically, how large a change in the outcome would be necessary in order for our models to be able to differentiate a true impact from forecast error?

We can calculate the required change using the AMSFE and the actual level of the outcome variable. The first row shows the AMSFE for the heteroskedastic model using data from 1985 through 1996 to estimate the two-year forecast period 1997 and 1998.³ The AMSFEs for each

³ As discussed in Section 6.5, a longer estimation period seems to provide the most reliability. As a result, we use a long estimation period to calculate the minimum required impact in this section. It is important to note that using a different estimation period and hence a different AMSFE will change the calculation of the minimum required impact.

outcome are shown in Table 6.7.a along with the mean of the actual value over the same two years. These values are used to calculate the minimum size of the change required to differentiate deviations from forecast error. We calculate the minimum size of changes individually for each county using the panel heteroskedastic model. There is considerable variation across counties in terms of how small an impact the model would be able to identify. Table 6.7.a. shows the mean, median, minimum and maximum values for the minimum required change across counties.⁴ Appendix Table A11 shows the actual values for individual counties. There seems to be no discernable pattern with which to distinguish between better performing counties from worse performing ones.

Table 6.7.a: Minimum Change Required to Identify Impact

	ARU	ARFG	TRU	TRFG
AMSFE 1997-1998 (Heteroskedastic)	2.15E-06	1.65E-05	5.14E-03	1.23E-03
Mean Actual Value 1997-1998	2.58E-03	8.34E-03	1.43E-01	9.81E-02
Minimum Required Change (Median)	74.23	59.26	72.24	58.57
Minimum Required Change (Mean)	87.43	67.29	73.19	59.87
Minimum Required Change (Max)	208.84	206.59	156.35	174.53
Minimum Required Change (Min)	33.12	23.26	23.96	16.45

Our 95% confidence interval indicates that we can only identify an impact greater than two standard deviations from the actual mean value. Consequently, we can identify an impact when comparing our forecasted to actual values if the impact on the AFDC-U accession rate is at least greater than 74%. Any smaller change cannot be differentiated from forecast error because it falls within two standard deviations. The required changes for the other outcomes are 59% for the AFDC-FG accession rate, 72% for the AFDC-U termination rate, and 59% for AFDC-FG termination rate.

The minimum required change was calculated for both the homoskedastic and heteroskedastic models using different estimation lengths (see Table 6.7.b). The results show that substantial minimum changes are required to identify an impact for AFDC-U and AFDC-FG accession and termination rates. The most problematic model is the AFDC-U accession rate where large changes are required for identification; for some estimation lengths, the minimum required change is three times larger than the level of the dependent variable.

The homoskedastic models generally require larger minimum changes than the heteroskedastic models in order to identify an impact. Homoskedastic models also have greater fluctuations in the required minimum change across estimation lengths. The required minimum changes calculated from the heteroskedastic models are more robust to changes in estimation length.

⁴ These minimum required changes are large relative to the calculations in Section 6.1. Minimum required change calculations are made by taking the square root of the AMSFE, since the AMSFE is a number smaller than one, the square root is large.

**Table 6.7.b: Minimum Change Required to Identify Impact by Estimation Length
(Percent of Mean Actual Value)**

		ARU	ARFG	TRU	TRFG
Year	Length	Homoskedastic			
1986	11	90.01	62.93	72.17	58.63
1987	10	93.46	62.00	76.05	55.95
1988	9	88.48	63.50	76.82	54.61
1989	8	102.64	56.14	68.54	49.74
1990	7	152.28	57.91	70.39	49.11
1991	6	149.20	59.27	73.66	50.35
1992	5	151.22	56.75	68.28	48.65
1993	4	198.94	60.75	69.16	49.80
1994	3	211.42	58.45	74.40	48.07
1995	2	300.22	57.98	61.03	64.65
1996	1	92.69	65.16	95.25	63.12
Year	Length	Heteroskedastic			
1986	11	74.24	58.64	72.38	57.97
1987	10	74.11	60.95	73.57	56.16
1988	9	74.31	60.98	74.58	50.75
1989	8	74.09	55.16	64.29	50.34
1990	7	76.71	56.10	73.06	48.94
1991	6	74.09	54.76	76.15	49.39
1992	5	73.76	57.47	64.58	47.77
1993	4	73.09	56.26	62.59	45.75
1994	3	75.95	57.68	67.12	42.31
1995	2	85.33	51.47	58.77	42.45
1996	1	68.21	59.44	72.50	49.59

6.7 AMSFE: The Variance-Bias Tradeoff

The AMSFE can be decomposed into two parts: the variance and the square of the bias. The variance measures the variability of the forecasts while the bias indicates whether the model consistently over- or under-predicts the outcome variables.

$$\text{AMSFE} = \text{Variance} + \text{Bias}^2$$

$$\text{AMSFE} = \frac{\sum_{t=1}^T \sum_{i=1}^N (\varepsilon_{it})^2}{N*T} = \frac{\sum_{t=1}^T (\varepsilon_{it} - \mu_i)^2}{N*T} + \frac{\sum_{t=1}^T (\mu_i)^2}{N*T}$$

$$\text{where } \mu_i = \frac{\sum_{t=1}^T (\varepsilon_{it})}{T}$$

Table 6.8 shows the decomposition of the AMSFE into its variance and bias components for the heteroskedastic model estimated over the entire period from October 1985 through December

1996 for all four outcome measures. Bias is always positive indicating that our models consistently under-predict accession and termination rates. We also find that the variance and bias components are roughly equal. Appendix Tables A12.a through A12.b show results of the decomposition for individual counties for each outcome variable.

**Table 6.8: Decomposition of AMSFE into Variance and Bias Components
(Means across Counties)**

	ARU	ARFG	TRU	TRFG
Variance (normalized)*	2.85E-01	2.27E-01	2.58E-01	1.93E-01
Bias (normalized)*	2.79E-01	2.12E-01	2.28E-01	2.05E-01
AMSFE (normalized)*	4.37E-01	3.36E-01	3.66E-01	2.99E-01

* components normalized by the average of the dependent variable for each county

6.8 Summary

In this chapter, we estimated our time-series and panel data multivariate models over varying time periods to investigate changes in forecast accuracy. We found that forecast accuracy depends on a variety of factors including the variation in the outcome variable over the forecasted period, the similarity of the estimation period to the forecasted period, and the length of the estimation period.

Using an eleven-year estimation period from 1985 to 1996 we found that forecast accuracy varied across the two forecast years: 1997 and 1998. We then forecasted the same time period from 1997-1998 with ever-shortening estimation periods. We found that the most accurate forecasts were produced when very short and very long estimation periods were used. This outcome may be a product of the sample period used. Short periods worked well because 1996 was very similar to 1997-1998. Long periods worked well because there was sufficient information about movements in the outcome across long periods to predict outcomes accurately.

We showed how AMSFE can be normalized by the mean, median, or standard deviation of the dependent variable. All three normalizations show that relative to some measure of the dependent variable, the AMSFE is low, less than 10%.

The above results may be sensitive to our choice of 1997-98 as the forecast period. To address this issue, we then forecasted all years. The result regarding similarity of the estimation and forecast periods was reinforced by two-year forecasts made for all years using the previous two-years as the estimation period. The AMSFE was much smaller when the estimation period was similar to the forecast period. The result regarding length of the estimation period was reinforced by calculations of optimal estimation length. We created two-year forecasts for all years using estimation periods of varying lengths. We again found that shorter estimation periods work well in stable periods and longer estimation periods work well in less stable periods.

The optimal estimation period varies. If we believe that the forecast period is similar to the previous period, then shorter estimation periods may suffice. However, if we believe that there

is a change in the movement of the outcome variable from the previous period, then longer estimation periods improve the forecast. Given that the stability of the forecast period is often unclear, longer estimation periods seem most useful despite the success of shorter estimation periods in stable periods.

We can construct 95% confidence intervals around our out-of-sample forecasts. In order to be able to identify the impact of a change in policy from forecast error, the minimum size of the impact must exceed two standard deviations from the actual level. That is, it must fall outside of our confidence interval. On average, for our heteroskedastic panel data model, only large impacts of approximately 74% and 59% of the actual values are identifiable with our AFDC-U and AFDC-FG accession rate models, respectively. Similarly only large impacts of the same magnitude can be identified using our AFDC-U (72%) and AFDC-FG (59%) termination rate models.

Finally, we can decompose the AMSFE into its two components, variance and bias. We find that these two components are roughly equal in heteroskedastic models of all four outcome variables. Generally our models under-predict the accession and termination rates.

Conclusions

This report has considered methods for establishing causal relationships in welfare research. To a large extent, research of any kind is about establishing causality. By identifying causal relationships, we can explain events such as movements from welfare to work. We can predict outcomes such as caseload reductions, or we can prevent problems such as ineffective programs. In social science and welfare research the quest for explanation and understanding makes causality especially crucial. While experimental methodologies have been the tried and true method of examining cause and effect in welfare research, the current environment does not always allow for social experimentation. The hunt for causality must continue within a non-experimental framework. This report has shown how we can do that.

Summary of the Report -- Chapter 1 discussed ways to determine causality. The neo-Humean, counterfactual, manipulation, and mechanism approach to causality were discussed. The neo-Humean approach requires the correlation of causes and effects and the temporal precedence of causes before effects. Problems arise in demonstrating that a correlation between a cause and an effect is not spuriously produced by some factor other than the putative cause, and temporal precedence is sometimes hard to establish. Counterfactual approaches require verifying not only that a cause leads to an effect but also that the absence of the cause leads to the absence of the effect in a world in which only the cause has been eliminated. The difficulty here is determining what the world would look like if only the cause were suppressed. The manipulation approach requires the actual manipulation of a cause and the demonstration that the manipulation is followed by the effect. Actual manipulations of causes, however, are not always possible, and it is often hard to show that the manipulation is responsible for the effect. Finally, the mechanism approach requires showing that some basic processes tie the cause to the effect. For example, epidemiologists who showed that cigarette smoking was correlated with lung cancer were challenged to show that some biological mechanism linked smoking to tumors. But demonstrating such connections can often be very difficult.

Rather than support one or the other of these theories of causality, we argued that each of them captured important aspects of causality. We used all four to develop general guidelines for making good causal inferences.

Chapter 2 continued this theoretical discussion by examining experimental and non-experimental approaches that can be used to support causal arguments. After more than three decades of experience with experiments in welfare research, randomized experiments are the gold standard for establishing causality because they have the advantages of a clear counterfactual and a clear manipulation, but even with experiments, it is often hard to establish causality.

When only observational studies are possible and experiments cannot be performed, it is even harder to determine causality. Two half-way houses between experimentation and observational studies are quasi-experiments and natural experiments where there is a clear manipulation (quasi-experiments) or a clear counterfactual (natural experiments) but some compromise on the other desired characteristic. When experimentation, natural experiments, or quasi-experiments are not possible, we are left with observational studies in which investigators must work simply to establish genuine correlation between causes and effects. Increasingly sophisticated statistical,

econometric, and logical tools have been developed to improve the interpretation of observational studies, but establishing causality is very hard.

Chapter 3 reviewed some of the econometric methodology that can be and has been applied to determine causal effects in welfare and other research. The chapter begins with the most basic model and builds towards a panel data model that corrects for heteroskedasticity. The time-series cross-sectional data analyzed by panel methods make it possible to implement a “difference-in-differences” methodology comparing over time one group that gets a treatment to another group that does not get it. The groups are first compared before they get the treatment to determine their pre-existing differences. Then changes that occur over time--starting from the baseline for each group--are compared to see if the treatment makes a genuine difference. This econometric method is very powerful, but it depends upon the proper specification of the econometric model.

Chapter 3 also discussed the statistical problems that arise in econometric modeling and the modeling strategies employed by various analysts. This chapter discusses how models can be evaluated in terms of their forecasting capabilities. It notes that models can fit the data very well for the set of observations from which they have been produced, but they can sometimes forecast very badly for observations out-of-sample because they have missed some important factors. Hence, out-of-sample forecasts are an especially stringent and useful test of a model, and we use them extensively in Chapter 6.

Chapter 4 described the data employed in our study. The large number of counties and the diversity of California make it an interesting state to analyze. This report utilizes county level monthly accession and termination data for the state for over fourteen years supplemented with substantial county level demographic and employment information. The chapter describes the strengths and limitations of these data.

In Chapter 5 these data were used to do regression analysis within and across counties, starting with the most basic correlational model and building up to a sophisticated multivariate model. These models were developed for four outcome variables, accession and termination rates for the program for families with one parent (AFDC-FG) and for families with two parents (AFDC-U).

Chapter 6 extended the models built in Chapter 5 to provide out-of-sample forecasts. This chapter includes an extended discussion of Average Mean Squared Forecast Error (AMSFE) and its decomposition into variance and bias components. The chapter changes the forecast period and the amount of data used in the estimation to determine which model works best based upon the AMSFE. In order to facilitate the interpretation of AMSFE, the chapter presents three different normalizations of it, but all three lead to the conclusion that the forecast error is quite substantial for our models. We also discuss how big an impact a treatment would have to make before models such as the ones we have developed could discern a statistically significant impact. We find that substantial changes in accession and termination rates would have to occur before we could use our models to declare reliably that a policy change has had an impact.

How Well Have We Done? -- One way to answer this question is to review the causality checklist of Chapter 2 to see how far we have gone toward establishing causality.

First consider the "General Issues" listed on the checklist. This report has tried to examine a specific cause and effect. We examined how a change in benefit levels would affect accession and termination rates. Causes were defined either as an indicator variable for an increase in benefits or the actual benefit level itself. Effects were measured as changes in the accession or termination rates. By declaring what our expected signs were on the coefficients of the "cause" variables we were explicitly declaring that our causal statement was: "We expect that as benefits increase, accession rates should increase and termination rates should decline." The counterfactual was that benefits did not matter to either outcome variable, and by performing statistical significance tests on the coefficients we were explicitly testing this counterfactual as a viable alternative to our causal statement. We explicitly stated that our analysis was for a specific location, either the entire state of California or each county taken individually, and we specified that the manipulation was a direct action of the state's Department of Social Services. Thus, we produced a well-defined causal assertion.

With this hypothesis in hand, we can consider whether our results meet the requirements of the four causal theories. We begin with the requirements of the Neo-Humean Theory, and we consider constant conjunction. When we estimated our multivariate models county by county we found that we always had coefficients of the expected sign, but they were not often significant. We explored the possibilities of alternative explanations for our findings by the inclusion of covariates. In our best model, we found that changing the benefit level had the expected impact on accessions and terminations, but the evidence was not very strong.

Meeting the requirements of the Counterfactual Theory poses even more difficulties. The only way that we could compare situations with and without the putative cause was to use covariates in the model to perform the role of adjusting our observational data to create facsimiles of the "most similar" alternative world. By controlling for the influence of other factors we created the best possible comparable counterfactual. One place where we obviously failed to do a very good job was with the use of dummy variables for each year. These variables help make adjustments from one year to the next, but they do nothing to help us identify the specific factors that cause one year to be different from another.

A major demand of the Manipulation Theory of causality is the explicit description of the cause under investigation. There is an important lesson for research in this requirement. Research that does not specify a manipulation cannot claim to have identified a causal relationship. One way to identify a manipulation is through studies of the treatment implementation that are designed to measure how (and whether) the manipulation actually occurred. Like most observational studies, our models do not provide this kind of check on the manipulation. If we had some instrumental variable available which was correlated with the treatment, then we could check on the manipulation, but there is no obvious candidate for such an instrumental variable.

Mechanism Theories of causality ask about the mechanism by which the cause brings about the effect. Looking for the mechanism is important in establishing causality. We would like to know the mechanism by which benefit levels can influence accession and termination rates. Presumably this mechanism involves some calculation on the part of current or prospective welfare recipients who encounter a change in benefits and then make decisions about signing up

for or leaving welfare. If we believe that the mechanism is the need for living expenses then it would be useful to know more about how people learn about benefits and how much difference small changes might make in their decisions. Does it really make sense, for example, that any change in the benefit level, even as little as a dollar (or for that matter one penny) could push someone onto or away from welfare? Our use of the dichotomous treatment variable relies on this assumption implicitly although our continuous variable is more refined. Careful consideration of the mechanism is important to understanding the nature of causality.

Although we attempted to follow the guidelines for establishing causality discussed in the first chapter, we found that it was very hard to meet these requirements with observational data for welfare accessions and terminations. The data themselves and the problems of fully specifying a suitable model limit how much we are actually able to do. These problems typically occur for any observational study. In the end, our actual empirical findings are disappointing in terms of our ability to define a clear counterfactual or manipulation and to find impacts, but we have learned a great deal about the assumptions (both explicit and implicit) that we and other researchers on welfare make.

Appendix

Table A1.a: ARU Results Within County - Across Coefficients (Dichotomous Aid)

County Number	County Name	% Right	%Wrong	%Right & Significant	%Wrong & Significant	AMSFE	Adusted R-squared	Number of Observations
1	Alameda	61.90	38.10	14.29	0.00	1.54E-07	0.33	117
2	Alpine							
3	Amador	57.14	42.86	9.52	0.00	4.18E-05	0.27	135
4	Butte	66.67	33.33	0.00	0.00	8.13E-05	0.46	135
5	Calaveras	52.38	47.62	4.76	0.00	2.65E-06	0.37	117
6	Colusa	61.90	38.10	14.29	4.76	1.68E-05	0.80	135
7	Contra Costa	80.95	19.05	0.00	0.00	2.98E-07	0.56	117
8	Del Norte	66.67	33.33	4.76	14.29	4.75E-06	0.68	135
9	El Dorado	57.14	42.86	0.00	0.00	1.26E-06	0.50	135
10	Fresno	61.90	38.10	0.00	0.00	8.07E-06	0.65	135
11	Glenn	71.43	28.57	14.29	9.52	3.11E-03	0.36	135
12	Humboldt	57.14	42.86	4.76	4.76	1.80E-05	0.42	135
13	Imperial	61.90	38.10	9.52	0.00	6.83E-05	0.61	135
14	Inyo	47.62	52.38	4.76	0.00	2.11E-06	0.06	117
15	Kern	42.86	57.14	9.52	4.76	2.42E-04	0.91	135
16	Kings	57.14	42.86	4.76	4.76	3.93E-06	0.50	135
17	Lake	57.14	42.86	14.29	0.00	2.07E-04	0.36	135
18	Lassen	38.10	61.90	0.00	0.00	5.40E-05	0.36	135
19	Los Angeles	52.38	47.62	9.52	0.00	3.01E-04	0.81	135
20	Madera	42.86	57.14	4.76	9.52	2.02E-05	0.65	135
21	Marin	66.67	33.33	0.00	0.00	7.29E-08	0.55	117
22	Mariposa	66.67	33.33	4.76	0.00	1.06E-05	0.14	90
23	Mendocino	47.62	52.38	0.00	4.76	8.60E-05	0.77	135
24	Merced	61.90	38.10	9.52	0.00	9.02E-05	0.57	135
25	Modoc	57.14	42.86	9.52	0.00	1.95E-05	0.22	135
26	Mono	80.95	19.05	0.00	0.00	6.99E-07	0.11	107
27	Monterey	61.90	38.10	19.05	4.76	5.67E-06	0.87	135
28	Napa	52.38	47.62	4.76	4.76	8.59E-06	0.56	117
29	Nevada	66.67	33.33	0.00	0.00	4.59E-06	0.48	135
30	Orange	66.67	33.33	4.76	0.00	1.66E-05	0.95	135
31	Placer	66.67	33.33	9.52	0.00	1.49E-07	0.46	135
32	Plumas	61.90	38.10	14.29	0.00	2.20E-05	0.62	135
33	Riverside	57.14	42.86	0.00	0.00	6.09E-07	0.85	117
34	Sacramento	61.90	38.10	9.52	9.52	3.77E-05	0.65	135
35	San Benito	61.90	38.10	0.00	0.00	9.09E-07	0.52	135
36	San Bernardino	42.86	57.14	0.00	4.76	6.52E-08	0.69	117
37	San Diego	66.67	33.33	4.76	0.00	1.23E-05	0.84	135
38	San Francisco	47.62	52.38	4.76	0.00	1.38E-08	0.49	117
39	San Joaquin	57.14	42.86	14.29	4.76	6.90E-06	0.80	135
40	San Luis Obispo	61.90	38.10	0.00	4.76	5.04E-06	0.76	135
41	San Mateo	52.38	47.62	0.00	0.00	1.40E-07	0.81	117
42	Santa Barbara	66.67	33.33	0.00	4.76	1.15E-04	0.88	135
43	Santa Clara	71.43	28.57	0.00	0.00	5.06E-07	0.87	135
44	Santa Cruz	61.90	38.10	14.29	0.00	3.04E-07	0.79	135
45	Shasta	71.43	28.57	4.76	0.00	1.79E-06	0.38	135
46	Sierra							
47	Siskiyou	52.38	47.62	0.00	4.76	2.32E-06	0.23	135
48	Solano	52.38	47.62	14.29	4.76	4.94E-08	0.41	117
49	Sonoma	66.67	33.33	0.00	4.76	2.54E-08	0.51	135
50	Stanislaus	76.19	23.81	0.00	0.00	4.51E-05	0.76	135
51	Sutter	52.38	47.62	4.76	4.76	2.66E-04	0.66	45
52	Tehama	52.38	47.62	0.00	0.00	5.79E-05	0.44	135
53	Trinity	61.90	38.10	4.76	4.76	1.78E-05	0.20	117
54	Tulare	57.14	42.86	4.76	0.00	6.83E-06	0.74	135
55	Tuolumne	57.14	42.86	0.00	0.00	4.87E-05	0.25	135
56	Ventura	61.90	38.10	4.76	4.76	1.66E-05	0.71	135
57	Yolo	52.38	47.62	0.00	0.00	1.01E-05	0.58	135
58	Yuba	71.43	28.57	23.81	14.29	7.65E-03	0.80	45

Table A1.b: ARU Results AcrossCounty - Within Coefficients (Dichotomous Aid)

	# Right	# Right & Significant	% Right	%Right & Significant
Dichotomous Aid	56	12	100.00%	21.43%
Farm	33	6	58.93%	10.71%
Farm (t-1)	35	3	62.50%	5.36%
Farm (t-2)	24	1	42.86%	1.79%
Farm (t-3)	27	2	48.21%	3.57%
Service	23	0	41.07%	0.00%
Service (t-1)	29	3	51.79%	5.36%
Service (t-2)	28	2	50.00%	3.57%
Service (t-3)	35	1	62.50%	1.79%
Retail	32	1	57.14%	1.79%
Retail (t-1)	28	3	50.00%	5.36%
Retail (t-2)	32	2	57.14%	3.57%
Retail (t-3)	26	3	46.43%	5.36%
Other	0	0	0.00%	0.00%
Other (t-1)	31	1	55.36%	1.79%
Other (t-2)	26	1	46.43%	1.79%
Other (t-3)	30	1	53.57%	1.79%
CA Min Wage	41	6	73.21%	10.71%
Percent Hispanic	25	2	44.64%	3.57%
Percent Black	56	6	100.00%	10.71%
Birth Rate	56	7	100.00%	12.50%

Table A2.a: ARFG Results Within County - Across Coefficients (Dichotomous Aid)

County Number	County Name	% Right	%Wrong	%Right & Significant	%Wrong & Significant	AMSFE	Adusted R-squared	Number of Observations
1	Alameda	57.14	42.86	14.29	4.76	2.37E-05	0.40	117
2	Alpine							
3	Amador	61.90	38.10	0.00	4.76	1.83E-05	0.24	135
4	Butte	57.14	42.86	9.52	4.76	1.95E-04	0.61	135
5	Calaveras	61.90	38.10	0.00	0.00	2.30E-05	0.26	117
6	Colusa	66.67	33.33	9.52	4.76	2.11E-03	0.48	135
7	Contra Costa	61.90	38.10	0.00	0.00	4.25E-06	0.35	117
8	Del Norte	66.67	33.33	0.00	4.76	3.98E-04	0.32	135
9	El Dorado	47.62	52.38	4.76	0.00	1.82E-06	0.37	135
10	Fresno	57.14	42.86	0.00	0.00	1.82E-03	0.41	135
11	Glenn	61.90	38.10	9.52	14.29	2.05E-02	0.48	135
12	Humboldt	57.14	42.86	4.76	4.76	1.71E-04	0.26	135
13	Imperial	66.67	33.33	4.76	4.76	1.43E-04	0.43	135
14	Inyo	42.86	57.14	4.76	0.00	3.32E-05	0.30	117
15	Kern	38.10	61.90	4.76	0.00	3.67E-04	0.91	135
16	Kings	47.62	52.38	4.76	0.00	2.65E-03	0.35	135
17	Lake	57.14	42.86	0.00	0.00	1.17E-03	0.22	135
18	Lassen	47.62	52.38	0.00	9.52	7.43E-04	0.51	135
19	Los Angeles	66.67	33.33	19.05	4.76	4.72E-04	0.81	135
20	Madera	61.90	38.10	0.00	14.29	1.24E-05	0.47	135
21	Marin	76.19	23.81	4.76	0.00	2.94E-07	0.53	117
22	Mariposa	66.67	33.33	9.52	4.76	1.08E-05	0.16	90
23	Mendocino	47.62	52.38	0.00	0.00	4.34E-04	0.76	135
24	Merced	66.67	33.33	0.00	0.00	4.66E-04	0.58	135
25	Modoc	57.14	42.86	0.00	0.00	1.01E-04	-0.03	135
26	Mono	52.38	47.62	0.00	4.76	3.92E-06	0.05	107
27	Monterey	61.90	38.10	0.00	4.76	1.23E-04	0.73	135
28	Napa	47.62	52.38	0.00	0.00	1.17E-05	0.52	117
29	Nevada	61.90	38.10	0.00	4.76	9.60E-05	0.61	135
30	Orange	66.67	33.33	4.76	0.00	4.87E-05	0.90	135
31	Placer	76.19	23.81	0.00	0.00	3.45E-06	0.58	135
32	Plumas	52.38	47.62	4.76	4.76	2.40E-05	0.51	135
33	Riverside	52.38	47.62	4.76	0.00	1.04E-05	0.74	117
34	Sacramento	66.67	33.33	4.76	0.00	5.54E-05	0.76	135
35	San Benito	71.43	28.57	9.52	4.76	1.83E-05	0.18	135
36	San Bernardino	57.14	42.86	0.00	4.76	1.05E-05	0.43	117
37	San Diego	61.90	38.10	0.00	0.00	1.77E-04	0.49	135
38	San Francisco	57.14	42.86	0.00	4.76	1.54E-07	0.44	117
39	San Joaquin	57.14	42.86	9.52	4.76	6.77E-05	0.86	135
40	San Luis Obispo	42.86	57.14	0.00	0.00	3.41E-06	0.68	135
41	San Mateo	52.38	47.62	0.00	0.00	1.30E-06	0.34	117
42	Santa Barbara	57.14	42.86	4.76	4.76	2.91E-04	0.82	135
43	Santa Clara	47.62	52.38	0.00	4.76	4.45E-05	0.70	135
44	Santa Cruz	61.90	38.10	9.52	0.00	3.86E-06	0.57	135
45	Shasta	80.95	19.05	0.00	0.00	8.82E-04	0.41	135
46	Sierra							
47	Siskiyou	66.67	33.33	9.52	4.76	7.28E-04	0.38	135
48	Solano	66.67	33.33	4.76	0.00	7.00E-06	0.31	117
49	Sonoma	57.14	42.86	4.76	4.76	7.63E-06	0.55	135
50	Stanislaus	57.14	42.86	4.76	4.76	2.49E-03	0.83	135
51	Sutter	57.14	42.86	4.76	0.00	9.87E-03	-0.11	45
52	Tehama	52.38	47.62	0.00	0.00	1.18E-05	0.52	135
53	Trinity	66.67	33.33	4.76	0.00	1.14E-04	0.06	117
54	Tulare	57.14	42.86	0.00	0.00	6.48E-04	0.34	135
55	Tuolumne	61.90	38.10	0.00	0.00	3.83E-06	0.27	135
56	Ventura	57.14	42.86	0.00	4.76	3.36E-05	0.40	135
57	Yolo	47.62	52.38	4.76	0.00	1.52E-06	0.51	135
58	Yuba	52.38	47.62	19.05	4.76	1.97E-02	0.79	45

Table A2.b: ARFG Results AcrossCounty - Within Coefficients (Dichotomous Aid)

	# Right	# Right & Significant	% Right	%Right & Significant
Dichotomous Aid	56	6	100.00%	10.71%
Farm	31	1	55.36%	1.79%
Farm (t-1)	28	2	50.00%	3.57%
Farm (t-2)	24	3	42.86%	5.36%
Farm (t-3)	30	0	53.57%	0.00%
Service	23	2	41.07%	3.57%
Service (t-1)	25	2	44.64%	3.57%
Service (t-2)	27	1	48.21%	1.79%
Service (t-3)	34	1	60.71%	1.79%
Retail	31	1	55.36%	1.79%
Retail (t-1)	32	1	57.14%	1.79%
Retail (t-2)	35	0	62.50%	0.00%
Retail (t-3)	23	1	41.07%	1.79%
Other	0	0	0.00%	0.00%
Other (t-1)	24	1	42.86%	1.79%
Other (t-2)	36	1	64.29%	1.79%
Other (t-3)	28	2	50.00%	3.57%
CA Min Wage	37	6	66.07%	10.71%
Percent Hispanic	29	2	51.79%	3.57%
Percent Black	56	4	100.00%	7.14%
Birth Rate	56	5	100.00%	8.93%

Table A3.a: TRU Results Within County - Across Coefficients (Dichotomous Aid)

County Number	County Name	% Right	%Wrong	%Right & Significant	%Wrong & Significant	AMSFE	Adjusted R-squared	Number of Observations
1	Alameda	61.90	38.10	14.29	0.00	1.54E-07	0.33	117
2	Alpine							
3	Amador	57.14	42.86	9.52	0.00	4.18E-05	0.27	135
4	Butte	66.67	33.33	0.00	0.00	8.13E-05	0.46	135
5	Calaveras	52.38	47.62	4.76	0.00	2.65E-06	0.37	117
6	Colusa	61.90	38.10	14.29	4.76	1.68E-05	0.80	135
7	Contra Costa	80.95	19.05	0.00	0.00	2.98E-07	0.56	117
8	Del Norte	66.67	33.33	4.76	14.29	4.75E-06	0.68	135
9	El Dorado	57.14	42.86	0.00	0.00	1.26E-06	0.50	135
10	Fresno	61.90	38.10	0.00	0.00	8.07E-06	0.65	135
11	Glenn	71.43	28.57	14.29	9.52	3.11E-03	0.36	135
12	Humboldt	57.14	42.86	4.76	4.76	1.80E-05	0.42	135
13	Imperial	61.90	38.10	9.52	0.00	6.83E-05	0.61	135
14	Inyo	47.62	52.38	4.76	0.00	2.11E-06	0.06	117
15	Kern	42.86	57.14	9.52	4.76	2.42E-04	0.91	135
16	Kings	57.14	42.86	4.76	4.76	3.93E-06	0.50	135
17	Lake	57.14	42.86	14.29	0.00	2.07E-04	0.36	135
18	Lassen	38.10	61.90	0.00	0.00	5.40E-05	0.36	135
19	Los Angeles	52.38	47.62	9.52	0.00	3.01E-04	0.81	135
20	Madera	42.86	57.14	4.76	9.52	2.02E-05	0.65	135
21	Marin	66.67	33.33	0.00	0.00	7.29E-08	0.55	117
22	Mariposa	66.67	33.33	4.76	0.00	1.06E-05	0.14	90
23	Mendocino	47.62	52.38	0.00	4.76	8.60E-05	0.77	135
24	Merced	61.90	38.10	9.52	0.00	9.02E-05	0.57	135
25	Modoc	57.14	42.86	9.52	0.00	1.95E-05	0.22	135
26	Mono	80.95	19.05	0.00	0.00	6.99E-07	0.11	107
27	Monterey	61.90	38.10	19.05	4.76	5.67E-06	0.87	135
28	Napa	52.38	47.62	4.76	4.76	8.59E-06	0.56	117
29	Nevada	66.67	33.33	0.00	0.00	4.59E-06	0.48	135
30	Orange	66.67	33.33	4.76	0.00	1.66E-05	0.95	135
31	Placer	66.67	33.33	9.52	0.00	1.49E-07	0.46	135
32	Plumas	61.90	38.10	14.29	0.00	2.20E-05	0.62	135
33	Riverside	57.14	42.86	0.00	0.00	6.09E-07	0.85	117
34	Sacramento	61.90	38.10	9.52	9.52	3.77E-05	0.65	135
35	San Benito	61.90	38.10	0.00	0.00	9.09E-07	0.52	135
36	San Bernardino	42.86	57.14	0.00	4.76	6.52E-08	0.69	117
37	San Diego	66.67	33.33	4.76	0.00	1.23E-05	0.84	135
38	San Francisco	47.62	52.38	4.76	0.00	1.38E-08	0.49	117
39	San Joaquin	57.14	42.86	14.29	4.76	6.90E-06	0.80	135
40	San Luis Obispo	61.90	38.10	0.00	4.76	5.04E-06	0.76	135
41	San Mateo	52.38	47.62	0.00	0.00	1.40E-07	0.81	117
42	Santa Barbara	66.67	33.33	0.00	4.76	1.15E-04	0.88	135
43	Santa Clara	71.43	28.57	0.00	0.00	5.06E-07	0.87	135
44	Santa Cruz	61.90	38.10	14.29	0.00	3.04E-07	0.79	135
45	Shasta	71.43	28.57	4.76	0.00	1.79E-06	0.38	135
46	Sierra							
47	Siskiyou	52.38	47.62	0.00	4.76	2.32E-06	0.23	135
48	Solano	52.38	47.62	14.29	4.76	4.94E-08	0.41	117
49	Sonoma	66.67	33.33	0.00	4.76	2.54E-08	0.51	135
50	Stanislaus	76.19	23.81	0.00	0.00	4.51E-05	0.76	135
51	Sutter	52.38	47.62	4.76	4.76	2.66E-04	0.66	45
52	Tehama	52.38	47.62	0.00	0.00	5.79E-05	0.44	135
53	Trinity	61.90	38.10	4.76	4.76	1.78E-05	0.20	117
54	Tulare	57.14	42.86	4.76	0.00	6.83E-06	0.74	135
55	Tuolumne	57.14	42.86	0.00	0.00	4.87E-05	0.25	135
56	Ventura	61.90	38.10	4.76	4.76	1.66E-05	0.71	135
57	Yolo	52.38	47.62	0.00	0.00	1.01E-05	0.58	135
58	Yuba	71.43	28.57	23.81	14.29	7.65E-03	0.80	45

Table A3.b: TRU Results AcrossCounty - Within Coefficients (Dichotomous Aid)

	# Right	# Right & Significant	% Right	%Right & Significant
Dichotomous Aid	56	12	100.00%	21.43%
Farm	28	0	50.00%	0.00%
Farm (t-1)	26	1	46.43%	1.79%
Farm (t-2)	35	1	62.50%	1.79%
Farm (t-3)	27	2	48.21%	3.57%
Service	22	3	39.29%	5.36%
Service (t-1)	30	5	53.57%	8.93%
Service (t-2)	25	0	44.64%	0.00%
Service (t-3)	34	4	60.71%	7.14%
Retail	37	4	66.07%	7.14%
Retail (t-1)	26	2	46.43%	3.57%
Retail (t-2)	20	1	35.71%	1.79%
Retail (t-3)	32	4	57.14%	7.14%
Other	0	0	0.00%	0.00%
Other (t-1)	25	2	44.64%	3.57%
Other (t-2)	39	1	69.64%	1.79%
Other (t-3)	30	3	53.57%	5.36%
CA Min Wage	33	5	58.93%	8.93%
Percent Hispanic	31	5	55.36%	8.93%
Percent Black	56	7	100.00%	12.50%
Birth Rate	56	10	100.00%	17.86%

Table A4.a: TRFG Results Within County - Across Coefficients (Dichotomous Aid)

County Number	County Name	% Right	%Wrong	%Right & Significant	%Wrong & Significant	AMSFE	Adusted R-squared	Number of Observations
1	Alameda	61.90	38.10	9.52	0.00	1.98E-04	0.54	117
2	Alpine							
3	Amador	52.38	47.62	0.00	0.00	4.91E-02	-0.07	135
4	Butte	61.90	38.10	0.00	4.76	1.86E-02	0.58	135
5	Calaveras	57.14	42.86	0.00	0.00	2.93E-03	0.10	117
6	Colusa	61.90	38.10	9.52	4.76	4.99E-02	0.54	135
7	Contra Costa	61.90	38.10	4.76	0.00	3.74E-04	0.29	117
8	Del Norte	66.67	33.33	19.05	4.76	4.41E-03	0.45	135
9	El Dorado	57.14	42.86	9.52	0.00	2.47E-03	0.43	135
10	Fresno	57.14	42.86	0.00	0.00	5.65E-04	0.38	135
11	Glenn	61.90	38.10	9.52	0.00	1.46E+00	0.59	135
12	Humboldt	66.67	33.33	0.00	0.00	3.51E-03	0.41	135
13	Imperial	61.90	38.10	0.00	9.52	7.53E-03	0.53	135
14	Inyo	66.67	33.33	0.00	0.00	4.34E-03	0.19	117
15	Kern	61.90	38.10	4.76	0.00	5.69E-04	0.87	135
16	Kings	66.67	33.33	0.00	0.00	3.68E-02	0.47	135
17	Lake	52.38	47.62	0.00	4.76	1.27E-02	0.33	135
18	Lassen	71.43	28.57	4.76	0.00	7.89E-03	0.47	135
19	Los Angeles	52.38	47.62	9.52	9.52	9.42E-02	0.73	135
20	Madera	76.19	23.81	4.76	0.00	8.48E-04	0.68	135
21	Marin	57.14	42.86	0.00	4.76	2.10E-03	0.39	117
22	Mariposa	52.38	47.62	4.76	0.00	3.96E-03	0.17	90
23	Mendocino	71.43	28.57	0.00	0.00	2.30E-03	0.74	135
24	Merced	61.90	38.10	0.00	4.76	3.12E-03	0.60	135
25	Modoc	61.90	38.10	4.76	9.52	1.13E-03	0.13	135
26	Mono	76.19	23.81	0.00	0.00	2.40E-02	-0.10	107
27	Monterey	61.90	38.10	0.00	0.00	2.54E-03	0.64	135
28	Napa	57.14	42.86	0.00	0.00	5.27E-04	0.51	117
29	Nevada	57.14	42.86	0.00	0.00	4.44E-03	0.50	135
30	Orange	66.67	33.33	4.76	9.52	1.51E-03	0.47	135
31	Placer	52.38	47.62	4.76	0.00	6.93E-03	0.42	135
32	Plumas	61.90	38.10	0.00	0.00	1.53E-02	0.50	135
33	Riverside	52.38	47.62	4.76	0.00	2.66E-04	0.46	117
34	Sacramento	42.86	57.14	0.00	4.76	1.57E-02	0.49	135
35	San Benito	66.67	33.33	0.00	4.76	9.13E-04	0.23	135
36	San Bernardino	57.14	42.86	0.00	4.76	5.48E-04	0.45	117
37	San Diego	61.90	38.10	0.00	0.00	1.67E-03	0.84	135
38	San Francisco	61.90	38.10	4.76	0.00	2.33E-05	0.45	117
39	San Joaquin	57.14	42.86	23.81	4.76	1.07E-03	0.90	135
40	San Luis Obispo	71.43	28.57	4.76	0.00	1.46E-02	0.29	135
41	San Mateo	61.90	38.10	0.00	0.00	1.67E-03	0.59	117
42	Santa Barbara	52.38	47.62	0.00	0.00	1.93E-04	0.50	135
43	Santa Clara	61.90	38.10	14.29	14.29	2.11E-02	0.70	135
44	Santa Cruz	61.90	38.10	0.00	0.00	6.78E-02	0.71	135
45	Shasta	47.62	52.38	0.00	4.76	3.14E-02	0.63	135
46	Sierra							
47	Siskiyou	66.67	33.33	4.76	0.00	2.34E-02	0.37	135
48	Solano	52.38	47.62	9.52	14.29	3.89E-04	0.33	117
49	Sonoma	57.14	42.86	19.05	0.00	3.04E-03	0.69	135
50	Stanislaus	47.62	52.38	4.76	4.76	7.62E-03	0.91	135
51	Sutter	71.43	28.57	4.76	0.00	4.00E-01	0.04	45
52	Tehama	47.62	52.38	4.76	4.76	7.12E-03	0.61	135
53	Trinity	52.38	47.62	0.00	0.00	5.00E-03	0.23	117
54	Tulare	47.62	52.38	9.52	4.76	2.86E-02	0.42	135
55	Tuolumne	47.62	52.38	9.52	9.52	1.76E-02	0.45	135
56	Ventura	61.90	38.10	0.00	0.00	3.39E-02	0.44	135
57	Yolo	52.38	47.62	9.52	4.76	6.12E-03	0.56	135
58	Yuba	76.19	23.81	9.52	9.52	7.16E-02	0.89	45

Table A4.b: TRFG Results AcrossCounty - Within Coefficients (Dichotomous Aid)

	# Right	# Right & Significant	% Right	%Right & Significant
Dichotomous Aid	56	4	100.00%	7.14%
Farm	34	5	60.71%	8.93%
Farm (t-1)	23	3	41.07%	5.36%
Farm (t-2)	38	2	67.86%	3.57%
Farm (t-3)	32	2	57.14%	3.57%
Service	29	2	51.79%	3.57%
Service (t-1)	28	2	50.00%	3.57%
Service (t-2)	30	3	53.57%	5.36%
Service (t-3)	31	2	55.36%	3.57%
Retail	32	0	57.14%	0.00%
Retail (t-1)	28	3	50.00%	5.36%
Retail (t-2)	29	1	51.79%	1.79%
Retail (t-3)	30	1	53.57%	1.79%
Other	0	0	0.00%	0.00%
Other (t-1)	29	1	51.79%	1.79%
Other (t-2)	28	0	50.00%	0.00%
Other (t-3)	32	3	57.14%	5.36%
CA Min Wage	24	2	42.86%	3.57%
Percent Hispanic	31	3	55.36%	5.36%
Percent Black	56	6	100.00%	10.71%
Birth Rate	56	5	100.00%	8.93%

Table A5.a: ARU Results Within County - Across Coefficients (Continuous Aid)

County Number	County Name	% Right	%Wrong	%Right & Significant	%Wrong & Significant	AMSFE	Adusted R-squared	Number of Observations
1	Alameda	61.90	38.10	14.29	0.00	1.14E-07	0.33	117
2	Alpine							
3	Amador	52.38	47.62	0.00	0.00	3.50E-05	0.22	135
4	Butte	66.67	33.33	0.00	0.00	1.48E-04	0.46	135
5	Calaveras	52.38	47.62	0.00	0.00	6.21E-07	0.35	117
6	Colusa	71.43	28.57	14.29	4.76	9.40E-05	0.78	135
7	Contra Costa	80.95	19.05	4.76	0.00	3.01E-07	0.56	117
8	Del Norte	66.67	33.33	4.76	14.29	3.49E-05	0.68	135
9	El Dorado	57.14	42.86	4.76	0.00	4.61E-06	0.51	135
10	Fresno	52.38	47.62	0.00	4.76	8.19E-07	0.65	135
11	Glenn	66.67	33.33	4.76	14.29	2.75E-03	0.34	135
12	Humboldt	57.14	42.86	4.76	4.76	6.72E-06	0.42	135
13	Imperial	61.90	38.10	9.52	4.76	2.71E-05	0.56	135
14	Inyo	52.38	47.62	9.52	0.00	1.43E-05	0.11	117
15	Kern	47.62	52.38	9.52	4.76	5.38E-05	0.91	135
16	Kings	61.90	38.10	4.76	4.76	2.77E-05	0.50	135
17	Lake	61.90	38.10	4.76	0.00	1.39E-04	0.34	135
18	Lassen	38.10	61.90	0.00	0.00	1.08E-04	0.35	135
19	Los Angeles	57.14	42.86	4.76	0.00	1.81E-04	0.81	135
20	Madera	47.62	52.38	0.00	4.76	1.62E-04	0.59	135
21	Marin	66.67	33.33	0.00	0.00	5.27E-08	0.55	117
22	Mariposa	66.67	33.33	0.00	0.00	1.11E-04	0.14	90
23	Mendocino	52.38	47.62	0.00	4.76	3.17E-05	0.77	135
24	Merced	61.90	38.10	9.52	0.00	2.94E-05	0.56	135
25	Modoc	57.14	42.86	14.29	0.00	1.63E-04	0.23	135
26	Mono	66.67	33.33	0.00	0.00	9.98E-07	0.11	107
27	Monterey	61.90	38.10	19.05	4.76	1.11E-05	0.88	135
28	Napa	52.38	47.62	4.76	4.76	8.91E-06	0.56	117
29	Nevada	61.90	38.10	0.00	0.00	4.27E-06	0.48	135
30	Orange	66.67	33.33	4.76	4.76	4.15E-05	0.95	135
31	Placer	61.90	38.10	4.76	4.76	1.06E-07	0.44	135
32	Plumas	57.14	42.86	14.29	0.00	2.13E-05	0.61	135
33	Riverside	57.14	42.86	0.00	0.00	3.04E-07	0.85	117
34	Sacramento	57.14	42.86	4.76	4.76	2.51E-05	0.65	135
35	San Benito	66.67	33.33	0.00	0.00	9.35E-07	0.53	135
36	San Bernardino	47.62	52.38	0.00	4.76	4.95E-07	0.68	117
37	San Diego	57.14	42.86	4.76	0.00	1.69E-05	0.84	135
38	San Francisco	47.62	52.38	4.76	0.00	3.39E-08	0.49	117
39	San Joaquin	57.14	42.86	14.29	9.52	3.60E-06	0.83	135
40	San Luis Obispo	47.62	52.38	0.00	4.76	6.40E-06	0.76	135
41	San Mateo	47.62	52.38	0.00	0.00	4.80E-08	0.81	117
42	Santa Barbara	71.43	28.57	0.00	4.76	1.51E-04	0.88	135
43	Santa Clara	71.43	28.57	4.76	4.76	2.61E-06	0.87	135
44	Santa Cruz	61.90	38.10	14.29	0.00	1.21E-06	0.79	135
45	Shasta	71.43	28.57	0.00	0.00	3.38E-05	0.35	135
46	Sierra							
47	Siskiyou	61.90	38.10	0.00	4.76	1.47E-05	0.26	135
48	Solano	52.38	47.62	14.29	4.76	5.28E-08	0.41	117
49	Sonoma	66.67	33.33	0.00	4.76	4.38E-07	0.51	135
50	Stanislaus	76.19	23.81	0.00	0.00	3.90E-05	0.76	135
51	Sutter	52.38	47.62	14.29	14.29	1.37E-03	0.74	45
52	Tehama	52.38	47.62	0.00	0.00	5.84E-05	0.42	135
53	Trinity	61.90	38.10	4.76	4.76	4.00E-05	0.21	117
54	Tulare	57.14	42.86	4.76	4.76	1.03E-04	0.74	135
55	Tuolumne	61.90	38.10	4.76	0.00	9.13E-07	0.25	135
56	Ventura	61.90	38.10	4.76	4.76	1.49E-05	0.72	135
57	Yolo	52.38	47.62	0.00	0.00	2.47E-06	0.58	135
58	Yuba	71.43	28.57	14.29	4.76	7.54E-03	0.77	45

Table A5.b: ARU Results AcrossCounty - Within Coefficients (Continuous Aid)

	# Right	# Right & Significant	% Right	%Right & Significant
Continuous Benefits	56	3	100.00%	5.36%
Farm	31	6	55.36%	10.71%
Farm (t-1)	34	2	60.71%	3.57%
Farm (t-2)	22	2	39.29%	3.57%
Farm (t-3)	27	2	48.21%	3.57%
Service	24	0	42.86%	0.00%
Service (t-1)	31	3	55.36%	5.36%
Service (t-2)	27	2	48.21%	3.57%
Service (t-3)	34	2	60.71%	3.57%
Retail	34	1	60.71%	1.79%
Retail (t-1)	27	3	48.21%	5.36%
Retail (t-2)	32	2	57.14%	3.57%
Retail (t-3)	27	3	48.21%	5.36%
Other	0	0	0.00%	0.00%
Other (t-1)	31	1	55.36%	1.79%
Other (t-2)	26	1	46.43%	1.79%
Other (t-3)	31	1	55.36%	1.79%
CA Min Wage	45	5	80.36%	8.93%
Percent Hispanic	24	3	42.86%	5.36%
Percent Black	56	3	100.00%	5.36%
Birth Rate	56	11	100.00%	19.64%

Table A6.a: ARFG Results Within County - Across Coefficients (Continuous Aid)

County Number	County Name	% Right	%Wrong	%Right & Significant	%Wrong & Significant	AMSFE	Adusted R-squared	Number of Observations
1	Alameda	52.38	47.62	9.52	4.76	3.11E-05	0.40	117
2	Alpine							
3	Amador	66.67	33.33	0.00	0.00	4.02E-06	0.24	135
4	Butte	52.38	47.62	14.29	4.76	9.48E-04	0.62	135
5	Calaveras	61.90	38.10	0.00	9.52	4.87E-06	0.28	117
6	Colusa	66.67	33.33	4.76	4.76	1.79E-03	0.48	135
7	Contra Costa	61.90	38.10	0.00	0.00	4.29E-06	0.37	117
8	Del Norte	57.14	42.86	0.00	4.76	1.22E-03	0.33	135
9	El Dorado	52.38	47.62	14.29	9.52	4.07E-05	0.43	135
10	Fresno	57.14	42.86	0.00	0.00	3.33E-04	0.41	135
11	Glenn	57.14	42.86	9.52	14.29	1.94E-02	0.48	135
12	Humboldt	57.14	42.86	4.76	0.00	2.57E-04	0.26	135
13	Imperial	71.43	28.57	4.76	0.00	3.42E-04	0.42	135
14	Inyo	47.62	52.38	0.00	0.00	5.18E-06	0.24	117
15	Kern	38.10	61.90	4.76	0.00	2.48E-04	0.91	135
16	Kings	47.62	52.38	4.76	0.00	2.34E-03	0.35	135
17	Lake	61.90	38.10	0.00	0.00	3.63E-04	0.20	135
18	Lassen	47.62	52.38	4.76	9.52	5.43E-04	0.51	135
19	Los Angeles	71.43	28.57	23.81	9.52	5.55E-05	0.83	135
20	Madera	66.67	33.33	0.00	9.52	3.59E-06	0.45	135
21	Marin	66.67	33.33	4.76	0.00	4.09E-07	0.53	117
22	Mariposa	71.43	28.57	0.00	4.76	1.50E-04	0.11	90
23	Mendocino	47.62	52.38	0.00	0.00	7.32E-04	0.76	135
24	Merced	66.67	33.33	4.76	0.00	9.96E-04	0.57	135
25	Modoc	57.14	42.86	0.00	4.76	1.91E-04	-0.02	135
26	Mono	47.62	52.38	0.00	4.76	2.51E-06	0.05	107
27	Monterey	57.14	42.86	0.00	4.76	1.17E-04	0.73	135
28	Napa	42.86	57.14	0.00	0.00	1.78E-05	0.51	117
29	Nevada	66.67	33.33	0.00	4.76	9.62E-05	0.61	135
30	Orange	61.90	38.10	4.76	0.00	2.12E-04	0.90	135
31	Placer	80.95	19.05	0.00	0.00	1.49E-06	0.58	135
32	Plumas	52.38	47.62	4.76	4.76	5.02E-05	0.52	135
33	Riverside	52.38	47.62	4.76	0.00	8.71E-06	0.74	117
34	Sacramento	66.67	33.33	4.76	0.00	2.14E-05	0.76	135
35	San Benito	76.19	23.81	4.76	4.76	7.83E-07	0.12	135
36	San Bernardino	52.38	47.62	0.00	4.76	6.41E-06	0.44	117
37	San Diego	61.90	38.10	0.00	0.00	1.51E-04	0.49	135
38	San Francisco	57.14	42.86	0.00	4.76	3.16E-07	0.45	117
39	San Joaquin	57.14	42.86	14.29	4.76	7.11E-05	0.86	135
40	San Luis Obispo	57.14	42.86	0.00	0.00	7.65E-06	0.68	135
41	San Mateo	42.86	57.14	0.00	0.00	1.94E-07	0.34	117
42	Santa Barbara	57.14	42.86	9.52	9.52	8.89E-04	0.83	135
43	Santa Clara	52.38	47.62	0.00	9.52	1.65E-05	0.71	135
44	Santa Cruz	57.14	42.86	9.52	0.00	3.05E-04	0.58	135
45	Shasta	71.43	28.57	4.76	0.00	1.44E-03	0.41	135
46	Sierra							
47	Siskiyou	66.67	33.33	4.76	4.76	7.79E-04	0.37	135
48	Solano	71.43	28.57	4.76	0.00	4.52E-06	0.31	117
49	Sonoma	57.14	42.86	4.76	4.76	2.96E-05	0.55	135
50	Stanislaus	61.90	38.10	4.76	4.76	1.74E-03	0.83	135
51	Sutter	57.14	42.86	0.00	0.00	4.51E-03	-0.19	45
52	Tehama	52.38	47.62	0.00	0.00	1.23E-05	0.52	135
53	Trinity	66.67	33.33	0.00	0.00	7.32E-05	0.05	117
54	Tulare	57.14	42.86	0.00	0.00	3.36E-05	0.34	135
55	Tuolumne	66.67	33.33	0.00	0.00	1.26E-05	0.26	135
56	Ventura	57.14	42.86	4.76	0.00	2.27E-05	0.43	135
57	Yolo	47.62	52.38	4.76	0.00	2.10E-06	0.51	135
58	Yuba	57.14	42.86	28.57	23.81	1.38E-02	0.88	45

Table A6.b: ARFG Results AcrossCounty - Within Coefficients (Continuous Aid)

	# Right	# Right & Significant	% Right	%Right & Significant
Continuous Benefits	56	7	100.00%	12.50%
Farm	29	1	51.79%	1.79%
Farm (t-1)	32	2	57.14%	3.57%
Farm (t-2)	24	2	42.86%	3.57%
Farm (t-3)	31	2	55.36%	3.57%
Service	24	2	42.86%	3.57%
Service (t-1)	27	2	48.21%	3.57%
Service (t-2)	27	1	48.21%	1.79%
Service (t-3)	35	1	62.50%	1.79%
Retail	30	1	53.57%	1.79%
Retail (t-1)	30	1	53.57%	1.79%
Retail (t-2)	33	0	58.93%	0.00%
Retail (t-3)	24	1	42.86%	1.79%
Other	0	0	0.00%	0.00%
Other (t-1)	23	1	41.07%	1.79%
Other (t-2)	35	0	62.50%	0.00%
Other (t-3)	32	4	57.14%	7.14%
CA Min Wage	37	5	66.07%	8.93%
Percent Hispanic	28	0	50.00%	0.00%
Percent Black	56	6	100.00%	10.71%
Birth Rate	56	7	100.00%	12.50%

Table A7.a: TRU Results Within County - Across Coefficients (Continuous Aid)

County Number	County Name	% Right	%Wrong	%Right & Significant	%Wrong & Significant	AMSFE	Adusted R-squared	Number of Observations
1	Alameda	57.14	42.86	14.29	9.52	1.22E-03	0.62	117
2	Alpine							
3	Amador	47.62	52.38	0.00	0.00	3.25E-02	0.15	135
4	Butte	61.90	38.10	4.76	4.76	6.81E-03	0.66	135
5	Calaveras	57.14	42.86	0.00	0.00	2.25E-02	0.40	117
6	Colusa	61.90	38.10	9.52	0.00	5.63E-01	0.61	135
7	Contra Costa	57.14	42.86	4.76	4.76	1.55E-03	0.60	117
8	Del Norte	47.62	52.38	4.76	0.00	4.39E-02	0.43	135
9	El Dorado	61.90	38.10	4.76	0.00	8.17E-03	0.55	135
10	Fresno	57.14	42.86	4.76	4.76	1.20E-01	0.42	135
11	Glenn	76.19	23.81	4.76	4.76	4.89E-01	0.38	135
12	Humboldt	52.38	47.62	9.52	0.00	3.11E-02	0.45	135
13	Imperial	66.67	33.33	9.52	9.52	1.29E-01	0.84	135
14	Inyo	57.14	42.86	9.52	0.00	7.81E-03	0.19	117
15	Kern	71.43	28.57	4.76	0.00	4.07E-02	0.83	135
16	Kings	66.67	33.33	4.76	0.00	3.87E-02	0.64	135
17	Lake	57.14	42.86	0.00	0.00	1.02E-02	0.59	135
18	Lassen	66.67	33.33	14.29	0.00	2.29E-02	0.43	135
19	Los Angeles	57.14	42.86	19.05	9.52	2.65E-01	0.85	135
20	Madera	71.43	28.57	14.29	4.76	2.14E-01	0.71	135
21	Marin	52.38	47.62	4.76	0.00	2.13E-03	0.36	117
22	Mariposa	66.67	33.33	4.76	14.29	8.90E-01	0.15	90
23	Mendocino	76.19	23.81	0.00	0.00	9.84E-02	0.58	135
24	Merced	47.62	52.38	14.29	0.00	1.11E-01	0.70	135
25	Modoc	57.14	42.86	9.52	4.76	3.07E-02	0.34	135
26	Mono	47.62	52.38	4.76	9.52	1.53E-01	0.33	107
27	Monterey	76.19	23.81	0.00	0.00	2.68E-01	0.74	135
28	Napa	66.67	33.33	0.00	4.76	2.25E-02	0.28	117
29	Nevada	71.43	28.57	0.00	0.00	1.88E-01	0.36	135
30	Orange	61.90	38.10	9.52	4.76	1.30E-02	0.52	135
31	Placer	71.43	28.57	0.00	0.00	3.67E-03	0.51	135
32	Plumas	61.90	38.10	0.00	4.76	2.98E-01	0.47	135
33	Riverside	52.38	47.62	9.52	19.05	1.47E-03	0.80	117
34	Sacramento	47.62	52.38	4.76	9.52	3.58E-02	0.75	135
35	San Benito	57.14	42.86	0.00	0.00	3.87E-02	0.42	135
36	San Bernardino	57.14	42.86	0.00	0.00	3.38E-04	0.85	117
37	San Diego	57.14	42.86	4.76	0.00	4.99E-02	0.90	135
38	San Francisco	71.43	28.57	4.76	4.76	1.03E-04	0.41	117
39	San Joaquin	57.14	42.86	9.52	0.00	1.25E-02	0.76	135
40	San Luis Obispo	66.67	33.33	9.52	4.76	1.33E-01	0.58	135
41	San Mateo	52.38	47.62	0.00	4.76	5.59E-03	0.59	117
42	Santa Barbara	71.43	28.57	0.00	9.52	2.06E-01	0.74	135
43	Santa Clara	66.67	33.33	14.29	0.00	4.89E-02	0.79	135
44	Santa Cruz	52.38	47.62	4.76	0.00	1.20E-01	0.70	135
45	Shasta	42.86	57.14	4.76	0.00	8.49E-02	0.53	135
46	Sierra							
47	Siskiyou	57.14	42.86	0.00	0.00	6.19E-02	0.35	135
48	Solano	52.38	47.62	4.76	4.76	1.90E-03	0.53	117
49	Sonoma	61.90	38.10	4.76	0.00	1.16E-02	0.60	135
50	Stanislaus	42.86	57.14	9.52	4.76	7.79E-02	0.89	135
51	Sutter	42.86	57.14	4.76	4.76	8.64E-02	0.72	45
52	Tehama	42.86	57.14	14.29	9.52	1.92E-01	0.62	135
53	Trinity	66.67	33.33	4.76	0.00	3.23E-02	0.27	117
54	Tulare	61.90	38.10	0.00	0.00	7.48E-03	0.72	135
55	Tuolumne	47.62	52.38	0.00	4.76	6.22E-02	0.52	135
56	Ventura	38.10	61.90	0.00	4.76	6.59E-03	0.85	135
57	Yolo	66.67	33.33	0.00	0.00	1.46E-02	0.57	135
58	Yuba	71.43	28.57	19.05	9.52	5.18E-01	0.82	45

Table A7.b: TRU Results AcrossCounty - Within Coefficients (Continuous Aid)

	# Right	# Right & Significant	% Right	%Right & Significant
Continuous Benefits	56	7	100.00%	12.50%
Farm	27	0	48.21%	0.00%
Farm (t-1)	28	1	50.00%	1.79%
Farm (t-2)	32	1	57.14%	1.79%
Farm (t-3)	28	1	50.00%	1.79%
Service	25	1	44.64%	1.79%
Service (t-1)	28	5	50.00%	8.93%
Service (t-2)	25	0	44.64%	0.00%
Service (t-3)	33	2	58.93%	3.57%
Retail	38	3	67.86%	5.36%
Retail (t-1)	24	1	42.86%	1.79%
Retail (t-2)	22	2	39.29%	3.57%
Retail (t-3)	33	4	58.93%	7.14%
Other	0	0	0.00%	0.00%
Other (t-1)	24	3	42.86%	5.36%
Other (t-2)	36	1	64.29%	1.79%
Other (t-3)	29	2	51.79%	3.57%
CA Min Wage	38	5	67.86%	8.93%
Percent Hispanic	28	6	50.00%	10.71%
Percent Black	56	9	100.00%	16.07%
Birth Rate	56	9	100.00%	16.07%

Table A8.a: TRFG Results Within County - Across Coefficients (Continuous Aid)

County Number	County Name	% Right	%Wrong	%Right & Significant	%Wrong & Significant	AMSFE	Adjusted R-squared	Number of Observations
1	Alameda	61.90	38.10	9.52	0.00	1.97E-04	0.54	117
2	Alpine							
3	Amador	52.38	47.62	0.00	0.00	2.45E-02	-0.05	135
4	Butte	61.90	38.10	0.00	4.76	1.36E-02	0.58	135
5	Calaveras	57.14	42.86	0.00	0.00	3.14E-03	0.10	117
6	Colusa	61.90	38.10	14.29	4.76	9.07E-03	0.54	135
7	Contra Costa	61.90	38.10	4.76	0.00	3.63E-04	0.29	117
8	Del Norte	66.67	33.33	19.05	4.76	1.54E-02	0.44	135
9	El Dorado	61.90	38.10	14.29	0.00	1.14E-02	0.45	135
10	Fresno	57.14	42.86	0.00	0.00	3.81E-03	0.38	135
11	Glenn	61.90	38.10	9.52	0.00	1.38E+00	0.59	135
12	Humboldt	61.90	38.10	4.76	0.00	1.65E-03	0.43	135
13	Imperial	61.90	38.10	0.00	4.76	7.83E-03	0.52	135
14	Inyo	71.43	28.57	4.76	0.00	6.26E-03	0.20	117
15	Kern	61.90	38.10	9.52	4.76	4.48E-03	0.87	135
16	Kings	66.67	33.33	0.00	0.00	5.44E-02	0.47	135
17	Lake	47.62	52.38	0.00	4.76	9.47E-03	0.32	135
18	Lassen	76.19	23.81	4.76	0.00	2.88E-03	0.46	135
19	Los Angeles	57.14	42.86	9.52	9.52	1.07E-01	0.73	135
20	Madera	76.19	23.81	4.76	0.00	9.70E-03	0.68	135
21	Marin	57.14	42.86	4.76	4.76	1.26E-03	0.42	117
22	Mariposa	57.14	42.86	0.00	0.00	3.28E-03	0.17	90
23	Mendocino	71.43	28.57	0.00	0.00	4.38E-04	0.74	135
24	Merced	66.67	33.33	4.76	0.00	1.10E-04	0.60	135
25	Modoc	61.90	38.10	4.76	9.52	1.04E-03	0.13	135
26	Mono	71.43	28.57	0.00	0.00	5.14E-02	-0.08	107
27	Monterey	61.90	38.10	0.00	0.00	2.88E-03	0.64	135
28	Napa	61.90	38.10	0.00	0.00	6.79E-04	0.52	117
29	Nevada	57.14	42.86	0.00	0.00	2.04E-03	0.51	135
30	Orange	66.67	33.33	0.00	9.52	6.00E-04	0.46	135
31	Placer	57.14	42.86	0.00	0.00	2.28E-03	0.42	135
32	Plumas	66.67	33.33	0.00	0.00	1.08E-02	0.51	135
33	Riverside	52.38	47.62	4.76	0.00	8.12E-05	0.47	117
34	Sacramento	42.86	57.14	0.00	0.00	3.15E-04	0.51	135
35	San Benito	61.90	38.10	0.00	4.76	4.47E-03	0.22	135
36	San Bernardino	57.14	42.86	0.00	4.76	3.59E-04	0.45	117
37	San Diego	57.14	42.86	4.76	0.00	5.89E-02	0.85	135
38	San Francisco	66.67	33.33	4.76	0.00	2.92E-05	0.46	117
39	San Joaquin	57.14	42.86	19.05	4.76	2.94E-03	0.90	135
40	San Luis Obispo	57.14	42.86	4.76	0.00	3.53E-02	0.38	135
41	San Mateo	66.67	33.33	0.00	4.76	2.51E-03	0.60	117
42	Santa Barbara	52.38	47.62	0.00	0.00	4.19E-03	0.50	135
43	Santa Clara	61.90	38.10	9.52	14.29	2.77E-02	0.71	135
44	Santa Cruz	61.90	38.10	0.00	0.00	1.45E-02	0.71	135
45	Shasta	52.38	47.62	4.76	4.76	6.02E-02	0.65	135
46	Sierra							
47	Siskiyou	66.67	33.33	9.52	0.00	2.19E-02	0.37	135
48	Solano	52.38	47.62	9.52	14.29	4.41E-04	0.33	117
49	Sonoma	61.90	38.10	23.81	0.00	1.91E-04	0.69	135
50	Stanislaus	47.62	52.38	9.52	4.76	1.51E-02	0.91	135
51	Sutter	71.43	28.57	0.00	0.00	7.42E-01	-0.04	45
52	Tehama	47.62	52.38	4.76	4.76	1.00E-02	0.61	135
53	Trinity	52.38	47.62	0.00	0.00	1.00E-02	0.24	117
54	Tulare	47.62	52.38	4.76	0.00	2.14E-03	0.42	135
55	Tuolumne	52.38	47.62	9.52	9.52	2.11E-02	0.45	135
56	Ventura	52.38	47.62	0.00	0.00	3.18E-02	0.45	135
57	Yolo	47.62	52.38	9.52	4.76	5.90E-03	0.56	135
58	Yuba	61.90	38.10	28.57	14.29	4.52E-02	0.93	45

Table A8.b: TRFG Results AcrossCounty - Within Coefficients (Continuous Aid)

	# Right	# Right & Significant	% Right	%Right & Significant
Continuous Benefits	56	6	100.00%	10.71%
Farm	34	3	60.71%	5.36%
Farm (t-1)	26	3	46.43%	5.36%
Farm (t-2)	37	3	66.07%	5.36%
Farm (t-3)	31	3	55.36%	5.36%
Service	29	1	51.79%	1.79%
Service (t-1)	28	2	50.00%	3.57%
Service (t-2)	29	2	51.79%	3.57%
Service (t-3)	31	3	55.36%	5.36%
Retail	33	1	58.93%	1.79%
Retail (t-1)	28	2	50.00%	3.57%
Retail (t-2)	30	2	53.57%	3.57%
Retail (t-3)	28	1	50.00%	1.79%
Other	0	0	0.00%	0.00%
Other (t-1)	29	1	51.79%	1.79%
Other (t-2)	28	0	50.00%	0.00%
Other (t-3)	32	3	57.14%	5.36%
CA Min Wage	26	2	46.43%	3.57%
Percent Hispanic	29	5	51.79%	8.93%
Percent Black	56	7	100.00%	12.50%
Birth Rate	56	9	100.00%	16.07%

**Table A9.a: Panel Data Models - Aggregated Employment Coefficients (Homoskedastic)
(Varying Estimation Lengths - Estimation Period Ends 1996)**

	Start Year	Method 1	Method 2				Method 3			
		Total Employment	Farm	Retail	Service	Other	Current (t)	Lag 1 (t-1)	Lag 2 (t-2)	Lag 3 (t-3)
ARU	1985	Correct	Correct		Correct	Correct	Correct	Correct		
	1986	Correct	Correct		Correct	Correct	Correct	Correct		
	1987		Correct		Correct	Correct	Correct	Correct		
	1988	Correct	Correct		Correct	Correct	Correct	Correct		
	1989	Correct	Correct		Correct	Correct	Correct	Correct		
	1990	Correct	Correct		Correct	Correct	Correct	Correct		
	1991	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	
	1992	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
	1993	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
	1994	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	
	1995	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
	1996	Correct	Correct	Correct	Correct	Correct	Correct		Correct	Correct
ARFG	1985				Correct	Correct	Correct			
	1986				Correct	Correct	Correct			
	1987				Correct	Correct	Correct			
	1988				Correct		Correct			
	1989				Correct		Correct			
	1990				Correct		Correct			
	1991				Correct	Correct	Correct	Correct		
	1992			Correct	Correct			Correct		
	1993			Correct			Correct	Correct		
	1994			Correct				Correct		
	1995					Correct		Correct		
	1996	Correct		Correct	Correct		Correct	Correct		Correct
TRU	1985	Correct	Correct	Correct	Correct		Correct	Correct		
	1986	Correct	Correct	Correct	Correct		Correct	Correct		
	1987	Correct	Correct	Correct	Correct		Correct	Correct		
	1988	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	
	1989	Correct	Correct		Correct	Correct	Correct	Correct	Correct	Correct
	1990	Correct	Correct		Correct	Correct	Correct	Correct	Correct	
	1991	Correct	Correct	Correct	Correct	Correct	Correct		Correct	
	1992	Correct	Correct		Correct	Correct	Correct	Correct	Correct	Correct
	1993	Correct	Correct		Correct	Correct	Correct	Correct		
	1994	Correct	Correct		Correct	Correct	Correct	Correct		
	1995	Correct	Correct		Correct	Correct	Correct	Correct		
	1996		Correct		Correct	Correct		Correct		
TRFG	1985	Correct	Correct	Correct		Correct	Correct			Correct
	1986	Correct	Correct	Correct		Correct	Correct			Correct
	1987	Correct	Correct	Correct		Correct	Correct			Correct
	1988	Correct	Correct	Correct	Correct	Correct	Correct			Correct
	1989	Correct	Correct	Correct	Correct	Correct	Correct			Correct
	1990	Correct	Correct	Correct	Correct	Correct	Correct	Correct		Correct
	1991	Correct	Correct	Correct	Correct	Correct	Correct			Correct
	1992	Correct	Correct		Correct	Correct	Correct	Correct	Correct	Correct
	1993	Correct	Correct		Correct	Correct	Correct	Correct	Correct	
	1994	Correct	Correct		Correct	Correct	Correct	Correct		Correct
	1995	Correct	Correct	Correct	Correct	Correct	Correct			Correct
	1996		Correct		Correct	Correct	Correct		Correct	Correct

**Table A9.b: Panel Data Models - Aggregated Employment Coefficients (Heteroskedastic)
(Varying Estimation Lengths - Estimation Period Ends 1996)**

	Start Year	Method 1	Method 2				Method 3			
		Total Employment	Farm	Retail	Service	Other	Current (t)	Lag 1 (t-1)	Lag 2 (t-2)	Lag 3 (t-3)
ARU	1985	Correct	Correct		Correct	Correct	Correct	Correct		
	1986	Correct	Correct		Correct	Correct	Correct	Correct		
	1987	Correct	Correct		Correct	Correct	Correct	Correct		
	1988	Correct	Correct		Correct	Correct	Correct	Correct		
	1989	Correct	Correct		Correct	Correct	Correct	Correct		
	1990	Correct	Correct	Correct	Correct	Correct	Correct	Correct		
	1991	Correct	Correct	Correct	Correct	Correct	Correct	Correct		
	1992	Correct	Correct	Correct	Correct	Correct	Correct	Correct		
	1993	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	
	1994	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	
	1995	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
	1996	Correct	Correct	Correct	Correct	Correct	Correct		Correct	Correct
ARFG	1985				Correct		Correct	Correct		
	1986				Correct		Correct	Correct		
	1987			Correct	Correct		Correct	Correct		
	1988			Correct	Correct		Correct	Correct		
	1989			Correct	Correct		Correct	Correct		
	1990			Correct	Correct		Correct	Correct		
	1991				Correct			Correct		
	1992				Correct			Correct		
	1993	Correct			Correct	Correct		Correct		
	1994	Correct			Correct	Correct		Correct		
	1995				Correct	Correct		Correct		
	1996	Correct		Correct	Correct		Correct	Correct		
TRU	1985	Correct	Correct	Correct	Correct		Correct	Correct		Correct
	1986	Correct	Correct	Correct	Correct		Correct	Correct		Correct
	1987	Correct	Correct	Correct	Correct		Correct	Correct		Correct
	1988	Correct	Correct	Correct	Correct		Correct	Correct		Correct
	1989	Correct	Correct		Correct		Correct	Correct		
	1990	Correct	Correct	Correct	Correct	Correct	Correct	Correct		Correct
	1991	Correct	Correct	Correct	Correct	Correct	Correct	Correct		
	1992	Correct	Correct		Correct	Correct	Correct	Correct	Correct	
	1993	Correct	Correct	Correct	Correct	Correct	Correct	Correct		
	1994	Correct	Correct		Correct	Correct	Correct	Correct		Correct
	1995	Correct	Correct	Correct	Correct		Correct	Correct		Correct
	1996	Correct	Correct		Correct	Correct	Correct	Correct		
TRFG	1985	Correct	Correct	Correct		Correct				Correct
	1986	Correct	Correct	Correct		Correct		Correct	Correct	
	1987	Correct	Correct	Correct		Correct		Correct	Correct	
	1988	Correct	Correct	Correct		Correct			Correct	
	1989	Correct	Correct		Correct	Correct			Correct	
	1990	Correct	Correct	Correct	Correct	Correct			Correct	
	1991	Correct	Correct	Correct	Correct	Correct			Correct	
	1992	Correct	Correct		Correct	Correct			Correct	
	1993	Correct	Correct		Correct	Correct			Correct	
	1994	Correct	Correct	Correct	Correct	Correct		Correct		Correct
	1995	Correct	Correct	Correct	Correct	Correct		Correct		Correct
	1996		Correct		Correct	Correct	Correct			Correct

Figure A1.a: Variance of AFDC-U Accession Rate by County
November Only

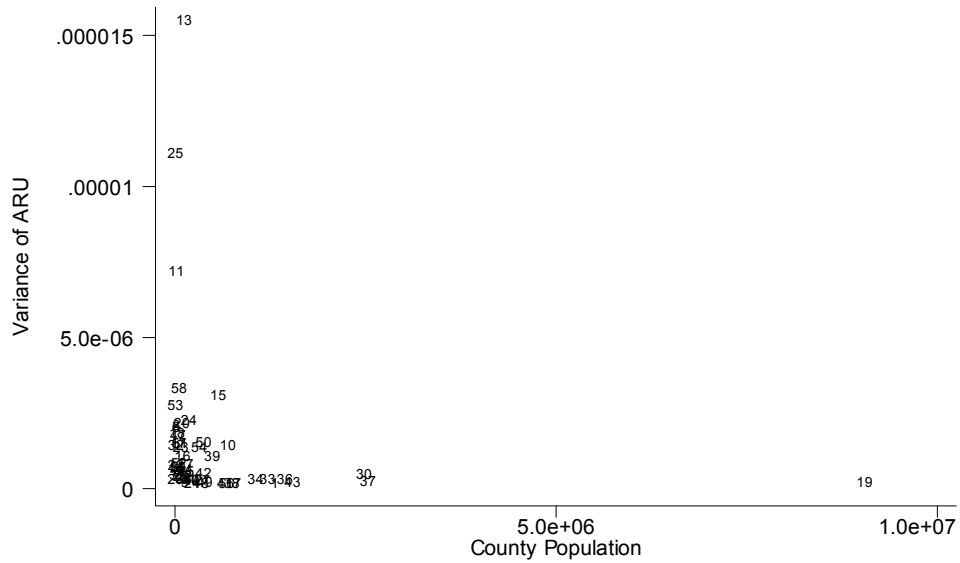


Figure A1.b: Variance of AFDC-FG Accession Rate by County
November Only

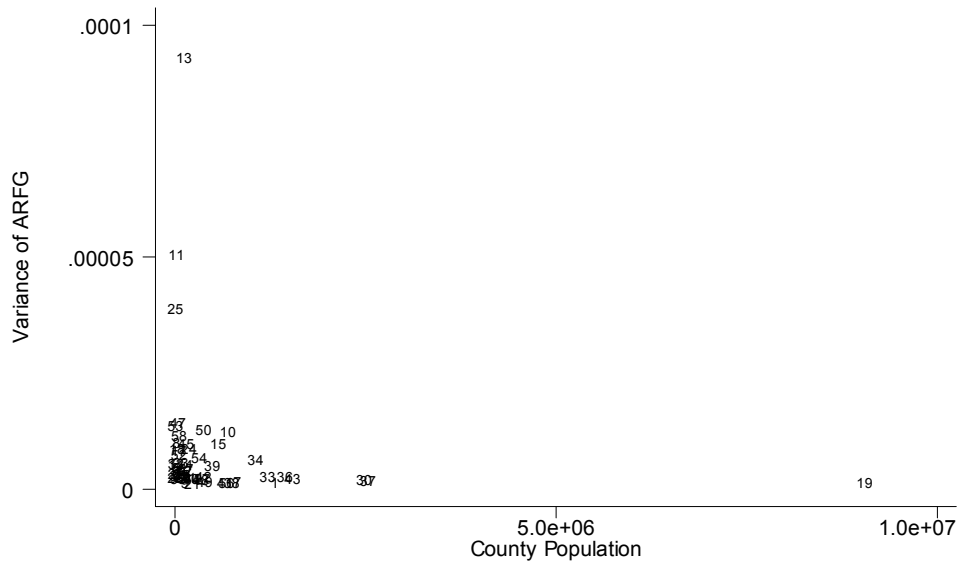


Figure A1.c: Variance of AFDC-U Termination Rate by County
November Only

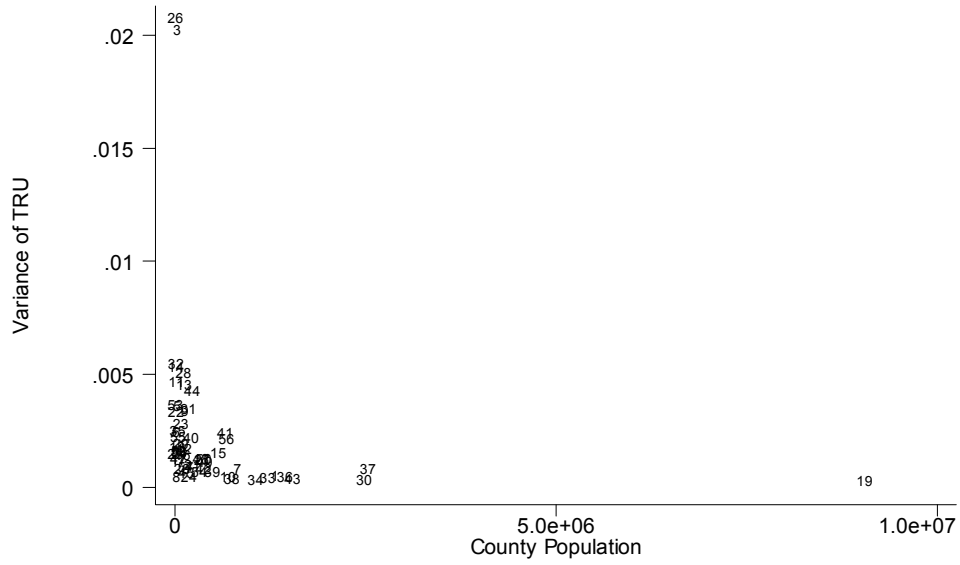


Figure A1.d: Variance of AFDC-FG Termination Rate by County
November Only

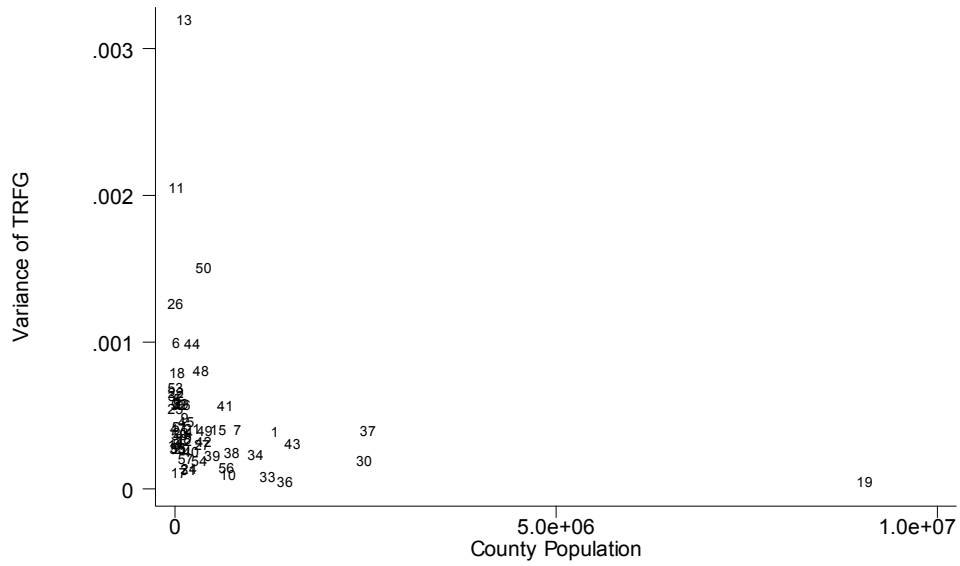


Figure A2.a: Variance of AFDC-U Accession Rate by County
November Only excluding LA County

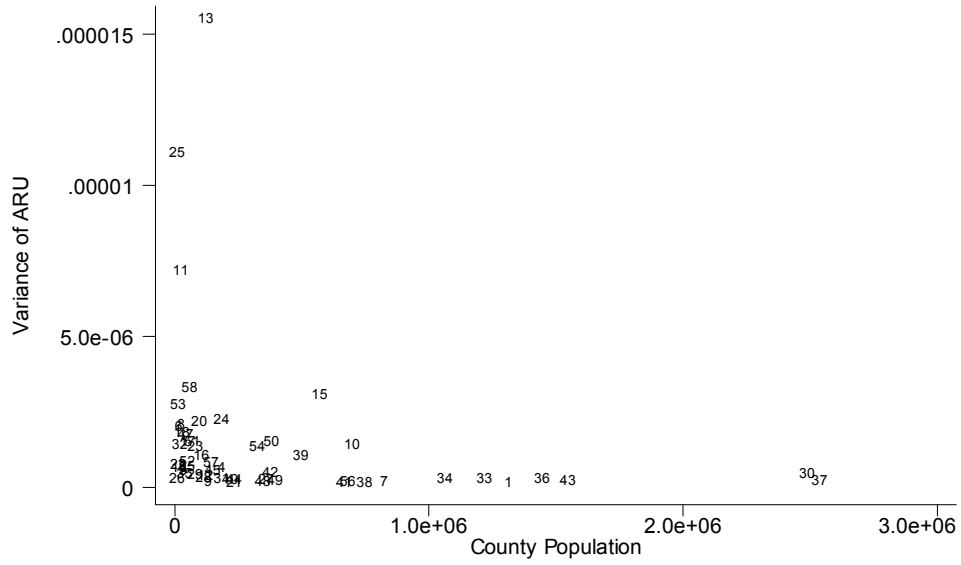


Figure A2.b: Variance of AFDC-FG Accession Rate by County
November Only excluding LA County

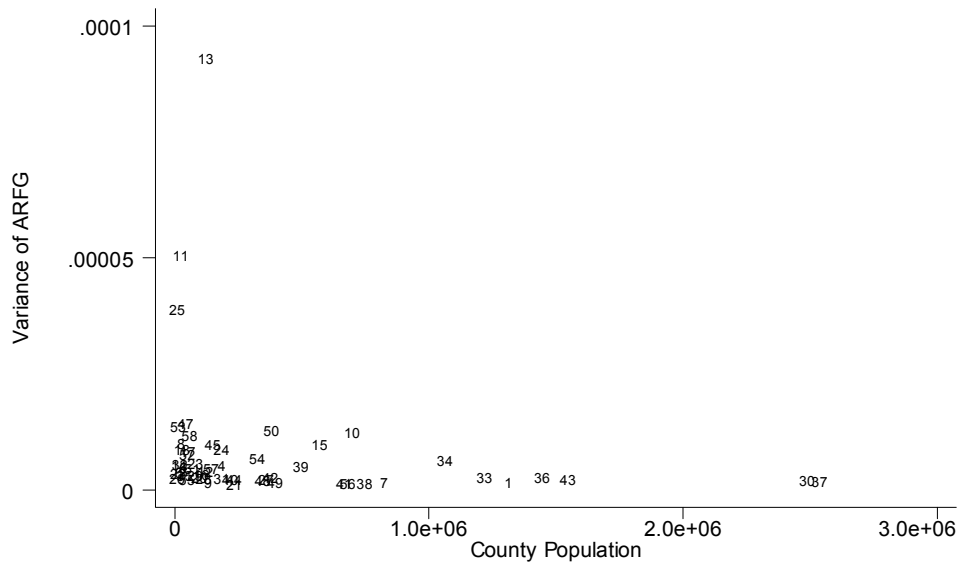


Figure A2.c: Variance of AFDC-U Termination Rate by County
November Only excluding LA County

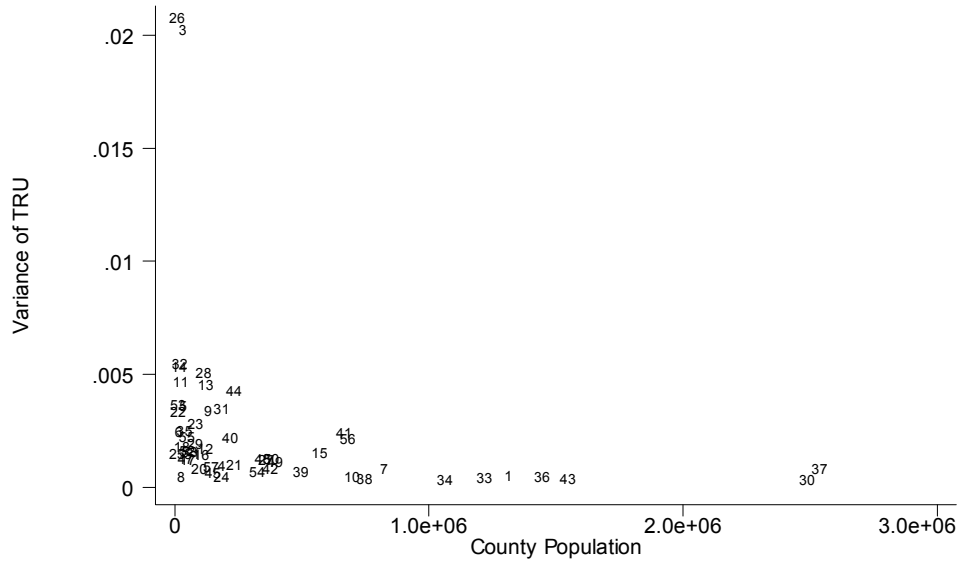


Figure A2.d: Variance of AFDC-FG Termination Rate by County
November Only excluding LA County

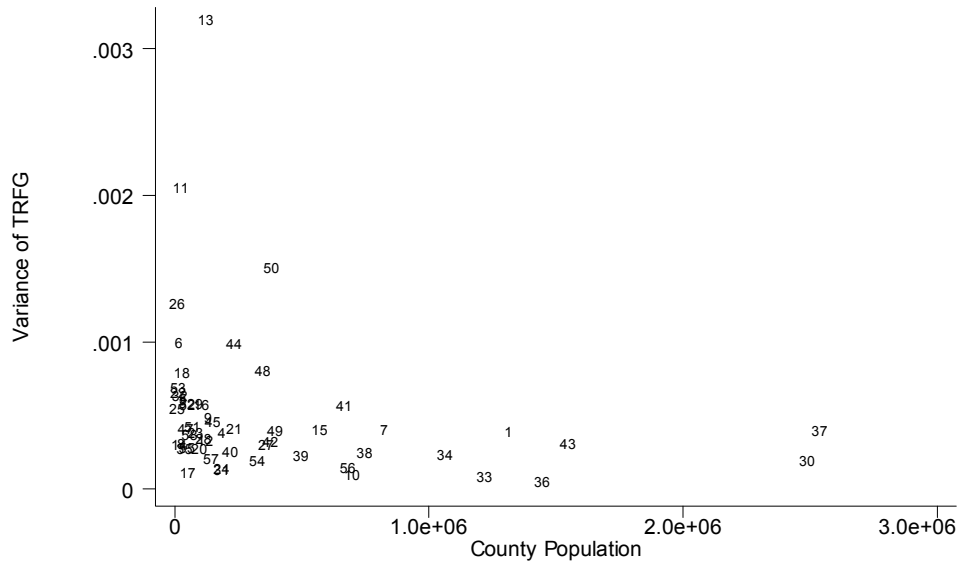


Figure A3.a: Variance of AFDC-U Accession Rate Across Counties
November Only

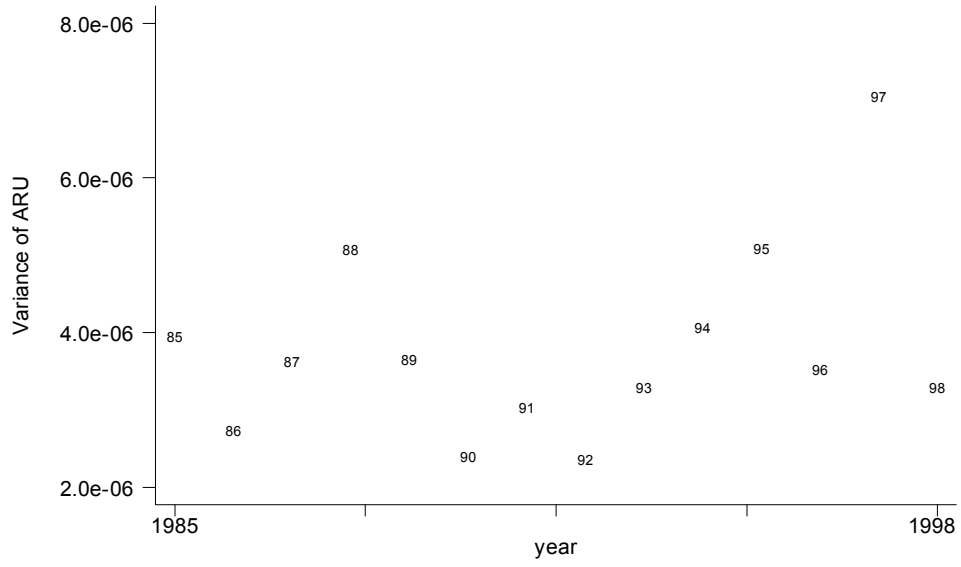


Figure A3.b: Variance of AFDC-FG Accession Rate Across Counties
November Only

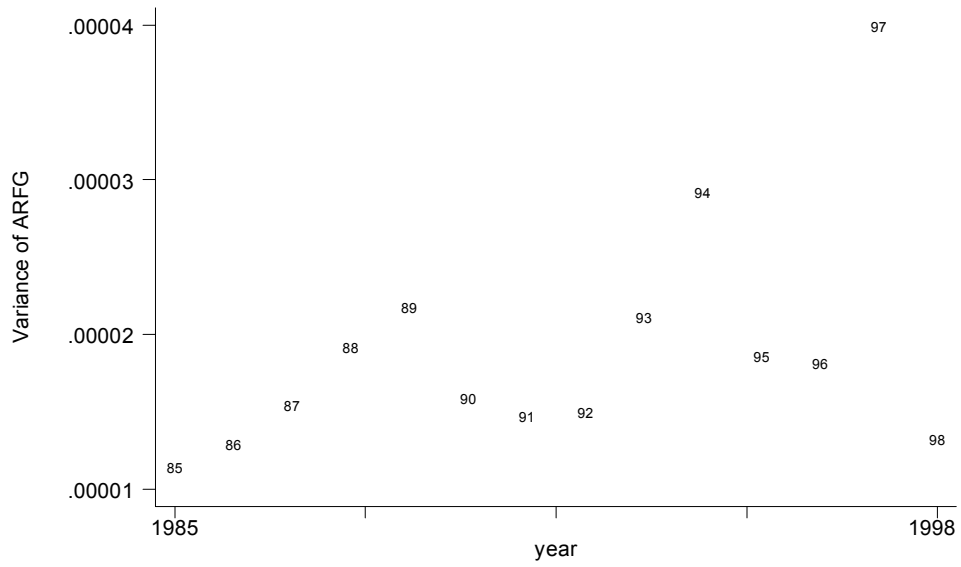


Figure A3.c: Variance of AFDC-U Termination Rate Across Counties
November Only

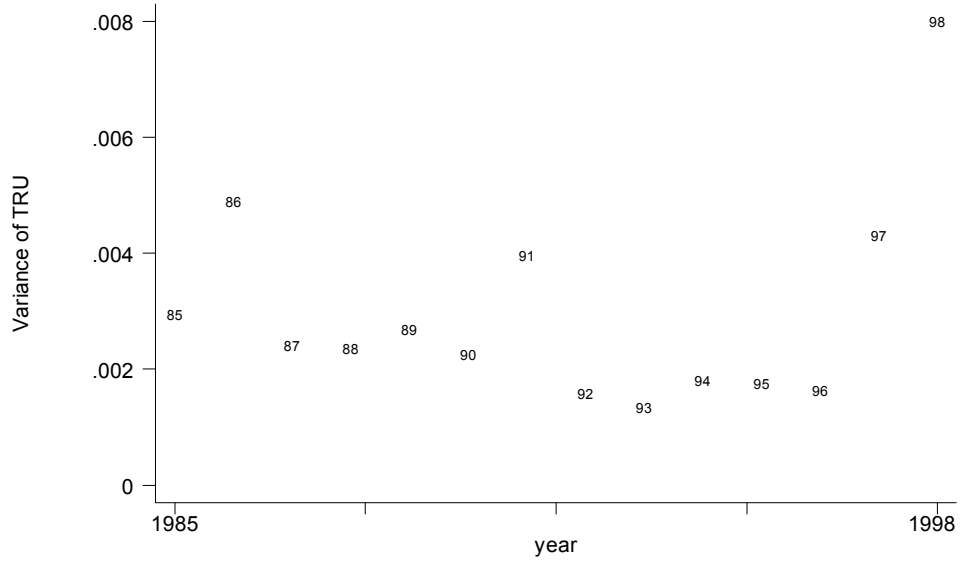


Figure A3.d: Variance of AFDC-FG Termination Rate Across Counties
November Only

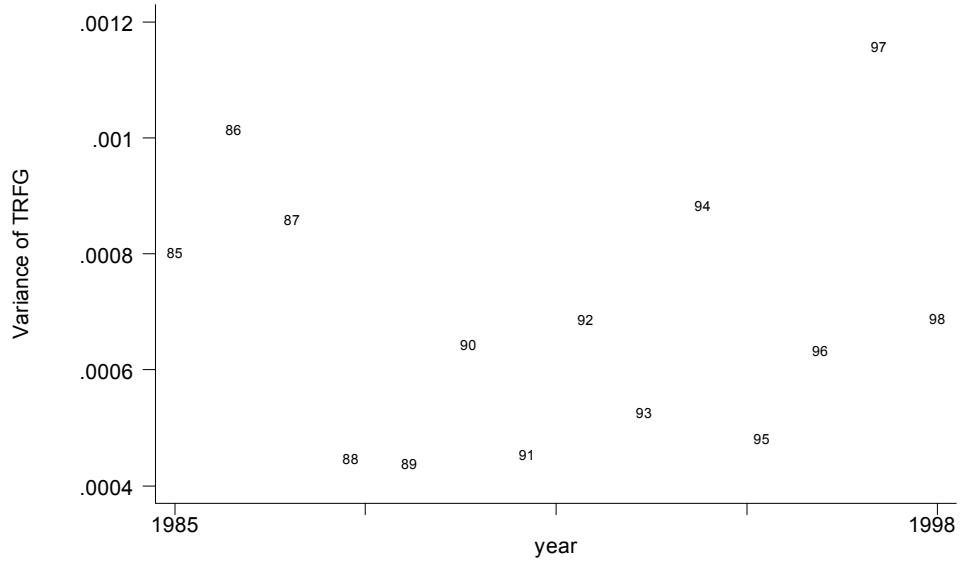


Table A10.a: AMSFE by Estimation Length and First Year of Forecast: AFDC-U Accession Rate

Length	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
2	5.79E-06	5.96E-06	4.08E-06	4.28E-06	3.58E-06	5.46E-06	7.18E-06	8.67E-06	2.04E-05	1.72E-05
3		5.05E-06	4.55E-06	3.59E-06	5.50E-06	1.23E-05	8.33E-06	7.77E-06	1.11E-05	1.73E-05
4			4.70E-06	3.95E-06	4.02E-06	1.62E-05	2.04E-05	7.80E-06	1.14E-05	1.66E-05
5				3.95E-06	4.19E-06	5.38E-06	8.30E-06	8.33E-06	1.11E-05	1.64E-05
6					4.75E-06	5.54E-06	8.11E-06	1.87E-05	1.18E-05	1.67E-05
7						5.65E-06	7.25E-06	8.04E-06	1.12E-05	1.65E-05
8							7.42E-06	1.05E-05	1.12E-05	1.68E-05
9								1.29E-05	1.22E-05	1.67E-05
10									1.47E-05	1.65E-05
11										2.03E-05

Table A10.b: AMSFE by Estimation Length and First Year of Forecast: AFDC-FG Accession Rate

Length	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
2	1.45E-06	9.04E-07	1.01E-06	1.18E-06	1.22E-06	1.14E-06	1.53E-06	2.06E-06	2.34E-06	2.44E-06
3		1.21E-06	1.06E-06	1.15E-06	1.24E-06	1.27E-06	1.70E-06	1.93E-06	2.08E-06	2.37E-06
4			1.06E-06	1.06E-06	1.33E-06	1.22E-06	1.64E-06	2.17E-06	2.11E-06	2.47E-06
5				1.16E-06	1.20E-06	1.44E-06	1.47E-06	2.07E-06	2.23E-06	2.16E-06
6					1.07E-06	1.51E-06	1.50E-06	2.00E-06	2.40E-06	2.37E-06
7						1.35E-06	1.50E-06	2.07E-06	2.17E-06	3.07E-06
8							1.50E-06	2.09E-06	2.33E-06	2.29E-06
9								2.09E-06	2.23E-06	2.14E-06
10									2.23E-06	2.22E-06
11										2.20E-06

Table A10.c: AMSFE by Estimation Length and First Year of Forecast: AFDC-U Termination Rate

Length	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
2	3.89E-04	2.91E-04	2.28E-04	4.24E-04	2.67E-04	6.07E-04	3.92E-04	4.71E-04	5.59E-04	1.05E-03
3		3.24E-04	3.01E-04	2.48E-04	3.03E-04	3.40E-04	5.38E-04	4.33E-04	7.00E-04	8.58E-04
4			2.64E-04	2.48E-04	2.92E-04	5.61E-04	4.73E-04	5.55E-04	6.39E-04	9.84E-04
5				2.30E-04	2.75E-04	4.76E-04	3.94E-04	4.79E-04	7.45E-04	8.03E-04
6					2.31E-04	2.90E-04	4.86E-04	4.41E-04	5.72E-04	9.48E-04
7						2.81E-04	4.10E-04	5.02E-04	6.23E-04	9.15E-04
8							4.15E-04	4.71E-04	7.00E-04	9.53E-04
9								4.70E-04	6.41E-04	8.31E-04
10									7.23E-04	1.14E-03
11										1.07E-03

Table A10.d: AMSFE by Estimation Length and First Year of Forecast: AFDC-FG Termination Rate

Length	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
2	4.20E-03	1.97E-03	1.80E-03	4.15E-03	2.61E-03	4.62E-03	1.61E-03	2.16E-03	2.10E-03	3.58E-03
3		2.98E-03	1.85E-03	1.72E-03	3.77E-03	2.97E-03	2.53E-03	2.11E-03	3.58E-03	3.72E-03
4			1.82E-03	1.82E-03	3.60E-03	4.06E-03	4.02E-03	3.06E-03	3.75E-03	3.95E-03
5				1.79E-03	3.26E-03	4.67E-03	2.00E-03	9.30E-03	5.72E-03	4.82E-03
6					2.82E-03	4.53E-03	1.66E-03	2.23E-03	2.43E-03	7.94E-03
7						2.70E-03	1.68E-03	2.03E-03	5.57E-03	8.11E-03
8							3.16E-03	2.84E-03	2.64E-03	4.06E-03
9								2.89E-03	2.84E-03	5.27E-03
10									3.11E-03	4.07E-03
11										4.15E-03

Table A11: Minimum Required Change Required to Identify Impact

County Number	County Name	ARU	ARFG	TRU	TRFG
1	Alameda	59.91	31.98	92.08	48.56
2	Alpine	N/A	N/A	N/A	N/A
3	Amador	160.18	77.23	138.28	88.05
4	Butte	62.31	69.30	84.51	75.28
5	Calaveras	59.77	38.88	89.57	77.04
6	Colusa	111.78	63.13	96.42	72.02
7	Contra Costa	82.89	28.86	33.40	33.73
8	Del Norte	65.80	45.59	79.89	69.82
9	El Dorado	55.22	58.71	75.63	62.38
10	Fresno	33.12	52.11	39.08	16.45
11	Glenn	57.27	36.20	70.65	61.32
12	Humboldt	84.57	79.05	120.10	88.47
13	Imperial	170.96	206.59	152.05	170.68
14	Inyo	55.62	56.05	97.19	65.94
15	Kern	113.16	79.00	54.38	80.14
16	Kings	61.30	47.21	76.84	74.75
17	Lake	145.48	180.28	122.48	174.53
18	Lassen	39.02	43.64	90.67	66.07
19	Los Angeles	34.66	43.66	34.64	17.85
20	Madera	50.67	29.44	70.11	55.83
21	Marin	79.74	32.34	90.99	72.44
22	Mariposa	85.50	62.30	110.16	89.23
23	Mendocino	62.96	50.34	95.05	75.78
24	Merced	73.96	48.82	56.90	28.77
25	Modoc	154.12	141.91	109.61	64.68
26	Mono	156.83	95.85	156.35	100.49
27	Monterey	98.36	73.26	110.70	97.06
28	Napa	87.81	72.92	91.85	52.06
29	Nevada	106.75	57.44	97.02	42.97
30	Orange	133.02	42.94	34.81	21.64
31	Placer	131.34	98.45	39.77	22.52
32	Plumas	66.03	65.46	112.93	74.52
33	Riverside	60.12	39.33	23.96	17.81
34	Sacramento	39.18	65.85	36.93	52.57
35	San Benito	70.61	51.00	73.48	78.06
36	San Bernardino	130.93	94.45	28.27	28.97
37	San Diego	101.38	90.96	61.49	30.92
38	San Francisco	85.14	48.52	43.48	18.14
39	San Joaquin	66.25	59.82	92.16	75.66
40	San Luis Obispo	55.93	66.60	47.85	37.33
41	San Mateo	208.84	91.18	52.85	34.43
42	Santa Barbara	62.78	23.26	37.63	26.95
43	Santa Clara	100.28	76.81	25.36	41.08
44	Santa Cruz	61.53	67.02	52.39	31.21
45	Shasta	60.53	40.61	68.01	48.97
46	Sierra	N/A	N/A	N/A	N/A
47	Siskiyou	104.26	100.53	56.64	41.13
48	Solano	41.14	37.38	46.00	44.74
49	Sonoma	95.37	38.53	47.10	42.02
50	Stanislaus	162.72	137.03	28.10	70.52
51	Sutter	117.48	131.01	99.92	108.42
52	Tehama	74.51	67.07	94.45	84.08
53	Trinity	111.74	54.51	98.07	51.21
54	Tulare	60.72	52.53	32.74	18.81
55	Tuolumne	72.26	66.23	70.99	84.76
56	Ventura	121.90	43.49	45.35	52.88
57	Yolo	33.67	63.92	35.57	26.39
58	Yuba	56.46	51.56	75.64	64.44

Table A12.a: Variance-Bias Tradeoff - AFDC-U Accessions

County Number	County Name	Variance	Bias	AMSFE
1	Alameda	2.94E-01	5.85E-02	3.00E-01
2	Alpine			
3	Amador	7.16E-01	3.58E-01	8.01E-01
4	Butte	1.28E-01	2.84E-01	3.12E-01
5	Calaveras	2.83E-01	9.73E-02	2.99E-01
6	Colusa	4.58E-01	3.20E-01	5.59E-01
7	Contra Costa	1.49E-01	3.87E-01	4.14E-01
8	Del Norte	3.29E-01	1.27E-02	3.29E-01
9	El Dorado	2.74E-01	3.70E-02	2.76E-01
10	Fresno	1.55E-01	5.83E-02	1.66E-01
11	Glenn	2.47E-01	1.45E-01	2.86E-01
12	Humboldt	3.50E-01	2.37E-01	4.23E-01
13	Imperial	8.02E-01	2.95E-01	8.55E-01
14	Inyo	2.73E-01	5.20E-02	2.78E-01
15	Kern	1.23E-01	5.52E-01	5.66E-01
16	Kings	1.97E-01	2.34E-01	3.07E-01
17	Lake	7.16E-01	1.31E-01	7.27E-01
18	Lassen	1.61E-01	1.10E-01	1.95E-01
19	Los Angeles	1.37E-01	1.06E-01	1.73E-01
20	Madera	1.23E-01	2.22E-01	2.53E-01
21	Marin	2.73E-01	2.91E-01	3.99E-01
22	Mariposa	3.59E-01	2.32E-01	4.28E-01
23	Mendocino	1.19E-01	2.91E-01	3.15E-01
24	Merced	3.70E-01	2.20E-03	3.70E-01
25	Modoc	7.69E-01	5.23E-02	7.71E-01
26	Mono	6.87E-01	3.79E-01	7.84E-01
27	Monterey	4.05E-01	2.79E-01	4.92E-01
28	Napa	3.57E-01	2.56E-01	4.39E-01
29	Nevada	3.39E-01	4.12E-01	5.34E-01
30	Orange	1.32E-01	6.52E-01	6.65E-01
31	Placer	2.53E-01	6.06E-01	6.57E-01
32	Plumas	2.11E-01	2.54E-01	3.30E-01
33	Riverside	1.27E-01	2.73E-01	3.01E-01
34	Sacramento	1.80E-01	7.79E-02	1.96E-01
35	San Benito	3.26E-01	1.36E-01	3.53E-01
36	San Bernardino	9.00E-02	6.48E-01	6.55E-01
37	San Diego	1.14E-01	4.94E-01	5.07E-01
38	San Francisco	1.92E-01	3.80E-01	4.26E-01
39	San Joaquin	1.72E-01	2.83E-01	3.31E-01
40	San Luis Obispo	2.70E-01	7.34E-02	2.80E-01
41	San Mateo	4.88E-01	9.23E-01	1.04E+00
42	Santa Barbara	1.86E-01	2.53E-01	3.14E-01
43	Santa Clara	1.51E-01	4.78E-01	5.01E-01
44	Santa Cruz	2.16E-01	2.19E-01	3.08E-01
45	Shasta	1.58E-01	2.58E-01	3.03E-01
46	Sierra			
47	Siskiyou	2.60E-01	4.52E-01	5.21E-01
48	Solano	2.06E-01	5.86E-03	2.06E-01
49	Sonoma	1.94E-01	4.35E-01	4.77E-01
50	Stanislaus	2.09E-01	7.86E-01	8.14E-01
51	Sutter	5.58E-01	1.85E-01	5.87E-01
52	Tehama	1.91E-01	3.20E-01	3.73E-01
53	Trinity	5.18E-01	2.10E-01	5.59E-01
54	Tulare	1.79E-01	2.45E-01	3.04E-01
55	Tuolumne	2.94E-01	2.10E-01	3.61E-01
56	Ventura	1.60E-01	5.88E-01	6.10E-01
57	Yolo	1.48E-01	8.09E-02	1.68E-01
58	Yuba	1.82E-01	2.16E-01	2.82E-01

Table A12.b: Variance-Bias Tradeoff - AFDC-FG Accessions

County Number	County Name	Variance	Bias	AMSFE
1	Alameda	1.59E-01	1.59E-02	1.60E-01
2	Alpine			
3	Amador	3.09E-01	2.31E-01	3.86E-01
4	Butte	1.28E-01	3.22E-01	3.47E-01
5	Calaveras	1.67E-01	9.93E-02	1.94E-01
6	Colusa	2.66E-01	1.70E-01	3.16E-01
7	Contra Costa	1.38E-01	4.06E-02	1.44E-01
8	Del Norte	2.14E-01	7.81E-02	2.28E-01
9	El Dorado	2.06E-01	2.09E-01	2.94E-01
10	Fresno	1.75E-01	1.93E-01	2.61E-01
11	Glenn	1.73E-01	5.39E-02	1.81E-01
12	Humboldt	3.50E-01	1.83E-01	3.95E-01
13	Imperial	9.40E-01	4.29E-01	1.03E+00
14	Inyo	2.21E-01	1.72E-01	2.80E-01
15	Kern	9.47E-02	3.83E-01	3.95E-01
16	Kings	1.68E-01	1.65E-01	2.36E-01
17	Lake	9.00E-01	4.09E-02	9.01E-01
18	Lassen	1.79E-01	1.25E-01	2.18E-01
19	Los Angeles	8.32E-02	2.02E-01	2.18E-01
20	Madera	1.03E-01	1.06E-01	1.47E-01
21	Marin	1.29E-01	9.78E-02	1.62E-01
22	Mariposa	2.86E-01	1.22E-01	3.11E-01
23	Mendocino	1.32E-01	2.14E-01	2.52E-01
24	Merced	2.44E-01	2.95E-03	2.44E-01
25	Modoc	7.07E-01	6.24E-02	7.10E-01
26	Mono	3.26E-01	3.51E-01	4.79E-01
27	Monterey	2.85E-01	2.30E-01	3.66E-01
28	Napa	2.20E-01	2.91E-01	3.65E-01
29	Nevada	2.33E-01	1.68E-01	2.87E-01
30	Orange	1.18E-01	1.79E-01	2.15E-01
31	Placer	1.74E-01	4.60E-01	4.92E-01
32	Plumas	2.07E-01	2.53E-01	3.27E-01
33	Riverside	1.05E-01	1.66E-01	1.97E-01
34	Sacramento	2.58E-01	2.04E-01	3.29E-01
35	San Benito	1.76E-01	1.84E-01	2.55E-01
36	San Bernardino	1.12E-01	4.59E-01	4.72E-01
37	San Diego	1.02E-01	4.43E-01	4.55E-01
38	San Francisco	1.28E-01	2.06E-01	2.43E-01
39	San Joaquin	1.06E-01	2.80E-01	2.99E-01
40	San Luis Obispo	1.92E-01	2.72E-01	3.33E-01
41	San Mateo	2.19E-01	4.00E-01	4.56E-01
42	Santa Barbara	1.07E-01	4.48E-02	1.16E-01
43	Santa Clara	1.35E-01	3.60E-01	3.84E-01
44	Santa Cruz	2.81E-01	1.83E-01	3.35E-01
45	Shasta	1.34E-01	1.53E-01	2.03E-01
46	Sierra			
47	Siskiyou	2.03E-01	4.60E-01	5.03E-01
48	Solano	1.49E-01	1.12E-01	1.87E-01
49	Sonoma	1.44E-01	1.28E-01	1.93E-01
50	Stanislaus	1.72E-01	6.63E-01	6.85E-01
51	Sutter	6.21E-01	2.08E-01	6.55E-01
52	Tehama	1.45E-01	3.02E-01	3.35E-01
53	Trinity	2.71E-01	2.47E-02	2.73E-01
54	Tulare	1.47E-01	2.18E-01	2.63E-01
55	Tuolumne	1.86E-01	2.74E-01	3.31E-01
56	Ventura	2.13E-01	4.30E-02	2.17E-01
57	Yolo	1.58E-01	2.78E-01	3.20E-01
58	Yuba	1.81E-01	1.83E-01	2.58E-01

Table A12.c: Variance-Bias Tradeoff - AFDC-U Terminations

County Number	County Name	Variance	Bias	AMSFE
1	Alameda	2.76E-01	3.69E-01	4.60E-01
2	Alpine			
3	Amador	5.05E-01	4.72E-01	6.91E-01
4	Butte	1.53E-01	3.94E-01	4.23E-01
5	Calaveras	3.72E-01	2.50E-01	4.48E-01
6	Colusa	4.02E-01	2.66E-01	4.82E-01
7	Contra Costa	1.56E-01	5.98E-02	1.67E-01
8	Del Norte	3.03E-01	2.61E-01	3.99E-01
9	El Dorado	2.94E-01	2.38E-01	3.78E-01
10	Fresno	1.15E-01	1.58E-01	1.95E-01
11	Glenn	3.19E-01	1.52E-01	3.53E-01
12	Humboldt	3.08E-01	5.15E-01	6.01E-01
13	Imperial	5.99E-01	4.68E-01	7.60E-01
14	Inyo	3.75E-01	3.10E-01	4.86E-01
15	Kern	1.03E-01	2.52E-01	2.72E-01
16	Kings	2.03E-01	3.26E-01	3.84E-01
17	Lake	6.08E-01	7.19E-02	6.12E-01
18	Lassen	2.29E-01	3.91E-01	4.53E-01
19	Los Angeles	1.07E-01	1.36E-01	1.73E-01
20	Madera	2.27E-01	2.67E-01	3.51E-01
21	Marin	2.50E-01	3.80E-01	4.55E-01
22	Mariposa	3.90E-01	3.89E-01	5.51E-01
23	Mendocino	1.78E-01	4.41E-01	4.75E-01
24	Merced	1.74E-01	2.25E-01	2.85E-01
25	Modoc	5.48E-01	1.42E-02	5.48E-01
26	Mono	6.68E-01	4.06E-01	7.82E-01
27	Monterey	3.54E-01	4.26E-01	5.53E-01
28	Napa	2.00E-01	4.13E-01	4.59E-01
29	Nevada	3.33E-01	3.52E-01	4.85E-01
30	Orange	1.31E-01	1.15E-01	1.74E-01
31	Placer	1.87E-01	6.65E-02	1.99E-01
32	Plumas	3.37E-01	4.53E-01	5.65E-01
33	Riverside	1.11E-01	4.61E-02	1.20E-01
34	Sacramento	1.81E-01	3.78E-02	1.85E-01
35	San Benito	3.60E-01	7.56E-02	3.67E-01
36	San Bernardino	1.40E-01	1.61E-02	1.41E-01
37	San Diego	1.43E-01	2.72E-01	3.07E-01
38	San Francisco	1.41E-01	1.66E-01	2.17E-01
39	San Joaquin	1.16E-01	4.46E-01	4.61E-01
40	San Luis Obispo	2.39E-01	1.57E-02	2.39E-01
41	San Mateo	2.53E-01	7.51E-02	2.64E-01
42	Santa Barbara	1.88E-01	1.30E-02	1.88E-01
43	Santa Clara	1.23E-01	3.05E-02	1.27E-01
44	Santa Cruz	2.50E-01	7.76E-02	2.62E-01
45	Shasta	1.16E-01	3.20E-01	3.40E-01
46	Sierra			
47	Siskiyou	2.83E-01	1.48E-02	2.83E-01
48	Solano	1.97E-01	1.18E-01	2.30E-01
49	Sonoma	1.59E-01	1.74E-01	2.36E-01
50	Stanislaus	1.28E-01	5.79E-02	1.41E-01
51	Sutter	4.13E-01	2.82E-01	5.00E-01
52	Tehama	2.13E-01	4.22E-01	4.72E-01
53	Trinity	4.45E-01	2.07E-01	4.90E-01
54	Tulare	1.58E-01	4.31E-02	1.64E-01
55	Tuolumne	2.72E-01	2.28E-01	3.55E-01
56	Ventura	1.21E-01	1.92E-01	2.27E-01
57	Yolo	1.72E-01	4.40E-02	1.78E-01
58	Yuba	1.45E-01	3.49E-01	3.78E-01

Table A12.d: Variance-Bias Tradeoff - AFDC-FG Terminations

County Number	County Name	Variance	Bias	AMSFE
1	Alameda	1.58E-01	1.85E-01	2.43E-01
2	Alpine			
3	Amador	3.73E-01	2.35E-01	4.40E-01
4	Butte	1.14E-01	3.59E-01	3.76E-01
5	Calaveras	2.51E-01	2.92E-01	3.85E-01
6	Colusa	2.61E-01	2.48E-01	3.60E-01
7	Contra Costa	1.36E-01	1.00E-01	1.69E-01
8	Del Norte	2.50E-01	2.44E-01	3.49E-01
9	El Dorado	2.35E-01	2.05E-01	3.12E-01
10	Fresno	7.21E-02	3.96E-02	8.23E-02
11	Glenn	1.96E-01	2.36E-01	3.07E-01
12	Humboldt	2.41E-01	3.71E-01	4.42E-01
13	Imperial	7.03E-01	4.84E-01	8.53E-01
14	Inyo	2.26E-01	2.40E-01	3.30E-01
15	Kern	9.01E-02	3.90E-01	4.01E-01
16	Kings	1.66E-01	3.35E-01	3.74E-01
17	Lake	8.62E-01	1.37E-01	8.73E-01
18	Lassen	1.19E-01	3.08E-01	3.30E-01
19	Los Angeles	7.20E-02	5.27E-02	8.92E-02
20	Madera	9.84E-02	2.61E-01	2.79E-01
21	Marin	1.68E-01	3.21E-01	3.62E-01
22	Mariposa	3.35E-01	2.95E-01	4.46E-01
23	Mendocino	1.37E-01	3.53E-01	3.79E-01
24	Merced	9.86E-02	1.05E-01	1.44E-01
25	Modoc	3.23E-01	2.23E-03	3.23E-01
26	Mono	3.25E-01	3.83E-01	5.02E-01
27	Monterey	2.69E-01	4.04E-01	4.85E-01
28	Napa	1.34E-01	2.23E-01	2.60E-01
29	Nevada	1.34E-01	1.68E-01	2.15E-01
30	Orange	9.95E-02	4.24E-02	1.08E-01
31	Placer	9.47E-02	6.08E-02	1.13E-01
32	Plumas	1.94E-01	3.18E-01	3.73E-01
33	Riverside	8.52E-02	2.58E-02	8.90E-02
34	Sacramento	2.09E-01	1.59E-01	2.63E-01
35	San Benito	2.24E-01	3.20E-01	3.90E-01
36	San Bernardino	6.90E-02	1.27E-01	1.45E-01
37	San Diego	1.54E-01	1.13E-02	1.55E-01
38	San Francisco	4.32E-02	7.98E-02	9.07E-02
39	San Joaquin	9.50E-02	3.66E-01	3.78E-01
40	San Luis Obispo	1.86E-01	1.97E-02	1.87E-01
41	San Mateo	9.83E-02	1.41E-01	1.72E-01
42	Santa Barbara	1.10E-01	7.72E-02	1.35E-01
43	Santa Clara	9.37E-02	1.83E-01	2.05E-01
44	Santa Cruz	1.56E-01	3.72E-03	1.56E-01
45	Shasta	1.15E-01	2.16E-01	2.45E-01
46	Sierra			
47	Siskiyou	1.98E-01	5.54E-02	2.06E-01
48	Solano	1.57E-01	1.59E-01	2.24E-01
49	Sonoma	1.14E-01	1.76E-01	2.10E-01
50	Stanislaus	1.20E-01	3.32E-01	3.53E-01
51	Sutter	4.57E-01	2.91E-01	5.42E-01
52	Tehama	1.41E-01	3.96E-01	4.20E-01
53	Trinity	2.19E-01	1.32E-01	2.56E-01
54	Tulare	9.41E-02	1.30E-03	9.41E-02
55	Tuolumne	2.42E-01	3.48E-01	4.24E-01
56	Ventura	2.28E-01	1.34E-01	2.64E-01
57	Yolo	1.25E-01	4.17E-02	1.32E-01
58	Yuba	1.22E-01	2.98E-01	3.22E-01

Bibliography

- Abdullah, Dewan A. and Peter C. Rangazas**, "Money and the Business Cycle: Another Look (in Notes)," *The Review of Economics and Statistics*, Vol. 70, No. 4. (Nov., 1988), pp. 680-685.
- Achen, Christopher H.**, "Toward Theories of Data: The State of Political Methodology," in *Political Science: The State of the Discipline*, Ada Finifter (editor), Washington, D.C., American Political Science Association, 1983.
- Angrist, Joshua D.**, "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *The American Economic Review*, Vol. 80, No. 3. (Jun., 1990), pp. 313-336.
- Bartels, Larry M.**, *Presidential Primaries and the Dynamics of Public Choice*, Princeton, N.J.: Princeton University Press, 1988.
- Bartels, Larry and Henry E. Brady**, "The State of Quantitative Political Methodology," in *Political Science: The State of the Discipline, 2nd Edition*, Ada Finifter (editor), Washington, D.C.: American Political Science Association, 1993.
- Beauchamp, Tom L. and Alexander Rosenberg**, *Hume and the Problem of Causation*, New York: Oxford University Press, 1981.
- Bennett, Jonathan**, *Events and Their Names*, Indianapolis: Hackett Publishing Company, 1984.
- Bertrand, Russell**, "On the Notion of Cause," in *Mysticism and Logic and Other Essays*, New York: Longmans, Green and Co., 1918.
- Bloom, Howard S., Charles Michalopoulos, Carolyn J. Hill, Ying Lei**, "Can Nonexperimental Comparison Group Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs?," MDRC Working Papers on Research Methodology, Manpower Demonstration Research Corporation, June 2002.
- Brady, Henry E.**, "Knowledge, Strategy and Momentum in Presidential Primaries," in *Political Analysis*, John Freeman (editor), Ann Arbor: University of Michigan Press, 1996.
- Brady, Henry E., Michael C. Herron, Walter R. Mebane, Jasjeet Singh Sekhon, Kenneth W. Shotts, and Jonathan Wand**, "Law and Data: The Butterfly Ballot Episode," *PS: Political Science & Politics*, v34, n1 (2001), pp. 59-69.

Brady, Henry E., Mary H. Sprague, Fredric C. Gey and Michael Wiseman, “The Interaction of Welfare-Use and Employment Dynamics in Rural and Agricultural California Counties,” 2000.

California Work Pays Demonstration Project: County Welfare Administrative Data, Public Use Version 4.1, Codebook, Berkeley, California: UC DATA Archive and Technical Assistance, 2001.

Campbell, Donald T. and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research*, Chicago: Rand McNally, 1966.

Card, David and Alan B. Krueger, “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania,” *The American Economic Review*, Vol. 84, No. 4. (Sep., 1994), pp. 772-793.

Cartwright, Nancy, *Nature's Capacities and Their Measurement*, New York: Oxford University Press, 1989

Chatfield, Chris, “Model Uncertainty, Data Mining and Statistical Inference,” *Journal of the Royal Statistical Society Series A (Statistics in Society)*, Vol. 158, No. 3. (1995), pp. 419-466.

Congressional Budget Office, “Forecasting AFDC Caseloads, with an Emphasis on Economic Factors,” Washington, DC, CBO Staff Memorandum, 1993.

Cook, Thomas D. and Donald T. Campbell, *Quasi-Experimentation: Design & Analysis Issues for Field Settings*, Boston: Houghton Mifflin Company, 1979.

Cook, Thomas D. and Donald T. Campbell, “The Causal Assumptions of Quasi-Experimental Practice,” *Synthese*, Vol. 68, pp. 141-180, 1986

Cox, David Roxbee, “Causality: Some Statistical Aspects,” *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Vol. 155, No. 2 (1992), pp. 291-301.

Cox, Gary W., *Making Votes Count : Strategic Coordination in the World's Electoral Systems*, New York: Cambridge University Press, 1997.

Dessler, David, “Beyond Correlations: Toward a Causal Theory of War,” *International Studies Quarterly*, Vol. 35, No. 3. (Sep., 1991), pp. 337-355.

Elster, Jon, “A Plea for Mechanisms,” in *Social Mechanisms*, Peter Hedstrom and Richard Swedberg (editors), Cambridge: Cambridge University Press, 1998.

Ehring, Douglas, *Causation and Persistence: A Theory of Causation*. New York: Oxford University Press, 1997.

- Fearon, James D.**, “Counterfactuals and Hypothesis Testing in Political Science” in *World Politics*, Vol. 43, No. 2. (Jan 1991), pp. 169-195.
- Ferber, Robert and Werner Z. Hirsch**, “Social Experimentation and Economic Policy: A Survey,” *Journal of Economic Literature*, Volume 16, Issue 4 (Dec., 1978), pp. 1379-1414.
- Firebaugh, Glenn and Kevin Chen**, “Vote Turnout of Nineteenth Amendment Women: The Enduring Effect of Disenfranchisement,” *American Journal of Sociology*, Vol. 100, No. 4. (Jan., 1995), pp. 972-996.
- Fisher, Ronald Aylmer, Sir**, *The Design of Experiments*, Edinburgh, London: Oliver and Boyd, 1935.
- Fraker, Thomas and Rebecca Maynard**, The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs (in Symposium on the Econometric Evaluation of Manpower Training Programs),” *The Journal of Human Resources*, Vol. 22, No. 2 (1987), pp. 194-227.
- Franke, Richard Herbert and James D. Kaul**, “The Hawthorne Experiments: First Statistical Interpretation,” *American Sociological Review*, Vol. 43, No. 5 (1978), pp. 623-643.
- Freedman, David A.**, “Statistical Models and Shoe Leather” in *Sociological Methodology*, Vol. 21 (1991), pp. 291-313.
- Freedman, David A.**, “As Others See Us: A Case Study in Path Analysis” in *Journal of Educational Statistics*, Vol. 12. No. 2 (1987), pp. 101-223, with discussion.
- Freedman, David A.**, “From Association to Causation via Regression,” in V. R. McKim and S. P. Turner (editors), *Causality in Crisis?* Notre Dame IN: University of Notre Dame Press, 1997, pp. 113-161.
- Freedman, David A.**, “From Association to Causation: Some Remarks on the History of Statistics,” *Statistical Science*, Vol. 14 (1999), pp. 243–58.
- Galison, Peter Louis**, *How Experiments End*, Chicago : University of Chicago Press, 1987.
- Gasking, Douglas**, “Causation and Recipes,” *Mind, New Series*, Vol. 64, No. 256 (Oct. 1955), pp. 479-487.
- Glennan, Stuart S.**, “Mechanisms and the Nature of Causation,” *Erkenntnis*, Vol. 44 (1996), pp. 49-71.

- Goldthorpe, John H.**, "Causation, Statistics, and Sociology," *European Sociological Review*, Vol. 17, No. 1 (2001), pp. 1-20.
- Goodman, Nelson**, "The Problem of Counterfactual Conditionals," *Journal of Philosophy*, Vol. 44, No. 5. (Feb 1947), pp. 113-128.
- Greene, William H.**, *Econometric Analysis*, Upper Saddle River, New Jersey: Prentice-Hall, 1997.
- Harre, Rom and Edward H. Madden**, *Causal Powers: A Theory of Natural Necessity*. Imprint: Oxford, [Eng.]: B. Blackwell, c1975.
- Hausman, Daniel M**, *Causal Asymmetries*. Imprint: Cambridge, U.K.; New York: Cambridge University Press, 1998.
- Heckman, James J.**, "Sample Selection Bias as a Specification Error," *Econometrica*, Vol. 47, No. 1. (Jan., 1979), pp. 153-162.
- Heckman, James J.**, "Randomization and Social Policy Evaluation," in *Evaluating Welfare and Training Programs*, Charles F. Manski and Irwin Garfinkel (editors), Cambridge, MA: Harvard University Press, 1992.
- Heckman, James J. and V. Joseph Hotz**, "Choosing Among Alternative Non-experimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training: Rejoinder (in Applications and Case Studies), *Journal of the American Statistical Association*, Vol. 84, No. 408. (Dec., 1989), pp. 878-880.
- Heckman, James and Richard Robb**, "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, James Heckman and Burton Singer (editors), New York: Wiley, 1995.
- Heckman, James J. and Jeffrey A. Smith**, "Assessing the Case for Social Experiments," *The Journal of Economic Perspectives*, Volume 9, Issue 2 (Spring 1995), pp 85-110.
- Hedstrom, Peter and Richard Swedberg** (editors), *Social Mechanisms: An Analytical Approach to Social Theory*, New York : Cambridge University Press, 1988.
- Hempel, Carl G.**, *Aspects of Scientific Explanation*, New York: Free Press, 1965.
- Hill, A. Bradford**, "The Environment and Disease: Association or Causation?," Proceedings of the Royal Society of Medicine, Vol. 58 (1965), pp. 295-300.
- Holland, Paul W.**, "Statistics and Causal Inference (in Theory and Methods)," *Journal of the American Statistical Association*, Vol. 81, No. 396. (Dec., 1986), pp. 945-960.

- Holland, Paul W.**, “Causal Inference, Path Analysis, and Recursive Structural Equations,” *Sociological Methodology*, Vol. 18. (1988), pp. 449-484.
- Holland, Paul W. and Donald B. Rubin**, “Causal Inference in Retrospective Studies,” *Evaluation Review*, Vol. 12 (1988), pp. 203-231.
- Hotz V. Joseph, Guido W. Imbens and Jacob A. Klerman**, “The Long-Term Gains from GAIN: A Re-Analysis of the Impacts of the California GAIN Program,” September 2001.
- Hotz, V. Joseph, Charles H. Mullin, and Seth G. Sanders**, “Bounding Causal Effects Using Data From a Contaminated Natural Experiment: Analysis the Effects of Teenage Childbearing,” *The Review of Economic Studies*, Vol. 64, No. 4, Special Issue: Evaluation of Training and Other Social Programmes. (Oct., 1997), pp. 575-603.
- Hume, David**, *A Treatise of Human Nature (1739)*, edited by L. A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon Press, 1978.
- Jenkins, Jeffery A.**, “Examining the Bonding Effects of Party: A Comparative Analysis of Roll-Call Voting in the U.S. and Confederate Houses,” *American Journal of Political Science*, Vol. 43, No. 4. (Oct., 1999), pp. 1144-1165.
- Jones, Stephen R. G.**, “Was There a Hawthorne Effect?” *American Journal of Sociology*, Vol. 98. No. 3. (1992), pp. 451-468.
- Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lütkepohl and Tsoung-Chao Lee**, *Introduction to the Theory and Practice of Econometrics*, New York: John Wiley and Sons, 1988.
- Lakoff, George and Mark Johnson**, “Conceptual Metaphor in Everyday Language” in *The Journal of Philosophy*, Vol. 77, No. 8. (Aug., 1980), pp. 453-486, (1980a).
- Lakoff, George and Mark Johnson**, *Metaphors We Live By*, Chicago: University of Chicago Press, 1980 (1980b).
- Lakoff, George and Mark Johnson**, *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*, New York: Basic Books, 1999.
- LaLonde, Robert J.**, “Evaluating the Econometric Evaluations of Training Programs with Experimental Data,” *The American Economic Review*, Vol. 76, No. 4. (Sep., 1986), pp. 604-620.
- Lewis, David**, *Counterfactuals*, Cambridge: Harvard University Press, 1973 (1973a).

- Lewis, David**, "Causation," *Journal of Philosophy*, Vol. 70, No. 17, (Oct 1973), pp. 556-567 (1973b).
- Lewis, David**, "Counterfactual Dependence and Time's Arrow," *Noûs*, Vol. 13, No. 4, Special Issue on Counterfactuals and Laws, pp. 455-476 (Nov 1979).
- Lewis, David**, *Philosophical Papers*, New York: Oxford University Press, Vol. 2, 1986.
- Lichbach, Mark Irving**, *The Rebel's Dilemma*, Ann Arbor : University of Michigan Press, 1995.
- Lichbach, Mark Irving**, *The Cooperator's Dilemma*, Ann Arbor : University of Michigan Press, 1996.
- Lijphart, Arend**, *Electoral Systems and Party Systems : A Study of Twenty-seven Democracies, 1945-1990*, Oxford ; New York: Oxford University Press, 1994.
- Lorenz, Edward N**, "Deterministic Nonperiodic Flow", *Journal of the Atmospheric Sciences* 20 (1963), pp.130-141.
- Machamer, Peter, Lindley Darden, and Carl F. Craver**, "Thinking about Mechanisms" in *Philosophy of Science*, Vol. 67, No. 1 (2000), pp. 1-25.
- Mackie, John L.**, "Causes and Conditions," *American Philosophical Quarterly*, 2/4 (1965), pp. 245-64.
- Manski, Charles F.**, "Identification Problems in the Social Sciences," *Sociological Methodology*, Vol. 23. (1993), pp. 1-56.
- Manski, Charles F.**, *Identification Problems in the Social Sciences*, Cambridge, Mass.: Harvard University Press, 1995.
- Marini, Margaret Mooney, and Burton Singer**, "Causality in the Social Sciences," *Sociological Methodology*, Vol. 18 (1988), pp. 347-409.
- Mauldon, Jane, Jan Malvin, Jon Stiles, Nancy Nicosia and Eva Seto**, "Impact of California's Cal-Learn Demonstration Project: Final Report", UC DATA Archive and Technical Assistance, 2000.
- Mellors, D.H.**, *The Facts of Causation*, London: Routledge, 1995.
- Menzies, Peter and Huw Price**, "Causation as a Secondary Quality," *British Journal for the Philosophy of Science*, Vol. 44, No. 2 (1993), pp. 187-203.
- Metrick, Andrew**, "Natural Experiment in "Jeopardy!"," *The American Economic Review*, Vol. 85, No. 1. (Mar., 1995), pp. 240-253.

- Meyer, Bruce D., W. Kip Viscusi, and David L. Durbin**, "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment," *The American Economic Review*, Vol. 85, No. 3. (Jun., 1995), pp. 322-340.
- Mill, John Stuart**, *A System of Logic, Ratiocinative and Inductive*, 8th Edition, New York: Harper and Brothers, 1988.
- Pearson, Karl**, *The Grammar of Science*, 3rd Edition, Revised and Enlarged, Part 1. – Physical, London: Adam and Charles Black, 1911.
- Pindyck, Robert S. and Daniel L. Rubinfeld**, *Econometric Models and Economic Forecasts*, New York: McGraw-Hill, 1991.
- Ragin, Charles C.**, *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*. Imprint: Berkeley: University of California Press, 1987.
- Riccio, James et al.**, "GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program", Manpower Demonstration Research Corporation, 1994.
- Rosenbaum, Paul R. and Donald B. Rubin**, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, Vol. 70, No. 1. (Apr., 1983), pp. 41-55.
- Rosenweig, Mark R. and Kenneth I. Wolpin**, "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, Vol. 48, No. 1. (Jan., 1980), pp. 227-240.
- Rubin, Donald B.**, "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, Vol. 66. No. 5 (1974), pp. 688-701.
- Rubin, Donald B.**, "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics*, Vol. 6, No. 1. (Jan., 1978), pp. 34-58.
- Rubin, Donald B.**, "Statistics and Causal Inference: Comment: Which Ifs Have Causal Answers," *Journal of the American Statistical Association*, Volume 81, Issue 396, December 1986, pp.961-962.
- Rubin, Donald B.**, "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies." *Statistical Science*, Vol. 5, No. 4 (1990), pp. 472-480.
- Ruell, D.**, "Deterministic Chaos: The Science and the Fiction," The Claude Bernard Lecture, 1989, *Proceedings of the Royal Society of London, A, Mathematical and Physical Sciences*, Volume 427, Issue 1873, February 8, 1990, pp. 241-248.

- Salmon, Wesley C.**, *Four Decades of Scientific Explanation*, Imprint: Minneapolis: University of Minnesota Press, 1989.
- Sawyer, Alan G., John G. Lynch, Jr., David L. Brinberg**, "A Bayesian Analysis of the Information Value of Manipulation and Confounding Checks in Theory Tests," *The Journal of Consumer Research*, Volume 21, #4, March 1995, pp. 581-595.
- Sheffrin, Steven M.**, *Rational Expectations*, Cambridge [Cambridgeshire] ; New York : Cambridge University Press, 1983.
- Simon, Herbert A.**, "On the Definition of the Causal Relation," *The Journal of Philosophy*, Vol. 49, No. 16(Jul 1952), pp. 517-528.
- Simon, Herbert A. and Yumi Iwasaki**, "Causal Ordering, Comparative Statics, and Near Decomposability," *Journal of Econometrics*, Vol. 39 (1988), pp. 149-173.
- Sobel , Michael E.**, "Causal Inference in the Social and Behavioral Sciences," in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel (editors) New York: Plenum Press, 1995.
- Sorenson, Aage B.**, "Theoretical Mechanisms and the Empirical Study of Social Processes," in *Social Mechanisms*, Peter Hedstrom and Richard Swedberg (editors), Cambridge: Cambridge University Press, 1998.
- Sosa, Ernest and Michael Tooley**, *Causation*, edited by Ernest Sosa and Michael Tooley. Imprint: Oxford; New York: Oxford University Press, 1993.
- Sprinzak, Ehud**, "Weber's Thesis as an Historical Explanation," *History and Theory*, Vol. 11, No. 3. (1972), pp. 294-320.
- Tetlock, Philip E. and Aaron Belkin (editors)**, *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*, Imprint: Princeton, N.J.: Princeton University Press, 1996.
- von Wright, Georg Henrik**, *Explanation and Understanding*. Ithaca, N.Y.: Cornell University Press, 1971.
- von Wright, Georg Henrik**, *Causality and Determinism*, New York: Columbia University Press, 1974.
- Wand, Jonathan N., and Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Michael C. Herron, and Henry E. Brady**. "The butterfly did it: the aberrant vote for Buchanan in Palm Beach County, Florida," *American Political Science Review*, Vol. 95, No. 4 (Dec. 1991), pp. 793-810.

Wawro, Geoffrey, *The Austro-Prussian War : Austria's war with Prussia and Italy in 1866*, New York: Cambridge University Press, 1996.

Weber, Max, *Selections in Translation*, W.G. Runciman (editor) and E. Matthews (translator), Cambridge: Cambridge University Press, 1906 [1978].