ABSTRACT

BUTRY, DAVID THOMAS. Estimating the Efficacy of Wildfire Management Using Propensity Scores. (Under the direction of Dr. Walter N. Thurman and Dr. Marcia L. Gumpertz).

This research examines the effect wildfire mitigation has on broad-scale wildfire behavior. Each year, hundreds of million of dollars are spent on fire suppression and fuels management applications, yet little is known, quantitatively, of the returns to these programs in terms of their impact on wildfire extent and intensity. This is especially true when considering that wildfire management influences and reacts to several, often times confounding factors, including socioeconomic characteristics, values at risk, heterogeneous landscapes, and climate. Due to the endogenous nature of suppression effort and fuels management intensity and placement with wildfire behavior, least-square models may prove inadequate. Instead, I examine the applicability of two-stage least squares, propensity score blocking, and a newly developed technique, control score blocking in modeling wildfire. This research makes several significant contributions including: (1) applying techniques developed in labor economics and in epidemiology to evaluate the effects of natural resource policies on landscapes, rather than on individuals; (2) a better understanding of the relationship between wildfire mitigation strategies and their influence on broad-scale wildfire patterns; (3) quantifying the returns to suppression and fuels management on wildfire behavior.

ESTIMATING THE EFFICACY OF WILDFIRE MANAGEMENT USING PROPENSITY SCORES

by

DAVID THOMAS BUTRY

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

ECONOMICS

Raleigh

2006

APPROVED BY:

Dr. Walter N. Thurman Co-chair of Advisory Committee	Dr. Marcia L. Gumpertz Co-chair of Advisory Committee
Dr. Raymond B. Palmquist	Dr. Jeffrey P. Prestemon

DEDICATION

This dissertation is dedicated to my grandmother, Clara Butry (Babci) and my grandfather, Robert LaLonde, both who have inspired me throughout life with their spirit, determination, and individuality.

BIOGRAPHY

David T. Butry was born in Buffalo, NY in 1973. He graduated from the State University of New York at Buffalo with a Bachelor of Arts in Political Science in 1996. After moving to North Carolina in August 1996, he graduated from Duke University with a Master of Environmental Management in Resource Economics and Policy in 1999. He served as an economist with the USDA Forest Service from November 1999 to May 2006 and is currently employed as an economist with the National Institute of Standards and Technology. Most importantly, he married Charlotte Anne Lanteri in October 2002. David lives with his beautiful wife and their “baby,” Lola, in Rockville, Maryland.

ACKNOWLEDGEMENTS

I have enjoyed a great deal of support and guidance in this effort, from many wonderful individuals who deserve recognition. My wife Charlotte has been an amazing source of motivation, understanding, and love, without which this could never have been possible. I would like to thank my parents, David and Cathy, for always believing in me, for allowing me to dream, and for just "being there." Thanks to my sister, Erin, and my fabulous nieces, Sydney and Natalie who always make me smile. Thanks to Dan May for being a best friend. Thanks to Charlotte's parents, Vincent and Anne, for being a constant source of support and encouragement. I wish to acknowledge and thank Subhrendu K. Pattanayak for being a mentor—I owe you a great deal! Thanks to the members of the USDA Forest Service, Economics of Forest Protection and Management, SRS-4851 in Research Triangle Park, North Carolina—David Wear, Jeffrey Prestemon, Bobby Huggett, John Pye, Karen Abt, Evan Mercer, and Tom Holmes—for helping me cut my research teeth. Finally, I am indebted to my committee, my co-chairs, Marcia Gumpertz and Walter Thurman, and to Jeffrey Prestemon and Raymond Palmquist for their time, energy, and patience.

This research was funded by the National Fire Plan (USDA Forest Service) as part of a project titled, "Quantifying Trade-offs of Alternative Vegetation Management Strategies, Wildfire, and Suppression in Fire Prone Regions of the US." Any remaining errors in this dissertation are the responsibility of the author.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 OVERVIEW

Nationwide, hundreds of millions of dollars are spent fighting wildfires and still millions of acres of forestland burn each year. Additionally, hundreds of millions more are spent on fuels management programs (programs that mitigate wildfire risk by physically reducing the amount of available flammable vegetation within the forest). What is not clear is whether society is optimally funding these wildfire management (suppression or fuels management) programs. One reason for this uncertainty is that little information exists on the damages and benefits associated with wildland fire, although Butry (2001) estimates the economic damage of the 1998 Florida wildfires. Another reason is little research exists quantifying the effects of wildfire mitigation programs on fire behavior and damage.

The discipline of economics is well suited to examine issues concerning tradeoffs, optimality, and allocations of resources. Tradeoffs are inherent in wildland management. Tradeoffs exist between how much wildfire is allowed, how much is controlled, and how to determine the best mix of suppression and fuels management to achieve these goals. The idea of an "optimal wildfire size" was introduced by Sparhawk in 1925. He explains that the optimal size is one that yields the minimum 'cost plus loss' (cost of wildland management[1] plus loss due to wildfire damage). This is an important finding as it may justify, in some instances, "let burn" or limited-action strategies—when the wildfire damages pale in comparison to the cost of treatment. Rideout and Omi (1990) formalize innovations instituted into the least-

[1] His focus was on suppression.

cost model by Gorte and Gorte (1979) and Davis (1965), who contend that optimal wildfire size is that which corresponds with the minimum cost plus *net value change* (NVC—damages net of benefits), as a profit maximization type problem. The goal here is to maximize the value of wildfire avoidance given the costs of wildfire management.

However, even before one can determine the optimal allocation of fire suppression and fuels management resources, it is critical to understand the physical relationship between management strategies and wildfire behavior. For example, society must understand the effectiveness of fire management before it can decide how much is needed. Unfortunately, there is a real lack of research modeling wildfire behavior at broad, policy relevant spatial scales. While experimental research has focused on the relationship between fuel loads and fire risk, these are based on small-forested plots. Although fine scale analysis is informative, perhaps it is not appropriate for policymaking. Further, it often ignores spatial and spatio-temporal relationships between wildfire and physical intervention strategies. It also cannot comprehend many of the complexities of landscapes and larger scale weather and climate phenomena.

There is a vulnerability of fine scale and experimental laboratory analysis to ignoring the influence society and culture, institutions, and environmental phenomena have on ignition patterns and fire propagation. The wildland-urban interface (WUI) is the area where communities abut forested areas. In 2000, the WUI comprised 9.4% of the land in the coterminous US and contained 38.5% of all housing units (Stewart et al. 2005). These numbers are up substantially from 1990 levels, as the decade experienced a 19.2% and 22.3%

growth in WUI lands and homes within the WUI (Stewart et al. 2005). Figure 1.1 presents a stylized depiction of wildfire in the wildland-urban interface and highlights its interconnectedness with communities and the environment. In the center of figure 1.1 is the fire triangle, which is made up of three sides—heat, fuel, and oxygen. Remove anyone of these sides and fire cannot exist. Figure 1.1 illustrates that wildfire is influenced by several factors, including climate and weather, socioeconomic characteristics (population, housing density, roads, etc.), and wildland management (fire suppression and fuels management). Each one of these factors is related to one or more of the sides of the fire triangle (heat, fuel, oxygen).

Focusing on the wildfire management relationship, it is apparent that fire suppression targets all three sides of the fire triangle. Fire fighters use water to cool the fire and reduce its heat content; fire fighters use fire retardants to smother the flames and deprive fire of oxygen; fire fighters also build fire lines (fire barriers devoid of flammable materials) to separate fire from available fuels. Fuels management, such as prescribed fire fuels treatments, targets only one side of the triangle—fuels—by eliminating dangerous fuel loads before the start of the fire season. Not surprisingly, factors that affect wildfire also affect suppression and fuel management. In addition, wildfire conditions influence suppression and fuels management decisions through a series of feedback mechanisms. While fire fighting mitigates fire size and intensity, the initial fire conditions, including proximity to values at risk (e.g., populations, housing development, infrastructure) influence the size and swiftness of response, especially during times of multiple fires (requiring prioritization of wildfires). Fuels treatments are performed far in the advance of the fire season, but their locations are not distributed randomly across the landscape, nor are their size and intensity randomly selected. Instead, fuels treatments are strategically located to areas of higher wildfire risk, vulnerable to damage from fire.

In economics, it is rare to use experimentally controlled data due to expense and inability to replicate real-world conditions. Instead, the majority of applied economic studies rely on observational data—one observes individuals behaving a certain way. Evaluating the effect of programs using observational data is challenging because often there are underlying factors that influence whether or not an individual enters a program. In the case of wildfire, wildfires are not randomly selected to receive wildfire management. The decision depends on a number of factors, some of which also influence wildfire behavior. For instance, because prescribed fire is not randomly applied across the landscape, it is not unusual to observe areas with higher levels of prescribed fire to have also higher levels of wildfire. Yet, this does not necessarily imply that prescribed fire is ineffective or counterproductive, rather the relationship between prescribed fire and wildfire is complex. The true measure of the effectiveness of wildfire management can be determined if one compares the same group of wildfires having been managed with their untreated selves. Of course, this latter group, the counterfactual, is not observable.

Because factors influencing wildfire behavior also influence wildland management decisions, these factors must be accounted for in statistical models. Failing to acknowledge this complex relationship renders estimates from standard statistical models biased as it invalidates some of the required modeling assumptions. Much of the previous empirical wildfire research has focused on modeling either the probability of a wildfire occurrence (ignition) or wildfire size or intensity. The former research primarily examines the factors involved in the probability of fire ignition (fire risk modeling), while the latter examines the factors contributing to a fire's final size or degree of intensity. In general, previous research finds wildfire behavior (however defined—whether meaning frequency, occurrence, size, or intensity) to be related to four sets of factors: wildfire specific characteristics, climate and weather conditions, wildfire management and mitigation (including prescribed fire and suppression effort), and landscape attributes (including dominant landuse and landcover characteristics and socioeconomic characteristics), but only recently has modeling made explicit the endogeneity of wildfire management (Yoder 2004).

1.2 RESEARCH OBJECTIVES

The primary objectives of this research are threefold; (1) to examine several program evaluation econometric techniques when treatment is a continuous variable and to determine their appropriateness when modeling wildfire management; (2) to evaluate, through Monte Carlo simulation, the properties of the program evaluation estimators and to compare their absolute and relative performance (unbiasedness and precision) to specified changes to the structure of the data generation process; (3) to estimate the efficacy of wildfire management (wildfire suppression and fuels management activities) on wildfire behavior using St. Johns River Water Management District in Florida as a case study.

1.3 CONTRIBUTIONS

This research makes several contributions to the fields of statistics and econometrics, and to the field of wildfire economics. The asymptotic and small sample properties of several leading program evaluation methods (ordinary least squares, two-stage least squares, propensity score blocking) are compared across differences in the data generation process. Variations in the data generation are meant to mimic different problems encountered when using sampled observational data. The statistical estimators are evaluated for unbiasedness and precision under differing levels of correlation between the program treatment and confounders that underlie the program-outcome relationship (endogeneity), correlation between the program treatment and an instrumental variable for the program treatment (instrument quality), correlation between the instrumental variable and the confounders (instrument validity), and functional form assumptions. In addition, this research develops a new program evaluation method (control score blocking) based on a hybridization of propensity score blocking and control functions for use when evaluating the effectiveness of programs when treatment is a continuous variable.

This study is novel in that it is one of the first to apply these program evaluation techniques developed in the fields of labor, education, and epidemiology to environmental and natural resource economics. This research estimates the effectiveness of wildfire management using the St. Johns River Water Management District (SJRWMD), in northeast Florida as a case study. After accounting for endogenous and possible nonlinear relationships between fuels management (modeled using prescribed fire acres), suppression (modeled using fire crew response time to the wildfire) and wildfire behavior, the results suggest compelling evidence that quicker response times limit wildfire behavior, and that prescribed fire may provide beneficial effects against wildfire behavior for three years after its application.

Economically, the financial returns to wildfire management appear impressive. Burning was prescribed for 669,290 acres (for wildfire hazard reduction) in the SJRWMD over this period of analysis. Roughly, $17.6 million was spent on prescribed fire, but the model shows that prescribed fire applications reduced wildfire behavior by 1.3%, yielding $27.0 million in avoided wildfire damages. Thus, for every $1 spent on prescribed fire treatments, $1.53 in wildfire damage was avoided. While it is not possible to calculate a similar estimate for suppression response (it is not known how large a fire would become without any response), it is possible to estimate what the financial return would have been, over the period of analysis, if suppression response had been 10% quicker. Any increase in fire fighting resources yielding a reduction in fire crew response times by 10% would have yielded $62.0 million in avoided fire damages.

SECTION REFERENCES

Butry, D.T. Mercer, D.E., Prestemon, J.P., Pye, J.M., and T.P. Holmes. 2001. “What is the Price of Catastrophic Wildfire?” *Journal of Forestry*, 99(11):9-17.

Davis, L.S. and R.W. Cooper. “How Prescribed Burning Affects Wildfire Occurrence.” *Journal of Forestry* 61(12): 915-917.

Gorte, J.K. and R.W. Gorte. 1979. *Application of Economic Techniques to Fire Management: A Status Review and Evaluation*. General Technical Report INT-53, Ogden, UT: USDA Forest Service.

Rideout, D.B. and P.N. Omi. 1990. “Alternative Expressions for the Economic Theory of Forest Fire Management,” *Forest Science*, 36(3):614-624.

Sparhawk, W.N. 1925. “The Use of Liability Ratings in Planning Forest Fire Protection.” *Journal of Agricultural Resources*, 30, 693-762.

Stewart, S., V. Radeloff, R. Hammer, J. Fried, S. Holcomb, and J. McKeefry. 2005. Mapping the Wildland Urban Interface and Projecting its Growth to 2030. USDA Forest Service, North Central Research Station.

Yoder, J. 2004. “Playing with Fire: Endogenous Risk in Resource Management.” *American Journal of Agricultural Economics* 86, 933-948.

SECTION FIGURES

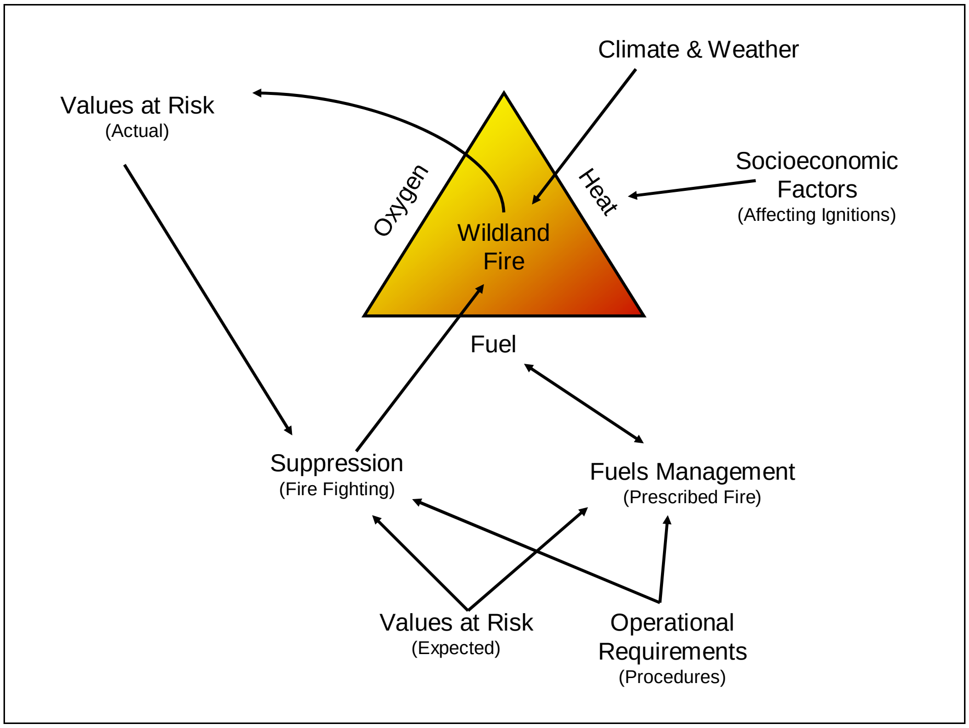

Figure 1.1. The fire triangle in the wildland-urban interface.

2. PROGRAM EVALUATION ECONOMETRICS

2.1 THE VALUE OF PROGRAMS

In general terms, programs (or their treatments) are designed to produce benefits for the participants, perhaps in the form of higher wages (job training program), higher educational achievement (education program), health enhancement (medicinal drug study), or in the case of wildfire, damage averted from wildfire (wildfire management program). In the program evaluation terminology, a program induces a treatment that affects outcome. For instance, wages, educational attainment, health attribute, and damage averted are all outcomes affected by the program treatment.

In much of the program evaluation literature the effect of a program or treatment is conceptualized as (for example, see Heckman and Hotz 1989; Dehejia and Wahba 1999; Smith and Todd 2005):

$$y_i = \tilde{y}_i + \Delta t_i \tag{2.1}$$

where y_i is the observed outcome for observation i, $\tilde{y}_i$ is the outcome if i is not selected into treatment (one can also think of $\tilde{y}_i$ as potential outcome before treatment), t_i denotes treatment (or participation in the program), and Δ is the treatment effect. It is the simple notion that if an observation undergoes treatment, the outcome value is enhanced by the treatment effect (Δ), whereas if an observation does not undergo treatment, then the outcome value is equal to the potential outcome $(\tilde{y})$.

Equation (2.1) is not a structural equation, in the sense that outcome is represented as a function of variables based on underlying theory. Essentially the term $\tilde{y}$ embodies all the variables that affect outcome, except for treatment. From a structural model viewpoint, one might envision that outcome is a function of treatment, a set of variables[2] that only directly affect outcome, *Z*, and a set of variables that directly affect outcome and treatment, *X* (this distinction will prove important later), and are uncorrelated with Z, so $Cov(X, t) \neq 0$ and $Cov(X, Z) = 0$, and *e* a random error component. In a structural model of program outcome, these components comprise $\tilde{y}_i$ ($\tilde{y}_i = \pi_z Z_i + \pi_x X_i + e_i$) from (2.1). Writing (2.1) as a structural equation, yields:

$$y_i = \pi_z Z_i + \pi_x X_i + \Delta t_i + e_i\,. \qquad (2.2)$$

The reason for distinguishing between (2.1) and (2.2) is because some program evaluation econometric techniques assume a structure similar to (2.1), for instance propensity score matching, while others, such as ordinary least squares, assume a structure similar to (2.2). I present equations (2.1) and (2.2) to demonstrate that they can represent the same concept.

[2] The existence of these (*Z*) variables are not required for any of the modeling approaches used here, they are simply included in an attempt to capture as much modeling realism as possible.

2.2 PROGRAM TREATMENT AS A BINARY VARIABLE

The vast majority of program evaluation econometrics focuses on programs with binary treatments, so I begin here. The case of treatment as a continuous variable is an extension of the binary case.

Mean Comparison

Perhaps the most straightforward program evaluation technique is a comparison of the mean outcomes between the treated group and another group that has not been treated. The mean comparison will produce an unbiased estimate of the treatment effect (average treatment effect) so long as the untreated group is similar to the treated in all ways except for treatment status (Rosenbaum and Rubin 1983). This implies that there are no unobserved confounders (underlying factors that influence both outcome and treatment status) to obscure the outcome-treatment relationship. The mean treatment effect for the treated can be shown by re-arranging (2.1) and taking its mathematical expectation conditional on treatment selection ($t_i = 1$):

$$E[\Delta t_i \mid t_i = 1] = \Delta = E[y_i \mid t_i = 1] - E[\tilde{y}_i \mid t_i = 1]. \qquad (2.3)$$

This is the true treatment effect on the treated observations. The true treatment effect on the treated is the difference of two components—the average outcome of the treated group, $E[y_i \mid t_i = 1]$, which is observable, and the average outcome for the same observations if they were not treated, $E[\tilde{y}_i \mid t_i = 1]$, which is not observable. For every observation it is either observed that it was treated or not, but never both. It is not known, for any treated observation, what its outcome would have been had it not participated or been selected into treatment (the counterfactual is unobserved). Using the untreated group as a counterfactual for the treated is only appropriate (unbiased), if the outcome-treatment relationship is unconfounded. Unconfoundedness implies that there are no unobservable factors that directly influence both outcome and treatment selection, thereby obscuring the true outcome-treatment relationship. The effect that would be estimated by comparing the mean outcomes between the observed treated and untreated groups is:

$$\tilde{\Delta} = E[y_i \mid t_i = 1] - E[\tilde{y}_i \mid t_i = 0] \qquad (2.4)$$

which only equals the true treatment effect for the treated observations if $E[\tilde{y}_i \mid t_i = 0] = E[\tilde{y}_i \mid t_i = 1]$. Note that $E[\tilde{y}_i \mid t_i = 0]$ is observable because $E[\tilde{y}_i \mid t_i = 0] = E[y_i \mid t_i = 0]$, which can be shown by taking expectations of (2.1), conditional on $t_i = 0$. This condition holds only if potential outcome prior to the treatment decision $(\tilde{y}_i)$ is independent of treatment assignment $(\tilde{y}_i \perp t_i)$. The lack of independence is a problem and is the central issue driving program evaluation econometrics (in the ordinary least squares context this is the "omitted variable" problem). If the estimated treatment effect does not equal the true treatment effect, then there exists an unobserved variable (or set of variables) that confounds the outcome-treatment relationship. Because *X* is a set of variables that directly affect both treatment assignment and outcome, *X* confounds the outcome-treatment relationship if

$$E[\tilde{y}_i \mid t_i = 1] - E[\tilde{y}_i \mid t_i = 0] \neq 0 \qquad (2.5)$$

but conditioning on *X* removes the differences between the groups (before applying the treatment) so that,

$$E[\tilde{y}_i \mid t_i = 1, X_i] - E[\tilde{y}_i \mid t_i = 0, X_i] = 0\,. \qquad (2.6)$$

Bias occurs because the probability that an observation will participate in a program is different between participants and non-participants due to these underlying factors, *X*, which if unobserved confound the outcome-treatment relationship.

2.2.1 Regression Approaches

It is the inability of (1) observing (or measuring) any or all of the variables that directly influence outcome and treatment or (2) specifying the relationship between confounders and outcome that complicates estimation of the true relationship between outcome and treatment. Re-writing (2.2) to illustrate the situation where *X* is a set of unobservable confounders yields:

$$y_i = \pi_z Z_i + \Delta t_i + \varepsilon_i \qquad (2.7)$$

where,

$$\varepsilon_i = \pi_x X_i + e_i \tag{2.8}$$

and $E[\varepsilon_i | t_i] \neq 0$ and $E[\varepsilon_i] = 0$.

Ordinary Least Squares

If there are no unobservables or if treatment is uncorrelated with the unobservables, then ordinary least squares estimation of (2.7) will yield an unbiased treatment effect estimate (Δ). If the confounders, *X*, are unobserved and correlated with treatment, estimation of (2.7) will yield a biased treatment effect estimate because $E[\varepsilon_i \mid t_i] \neq 0$ (Greene 2000).

Instrumental Variables

The method of instrumental variables (IV) was designed to eliminate the endogeneity present in simultaneous equations or to mitigate measurement error (Angrist and Krueger 2001). Recently IV has been used to evaluate the effects of programs with endogenous treatment variables (Angrist 1990; Angrist and Krueger 1991; Angrist and Krueger 2001; Angrist 1998; Angrist and Lavy 1999). Intuitively, IV eliminates the influence of the confounders of the outcome-treatment relationship by using only a part of the variability of the endogenous treatment regressor (the variability that is uncorrelated with the confounders) when estimating the outcome-treatment relationship (Angrist and Krueger 2001). An instrument for treatment is a variable that is correlated with the endogenous treatment variable, but uncorrelated with the confounders.

An example of an IV estimator when treatment is an endogenous binary variable is:

$$\begin{bmatrix} \hat{\pi}_z^{iv} \\ \hat{\Delta}^{iv} \end{bmatrix} = (K'D)^{-1} K'y \tag{2.9}$$

where Z is an N x K_Z matrix, t is an N x 1 vector, $D = [Z, t]$ and is a N x K_D matrix ($K_D = K_Z + 1$) , $K = [Z, h]$ and is a N x K_K matrix ($K_K = K_Z + 1$), h is an N x 1 vector instrument for treatment (Z instruments for itself[3] in K), and the "*iv*" superscript on the parameters denotes that they are estimated using instrumental variables. The instrument h is a valid instrument for treatment if $Cov(h, t) \neq 0$ and $Cov(h, \varepsilon) = 0$, where in the limit,

$$\begin{aligned} \text{plim} \begin{bmatrix} \hat{\pi}_z^{iv} \\ \hat{\Delta}^{iv} \end{bmatrix} &= \text{plim}\,[(K'D)^{-1} K'y] \\ &= \begin{bmatrix} \pi_z \\ \Delta \end{bmatrix} + \text{plim}\,[K'\varepsilon] \qquad . \\ &= \begin{bmatrix} \pi_z \\ \Delta \end{bmatrix} \end{aligned} \tag{2.10}$$

Unlike OLS, IV does not require the confounders correlated with treatment to be observable. Instrumental variables requires that the instrument(s) affects the outcome only through its effect on treatment and is uncorrelated with the confounders (the identification condition and

[3] Z is a valid instrument for itself since $Cov(Z, Z) = Var(Z) \neq 0$ and $Cov(Z, \varepsilon) = 0$, because Z is assumed orthogonal to X and e.

exogeneity assumption). Also, the IV asymptotic variance-covariance matrix for the estimated parameters is calculated differently than is the OLS variance-covariance matrix (see Greene 2000).

The challenge is finding valid instruments for the endogenous treatment variable. In Angrist, Imbens, and Rubin (1996), military service status (military service/no-military service) is instrumented with military draft lottery status (low lottery number/high lottery number) to assess the effect military service has on civilian mortality. They argue that draft lottery status is a good instrument for military service since an individual's draft lottery number is correlated with military status but is not presumed correlated with confounders that affect both military status and civilian mortality (e.g., risk aversion or socioeconomic status).

Control Functions

Control function models relate observed information (the observed variables) to the unobserved confounders by applying additional structure to the model so as to control for the nonzero conditional expectation of the error term (Heckman and Navarro-Lozano 2004). The intuition behind control function methods is such: because endogeneity is caused by unobservable confounders masking the true outcome-treatment relationship in OLS estimation, if one were to estimate the unobservables using observable information and introduce this estimate into the OLS equation, then treatment would no longer be endogenous to outcome due to unobservables, as they have now been accounted for (controlled for). Control function models are able to account for the unobservables by making assumptions regarding the relationship among the unobservables and observables, and their influence on treatment and the outcome.

The treatment effect is estimated using two estimation steps, and when treatment is binary, a bivariate probit selection model is used (Heckman 1978). The control function approach is an alternative to the IV method in the binary treatment case. Using Heckman's bivariate probit selection model, equation (2.7) is modeled in two estimation steps (2.11) then (2.12):

$$\begin{aligned} t_i^* &= \delta_0^{bcf} + \delta_h^{bcf} h_i + v_i^{bcf} \\ t_i &= 1 \text{ if } t_i^* > 0 \\ t_i &= 0 \text{ if } t_i^* \le 0 \end{aligned} \tag{2.11}$$

$$y_i = \pi_z^{bcf} Z_i + \Delta^{bcf} t_i + \rho\sigma_\varepsilon \left[\frac{\phi(\hat{t}_i^*)}{\Phi(\hat{t}_i^*)} t_i - \frac{\phi(\hat{t}_i^*)}{1-\Phi(\hat{t}_i^*)}(1-t_i) \right] + e_i \tag{2.12}$$

where ϕ is the standard normal pdf, Φ is the standard normal cdf, $\hat{t}_i^* = \hat{\delta}_0^{bcf} + \hat{\delta}_h^{bcf} h_i$, and v and ε are bivariate normally distributed as [0,0, σ_v,σ_ε,ρ], with $\sigma_v = 1$ by assumption (it cannot be estimated; Greene 1995), and the "*bcf*" superscript on the parameters denotes that they are estimated using a binary (treatment is binary) control function model, and e is a normally distributed error term. The assumption of bivariate normal error terms, v and ε, is what drives the Heckman model, as it relates the observable information with the unobservable confounders.

The two-step procedure requires estimation of (2.11) as a probit model then generating the bracketed term in (2.12) using the parameter estimates from the probit model. The bracketed term is inserted into (2.7) as an additional regressor, yielding (2.12). The parameters $\rho\sigma_\varepsilon$ are estimated jointly. By adding the bracketed term, ordinary nonlinear least squares can be used to estimate (2.12) since now the unobserved heterogeneity (from the unobservable confounders) causing bias (endogeneity) has been controlled. Note $E[\varepsilon_i \mid t_i, c_i] = 0$, where

$c_i = \left[\frac{\phi(\hat{t}_i^*)}{\Phi(\hat{t}_i^*)} t_i - \frac{\phi(\hat{t}_i^*)}{1-\Phi(\hat{t}_i^*)} (1-t_i) \right]$, because

$E[\varepsilon \mid t = 0] = E[\varepsilon \mid v < \delta_0^{bcf} + \delta_h^{bcf} h_i] = \frac{-\phi(\delta_0^{bcf} + \delta_h^{bcf} h_i)}{1-\Phi(\delta_0^{bcf} + \delta_h^{bcf} h_i)}$ and

$E[\varepsilon \mid t = 1] = E[\varepsilon \mid v > -\delta_0^{bcf} - \delta_h^{bcf} h_i] = \frac{\phi(\delta_0^{bcf} + \delta_h^{bcf} h_i)}{\Phi(\delta_0^{bcf} + \delta_h^{bcf} h_i)}$. As in the case with IV estimation, the OLS standard errors are not the appropriate standard errors for hypothesis testing (see Greene 1995).

Since the bracketed term in (2.12) is a nonlinear function of the regressor in (2.11), identification is satisfied even if all the regressor(s) in (2.11) are also all contained in (2.12). Identification occurs because of the assumed structure of the model. But like IV, the regressors in (2.11) must be uncorrelated with unobservables.

2.2.2 Propensity Score Matching

While IV estimates the true effect an endogenous treatment variable, *t*, has on outcome, it requires a few conditions to be satisfied. It requires the existence of instruments that are closely related to the endogenous explanatory variable but at the same time, uncorrelated with the error term. When there is correlation with the error term, OLS may outperform IV in terms of bias reduction (Bound, Jaeger, and Baker, 1995). In addition, IV requires an identification condition to be met; meaning that for each endogenous variable there must be at least one instrument that does not directly affect the outcome. While the bivariate probit selection model (control function) automatically satisfies the identification condition, this is due to the additional structure imposed. Finally, because traditional IV and control function models require functional form specification of the outcome equation, it has been shown that even minor misspecification will severely bias estimates (Bartels, 1991 and Yatchew and Griliches 1985). Ordinary least squares, instrumental variables, and control function models do allow nonlinear variable functional forms, so long as the functional form is known. If the functional form is unknown, splines or other semiparametric method might be appropriate.

The overwhelming bulk of the propensity score matching (PSM) literature involves a binary treatment and is applied in labor economics and in epidemiological studies (See Rosenbaum and Rubin [1983]; Heckman and Hotz [1989]; Dehejia and Wahba [1999]; Smith and Todd [2005] for a more detailed introduction). Propensity score matching eliminates selection bias by comparing matched participant and non-participant outcomes who have been paired based on their estimated probability of treatment. The difference between propensity score matching and matching on covariates (confounders) is that matching is performed on a single variable (the propensity score), hence the *curse of dimensionality* (Bellman 1961) is eliminated. Propensity score matching also allows for consistent estimation of the treatment effect between comparison groups that are not experimentally controlled (see Rosenbaum and Rubin [1983]; Rosenbaum and Rubin [1985a]; Rosenbaum and Rubin [1985b]; Dehejia and Wahba [1999]; Dehejia and Wahba [2002]). Propensity score matching requires a few conditions to be met in order to produce unbiased treatment effect estimates—strong ignorability (or "selection on observables" per Heckman and Robb [1985] or no unobserved confounders) and that the outcome after a treatment is applied is the same as the potential outcome to treatment (the "stable unit treatment value" assumption).

Propensity score matching (PSM) differs from IV and control function models with regard to data requirements in that the confounding variables must be observable. (OLS also makes this requirement). Unlike traditional IV and control function models (when the treatment variable is continuous), PSM does not require an identification condition to be satisfied. Also, because the treatment effect is estimated using a matching procedure the relationship between outcome, treatment, and the confounders need not be known. In fact, PSM has been found to be robust when the outcome is a nonlinear function (LaLonde 1986; Winship and Mare 1992; Joffe and Rosenbaum 1999; Rubin and Thomas 2000; Imai and van Dyk 2004).

Strong ignorability (Rosenbaum and Rubin 1983). If the outcome-treatment relationship is confounded by *X*, then treatment status is strongly ignorable if the potential outcome (before treatment) is independent of treatment assignment given *X*:

$$\tilde{y}, \tilde{y} + \Delta \perp t \mid X \qquad (2.13)$$

where "⊥" denotes independence. Strong ignorability means that conditional on the confounders, outcome before treatment is independent of (future) treatment status. In context of a linear model (such as equation (2.7)), strong ignorability implies $\varepsilon \perp t \mid X$. Related, strong ignorability requires the Stable-Unit-Treatment-Value-Assumption (SUTVA; Rubin 1990), meaning the outcome of one observation when it receives a treatment is the same as its potential outcome to that treatment before it actually receives the treatment. This means that the observation does not change between the time of assigning and receiving the treatment and also that the treatment of other observations does not change the response of an observation to treatment. Another important requirement that must be satisfied to implement the method is that $0 < pr(t = 1 \mid X) < 1$. It also requires that the probability of treatment to be strictly greater than zero and strictly less than unity (this condition proves important when matching on the propensity score—it ensures the possibility of matches between the treat and untreated observations). Matching on the confounders will eliminate the selection bias by pairing (matching) treated and non-treated observations based on the set of covariates that influence selection—pairs are created between observations with like covariate values. If X is observable and if $\tilde{y} \perp t \mid X$ holds, then:

$$\Delta = E_X\left[E[y \mid t = 1, X] - E[\tilde{y} \mid t = 0, X]\right] \qquad (2.14)$$

where E_X is the expectation operator taken with regard to the distribution of X. Because the interest here is on the treatment effect on the treated, only a weaker form of strong ignorability is needed, $\tilde{y} \perp t \mid X$, rather than $\tilde{y}, \tilde{y} + \Delta \perp t \mid X$ (Dehejia and Wahba 2002).

Note that as the number of covariates increases, it becomes more difficult to match (*the curse of dimensionality*).

Balancing score (Rosenbaum and Rubin 1983). A balancing score, $b(X)$, is defined as:

$$X \perp t \mid b(X)\,. \tag{2.15}$$

The balancing score eliminates the curse of dimensionality because it is a scalar function of the multivariate *X*. The balancing score is not simply any function of *X*; rather, the balancing function, *b*, is specified. Therefore, conditioned on the score, treatment status is independent of the confounders *X*. The notion of balance comes from matching treated and untreated observations based on the balancing score. If a score is a balancing score then the distribution of *X* between the matched treated and untreated observations should be the same or are said to be *balanced*.

The propensity score, $pr(t=1 \mid X) = p(X)$, is a balancing score , i.e.,

$$X \perp t \mid p(X)\,. \tag{2.16}$$

The propensity score is the probability of an observation being treated. In randomized experiments, the propensity score is known. However, in observational studies it is not and must be estimated. The estimated propensity score is only useful so long as it is a balancing score, meaning that conditional on the propensity score, the distribution of confounders between treatment status groups is the same. The novelty of the propensity score is it guides or helps motivate the specification of the balancing score—the probability of treatment given the confounders. Probability models, such as the probit specification, often are used to estimate the propensity score.

Strong ignorability given propensity score (Rosenbaum and Rubin 1983): If treatment status is strongly ignorable given the confounders, then the treatment status is strongly ignorable given the propensity score; that is,

If $\tilde{y}, \tilde{y} + \Delta \perp t \mid X$ and if $0 < pr(t = 1 \mid X) < 1$ for all X,

then $\tilde{y}, \tilde{y} + \Delta \perp t \mid p(X)$ and $0 < pr(t = 1 \mid p(X)) < 1$ for all $p(X)$.

Propensity score matching assumes that an observation's assignment to the program can be expressed as a function of observed variables X, where $0 < pr(t = 1 \mid X) < 1$ (the overlap assumption, per Imbens 2004). Note that the overlap assumption does not include the endpoints (0,1). In a randomized experiment, the researcher chooses the probability of treatment then randomly selects observations, based on this probability of selection, into treatment. If the researcher were to choose 0 or 1 then they would have to either admit no observations into treatment or they would have to admit all observations into treatment, respectively. Choosing either 0 or 1 would not produce a comparison group. With the propensity score method, the researcher does not choose the probability of treatment. Rather, the researcher matches paired observations (treated with control) with the same probability, which is estimated empirically. In theory, a researcher should not find a match between a treated and control pair with a treatment probability of 0 or 1, hence matching is confined to where overlap is possible.

Propensity score matching as an unbiased treatment effect estimator (Rosenbaum and Rubin 1983). Given strong ignorability, the propensity score matched mean treatment effect is unbiased:

$$\begin{aligned}\tilde{\Delta} &= E_X\left[E[y \mid t=1, X] - E[\tilde{y} \mid t=0, X]\right] \\ &= E_{p(X)}\left[E[y \mid t=1, p(X)] - E[\tilde{y} \mid t=0, p(X)]\right] \\ &= E_{p(X)}\left[E[y \mid p(X)] - E[\tilde{y} \mid p(X)]\right] \\ &= \Delta\end{aligned} \tag{2.17}$$

where $E_{p(X)}$ is the expectation operator taken with regard to the distribution of $p(X)$. Thus, given strong ignorability (observable confounders) and the propensity score as a balancing score, the expected treatment effect is equal to the mean difference between the matched participants' and non-participants' outcomes, where observations are matched by their estimated probability of treatment. The importance of score matching versus covariate matching should again be duly noted. Score matching allows for pairing based on a scalar (probability to participate), while covariate matching pairs observations based on vectors of *X*; thus the greater the number of confounders, the more difficult it becomes to match observations (Dehejia and Wahba 2002).

Implementation

In practice, the propensity score is usually estimated by logistic or probit regression. Several pairing/matching techniques exist, but the general form is (Smith and Todd 2005):

$$\tilde{\Delta} = \frac{1}{n_1}\sum_{i=1}^{n_1}\left[y_i - \sum_{j=1}^{n_0} W(i,j)\tilde{y}_j \right] \tag{2.18}$$

where W(i,j) is a weighting function, n_1 is the number of treated observations, n_0 is the number of untreated observations ($n_1 + n_0 = N$). The weighting function depends on the type of matching desired. As summarized by Smith and Todd (2005) several matching procedures are available including nearest neighbor, caliper, interval, and kernel matching[4]. Several matching techniques exist because exact score pairing may not be feasible. The closer matched scores are, however, the more selection bias is eliminated, with exact matches eliminating all selection bias. The researcher may trade bias for variance reduction with several of these matching techniques. Not only is having close matches important for bias reduction, but as Smith and Todd (2005) discuss, low bias will also result if there exists a rich set of confounders that explains treatment selection, if the comparison group comes from the same population as the treated group, and if the outcome variable is measured consistently across both comparison and treated groups. Other weighting schemes exist, as Lunceford and Davidian (2004) present an inverse weighting method that avoids the use of matching altogether.

---

[4] Smith and Todd (2005) find, however, little effect of choice of matching technique on treatment effect estimate.

The balancing condition (the propensity score is a balancing score) is tested by differencing the covariates between matched (based on the estimated propensity score) treated and non-treated observations. If the difference in the covariates is on average zero, then the covariate is deemed balanced. The point is to determine whether, conditioned on the estimated propensity score, the magnitude of covariates differ between treatment groups (tests for multivariate balance are desirable, but in practice rarely performed). If the estimated propensity score is a balancing score, the difference should be zero—conditioned on the estimated propensity score, treatment and the covariate (confounder) should no longer be related. If the estimated propensity score does not balance a covariate, a modified specification of the propensity score model is required. For instance, if a covariate is not balanced, higher order terms of the covariate in question may be added to the propensity score model (see Rosenbaum and Rubin 1984). After re-estimating the propensity score, the balance is re-checked.

2.3 TREATMENT AS A CONTINUOUS VARIABLE

Here the focus is on program evaluation econometric methods when treatment is no longer a binary variable, but instead a continuous one. No longer is the treatment decision to treat or not, but rather how much (intensity).

2.3.1 Regression Approaches

Ordinary Least Squares & Instrumental Variables

The OLS and IV estimators presented in Section 2.2.1 are generalizable to cases where treatment is a continuous variable. The only difference in presentation (from Section 2.2.1) is that treatment, *t*, is a continuous variable. The control function and propensity score methods are different, although the intuition behind each method is the same. Also, two-stage least squares (a special case of IV) is presented, which can be used when several instruments are available for the endogenous treatment variable or in cases where there are several endogenous explanatory variables.

Two-Stage Least Squares

The two-stage least squares (2SLS) estimator for (2.7) is constructed using two estimation steps (2.19) and (2.20):

$$\hat{D} = K(K'K)^{-1}K'D \tag{2.19}$$

$$\begin{bmatrix} \hat{\pi}_z^{2sls} \\ \hat{\Delta}^{2sls} \end{bmatrix} = (\hat{D}'\hat{D})^{-1}\hat{D}'y \tag{2.20}$$

where as before $D = [Z, t]$, $\hat{D} = [Z, \hat{t}\,]$, $K = [Z, h]$, $\hat{t}$ is the fitted value of treatment, the "*2sls*" superscript on the parameters denotes that they are estimated using two-stage least squares, and the instrument, *h*, may include more than one instrument for each endogenous regressor, although here I will assume *h* is still N x 1.

An example of instruments used in 2SLS program evaluation studies is the use of birth date as an instrument for years of schooling (Angrist and Krueger 1991). In Angrist and Krueger (1991), the effect of schooling on earnings is evaluated. Schooling may be correlated with other confounders that also affect earnings (e.g., motivation or status), so birth date quarter is used as an instrument for years of schooling. Because children are required to be enrolled in school the year that they become six, children begin school at different ages (some older than six and others younger), and because students are allowed to dropout once they become 16, some students are required to attend school longer than others. So for high school drop outs, birth quarter is correlated with the number of months of compulsory school attendance—those born in the first quarter are required fewer months in school than those born later—but is unlikely to be correlated with confounders such as motivation or status.

Control Functions

When treatment is a continuous variable, the control function model simplifies into a two-step estimation framework. The difference between the Heckman two-stage control function model and the continuous treatment variable two-stage control function is the construction of the added regressor in (2.11). Treatment is typically estimated as a linear function of the instrument using OLS:

$$t_i = \delta_o^{ccf} + \delta_h^{ccf} h_i + v_i^{ccf} \; . \qquad (2.21)$$

Using the results from (2.21), the outcome equation is estimated as:

$$y_i = \pi_z^{ccf} Z_i + \Delta^{ccf} t_i + \rho\sigma_\varepsilon [t_i - \hat{\delta}_0^{ccf} - \hat{\delta}_h^{ccf} h_i] + e_i^{ccf} \qquad (2.22)$$

where (2.22) is estimated using OLS, the "*ccf*" superscript on the parameters denotes that they are estimated using continuous (treatment is a continuous variable) control function model, and that the bracketed term equals $\hat{v}_i^{ccf}$. The bracketed term, like in the binary case, controls (the conditional expectation of the error in (2.22) is zero) for the endogeneity between treatment and outcome (Garen 1987). Note that 2SLS and control function models differ with respect to the regressors included in the second stage, with the fitted treatment values used in 2SLS and actual treatment and the estimated errors from the treatment equation used in the control function model. When treatment is a continuous variable, the control function treatment effect estimate is numerically equivalent to the two-stage least squares estimate (Garen 1987).

2.3.2 Propensity Score Blocking

The majority of the PSM literature focuses on a binary treatment decision or multi-valued, ordinal treatment (for instance, see Behrman, Cheng, and Todd 2004), but very recent work explores the use of propensity scores with treatment as a continuous variable (Imbens 2000; Lu et al. 2001; Hirano and Imbens 2004; Imai and van Dyk 2004). In the binary treatment case, treated observations are matched with non-treated observations based on their like propensity score (probability of treatment). When treatment is a continuous variable, rather than binary, all of the observations may have received some treatment and differ only with respect to their level of treatment. Unlike the binary case, there is no longer a control (untreated) group to compare with the treated group. However, some authors have taken the approach of discretizing the continuous treatment variable into several ordinal treatment groups (see Behrman, Cheng, and Todd 2004). Of course the boundaries for the groups may be arbitrary and meaningless. Also with a continuous treatment variable, it may be more interesting (and meaningful) to examine the marginal effect of treatment.

When treatment is a continuous variable, matching gives way to modeling the outcome as a function of treatment while conditioning on the propensity score (Imai and van Dyk 2004). The requirements that the propensity score achieve strong ignorability when treatment is a continuous variable necessitates only a slight modification to that shown for matching.

Strong ignorability given propensity score when treatment is a continuous variable (Rosenbaum and Rubin 1983; Imai and van Dyk 2004). If treatment status is strongly ignorable given the confounders, then the treatment status is strongly ignorable given the propensity score; that is,

If $\tilde{y}, \tilde{y} + \Delta \perp t \mid X$ and if $0 < E[t \mid X]$ for all X,

then $\tilde{y}, \tilde{y} + \Delta \perp t \mid E[t \mid X]$ and $0 < E[t \mid E[t \mid X]]$ for all $E[t \mid X]$.

In practical terms, the distribution of outcome is modeled using a parametric form while conditioning on the propensity score. The propensity score is defined as $E[t \mid X]$. Propensity score is also estimated parametrically, but no functional form is assumed for the relationship between the propensity score and the outcome. The two-step method proposed by Imai and van Dyk (2004) applied to (2.7) requires X to be observable (strong ignorability), so treatment can be estimated as a function of X, such as

$$E[t \mid X] = \delta_0^{psb} + \delta_x^{psb} X \tag{2.23}$$

where the δ's are estimated using OLS and the "*psb*" superscript on the parameters denotes that they are parameters in the propensity score blocking model. The linear form of (2.23) is not necessitated by the technique, only that the propensity score fitted from (2.23) is a balancing score. Blocking on the propensity score means that the observations are divided into P blocks (groups or sub-samples) and OLS estimation of y on Z and t are performed within each block. The second stage estimates equation (2.7) while blocking on the propensity score (the fitted values from (2.23)):

$$y_{pj} = \pi_z^{psb} Z_{pj} + \Delta^{psb} t_{pj} + \varepsilon_{pj}^{psb} \tag{2.24}$$

where j references observations contained in the propensity score block p, where $p = 1, \ldots, P$. The average treatment effect is a weighted average (weighted by the number of observations within each block), calculated as:

$$\hat{\Delta}^{psb} = \frac{1}{N} \sum_{p=1}^{P} \hat{\Delta}_p^{psb} n_p \tag{2.25}$$

where n_p is the number of observations in block p.

When treatment is a continuous variable, the balancing condition $X \perp t \mid p(X)$ is difficult to test in the manner used when treatment is a binary variable. This is because treatment may take on any number of values instead of only two. Imai and van Dyk (2004) explain that, if the propensity score is constructed as a linear function of the confounders, then the balancing condition may be tested by regressing a function of each of the continuous confounders on the treatment variable and the propensity score. Similar to the binary treatment case, the balancing test is a test of the propensity score model specification. A function of each confounder is used to test the propensity score specification, because by construction, regressing the untransformed covariate (the form used in the propensity score model) on treatment and the propensity score will yield an insignificant parameter estimate on treatment (if using simple linear regression). If the estimated parameter for treatment effect is significant, then treatment and the covariate are related and the specified propensity score is not a balancing score (the propensity score model is misspecified). However, any insignificance does not rule out the possibility that higher order transformations of the covariate are related to treatment (or a Type II error occurred in the test).

2.3.3 Control Score Blocking

Both propensity score matching and propensity score blocking require the confounders to be observable in order to ensure strong ignorability. This is a major point of departure from IV, 2SLS, and control function methods. Ordinary least squares also requires strong ignorability, but also that the modeler makes functional form assumptions. If the confounders are unobservables (or immeasurable), standard propensity score methods cannot be used, as the propensity score will be misspecified, if it can be estimated at all.

I propose a new technique, a modification of the propensity score approach that will produce consistent treatment effect estimates if the exclusion restriction used with IV, 2SLS, and control function holds (an instrument exists that is uncorrelated with the unobservable confounders).

As mentioned, the weakness of the propensity score based approaches is their (non-testable) requirement for observable confounders, which makes some prefer IV or control function approaches (e.g., Heckman and Navarro-Lozano 2004). Because the confounders are unobservable, the propensity score (the probability of treatment) cannot be a balancing score because it cannot be correctly specified. However, using the control function assumption that unobservable confounders are functions of observable variables, the control regressor (the bracketed term in equation (2.22)) is a balancing score. From the control function first stage model (equation (2.21)), the estimated error (the bracketed term in equation (2.22)) is constructed as:

$$\hat{v}_i^{ccf} = t_i - \hat{\delta}_o^{ccf} - \hat{\delta}_h^{ccf} h_i . \qquad (2.26)$$

The variable $\hat{v}^{ccf}$ is what I term the *control score*.

Control score as a balancing score. The control score is a balancing score if *X* and *t* are independent given the control score (*v*), $X \perp t \mid v$. If $P(X \mid t, v) = P(X \mid v)$, then *X* and *t* are independent given the control score, where *P* denotes probability distribution. Suppose treatment is a function of three sources of variation—unobserved confounders (*X*), an instrument (*h*), and random error (*u*)—all of which are uncorrelated with the others, then treatment can be represented as:

$$t_i = \delta_0 + \delta_h h_i + \delta_x X_i + u_i \tag{2.27}$$

$$t_i = \delta_0 + \delta_h h_i + v_i, \quad v_i = \delta_x X_i + u_i \tag{2.28}$$

Let $\hat{t}$ be the fitted value obtained by regressing *t* on *h* (2.28), and the control score be the residual $\hat{v}_i = t_i - \hat{t}_i$. Note that $\hat{v}$ estimates *v*. Given (2.27) and (2.28) it can be shown that the control score is a balancing score,

$$\begin{aligned} P(X \mid t, v) &= P(X \mid \delta_0 + \delta_h h + v, v) \\ &= P(X \mid \delta_0 + \delta_h h, v) \\ &= P(X \mid v) \quad if \;\; X \perp h \end{aligned} \tag{2.29}$$

Thus, the control score is a balancing score.

Strong ignorability given control score when treatment is a continuous variable. Strong ignorability given X implies strong ignorability given any balancing score.

The propensity score blocking framework remains the same, except that the propensity score is replaced with the control score. The second stage of the control score approach (first stage is equivalent to the first stage control function model) estimates equation (2.7) while blocking on the control score (the estimated error from (2.21)):

$$y_{cj} = \pi_z^{csb} Z_{cj} + \Delta^{csb} t_{cj} + \varepsilon_{cj}^{csb} \tag{2.30}$$

where j references observations contained in the control score block, c, where $c = 1, \ldots, C$, and the "*csb*" superscript on the parameters denotes that they are estimated using control score blocking. Again, the average treatment effect is a weighted average (weighted by the number of observations within each block), calculated as:

$$\hat{\Delta}^{csb} = \frac{1}{N} \sum_{p=1}^{C} \hat{\Delta}_c^{csb} n_c \tag{2.31}$$

where n_c is the number of observations in block c. The advantage of using the control score approach over the propensity score approach is that the confounders do not need to be observed. The advantage of the control score approach over the control function approach is, like the propensity score blocking approach, that it is robust to misspecification of the outcome equation.

2.4 DISCUSSION

Choosing the appropriate program evaluation econometric technique largely depends on the dataset available and the distribution of the treatment variable (binary or continuous). When an endogenous treatment variable exists due to unobserved or misspecified confounders, OLS estimation is biased and other estimation methods are preferred. Propensity score methods are superior to OLS when the confounders are observed and the functional form of the outcome equation is unknown. The problem with propensity score and OLS methods is knowing with certainty if the confounders are truly observable, as it is not testable with empirical data. Instrumental variables, two-stage least squares, and control function models all allow for unobservable confounders, so long as the identification condition is satisfied. Identification is satisfied only if valid instruments exist when confounders are unobservable. I have presented a new technique, control score blocking, that combines the strengths of propensity score blocking (robust to confounder misspecification) with the strengths of IV, 2SLS, and control function methods (confounders do not need to be observed) and offers a viable alternative to the traditional program evaluation approaches discussed above.

The degree to which each of the methods outperform the others (in terms of unbiasedness of the treatment effect estimate) depends on the strength of endogeneity, instrument quality (correlation between treatment and the instrument), instrument validity (correlation between the instrument and the unobserved confounders), and misspecification of the confounders. These technique are often presented using simplifying assumptions, for instance, PSM assumes that confounders are all observable, whereas IV assumes that instruments are strong and valid. What is more interesting and for empirical work, more realistic, is examination, in relative terms, when some of these assumptions hold and others do not. This is the subject of Section 3.

SECTION REFERENCES

Angrist, J.D. 1998. "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants." *Econometrica*, 66(2): 249-288.

Angrist, J.D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *The American Economic Review*, 80(3): 313-336.

Angrist, J.D., G.W. Imbens, and D.B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, 91(434): 444-455.

Angrist, J.D. and A.B. Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives*, 15(4): 69-85.

Angrist, J.D. and A.B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, 4: 979-1014.

Angrist, J.D. and V. Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *The Quarterly Journal of Economics*, May: 533-575.

Bartels, L.M. 1991. "Quasi-Instrumental Variables." *American Journal of Political Science*, 35, 777-800.

Bellman, R.E. 1961. *Adaptive Control Processes*. Princeton, N.J.: Princeton University Press.

Bound, J., Jaeger, D.A., and Baker, R.M. 1995. "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association*, 90, 443-450.

Behrman, J.R., Cheng, Y., and Todd, P.E. 2004. "Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach." *The Review of Economics and Statistics*, 86, 108-122.

Dawid, A.P. 1979. "Conditional Independence in Statistical Theory." *Journal of the Royal Statistical Society, Series B (Methodological)* 41(1): 1-31.

Dehejia, R.H. and S. Wahba. 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies," *The Review of Economics and Statistics*, 84(1): 151-161.

Dehejia, R.H. and Wahba, S. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of American Statistical Association*, 94, 1053-1062.

Garen, J.E. 1987. “Relationship Among Estimators of Triangular Econometric Models.” *Economic Letters*, 25: 39-41.

Greene, W.H. 2000. Econometric Analysis. Upper Saddle River, New Jersey, 1004 pages.

Greene, W.H. 1995. Limdep Version 7.0. Econometric Software, Inc, Plainview, 850 pages.

Heckman, J.J. 1978. “Dummy Endogenous Variables in a Simultaneous Equation System.” *Econometrica*, 46(4): 931-959.

Heckman, J.J. and Hotz, V.J. 1989. “Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training.” *Journal of the American Statistical Association*, 84, 862-874.

Heckman, J. and S. Navarro-Lozano. 2004. “Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models.” *The Review of Economics and Statistics*, 86(1): 30-57.

Heckman, J.J. and R. Robb. 1985. Alternative Methods for Evaluating the Impact of Interventions. In Heckman, J., Singer, B. (Eds.), Longitudinal Analysis of Labor Market Data. Cambridge University Press, New York, pp.156-246.

Hirano, K. and Imbens, G.W. 2004. The Propensity Score with Continuous Treatment. Draft of Chapter for Missing Data and Bayesian Methods in Practice: Contributions from Donald Rubin’s Statistical Family, Forthcoming from Wiley.

Imai, K. and van Dyk, D.A. 2004. “Causal Inference with General Treatment Regimes: Generalizing the Propensity Score.” *Journal of the American Statistical Association*, 99, 854-866.

Imbens, G.W. 2004. “Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review”, The *Review of Economics and Statistics* 86(1):4-29.

Imbens, G.W. 2000. “The Role of the Propensity Score in Estimating Does-Response Functions,” *Biometrika* 87(3): 706-710.

Joffe, M.M. and Rosenbaum, P.R. 1999. “Propensity Scores.” *American Journal of Epidemiology*, 150, 327-333.

LaLonde, R. 1986. “Evaluating the Econometric Evaluations of Training Programs with Experimental Data.” *American Economic Review*, 76, 604-620.

Lu, B., Zanutto, E., Hornik, R., and Rosenbaum, P.R. 2001. “Matching with Dose in an Observational Study of a Media Campaign Against Drug Abuse.” *Journal of the American Statistical Association*, 96, 1245-1253.

Lunceford, J.K. and M. Davidian. 2004. “Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study.” *Statistics in Medicine*, 23: 2937-2960.

Rosenbaum, P.R. and D.B. Rubin. 1985a. “The Bias Due to Incomplete Matching,” *Biometrics*, 41: 103-116.

Rosenbaum, P.R. and D.B. Rubin. 1985b. “Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score,” *The American Statistician*, 39(1): 33-38.

Rosenbaum, P.R. and Rubin, D.B. 1984. “Reducing Bias in Observational Studies Using Subclassification on the Propensity Score.” *Journal of the American Statistical Association*, 79, 516-524.

Rosenbaum, P.R. and Rubin, D.B. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika*, 70, 41-55.

Rubin, D.B. 1990. “Formal Modes of Statistical Inference for Causal Effects.” *Journal of Statistical Planning and Inference*, 25: 279-292.

Rubin, D.B. and Thomas, N. 2000. “Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates.” *Journal of the American Statistical Association*, 95, 573-585.

Smith, J. and Todd, P. 2005. “Does Matching Overcome LaLonde’s Critique of Nonexperimental Estimators?” *Journal of Econometrics*, 125, 305-353.

Winship, C. and Mare, R.D. 1992. “Models for Sample Selection Bias.” *Annual Review of Sociology*, 18, 327-350.

Yatchew, A., and Z. Griliches. 1985. “Specification Error in Probit Models.” *Review of Economics and Statistics* 67(1): 134-139.

3. SIMULATION OF AN ENDOGENOUS TREATMENT WHEN TREATMENT IS A CONTINUOUS VARIABLE

3.1 SIMULATION INTRODUCTION

The objective of this Monte Carlo simulation is to quantify, in terms of bias and mean squared error, the ability of ordinary least squares (OLS), two-stage least squares (2SLS)[5], propensity score blocking (PSB), and control score blocking (CSB) to estimate the treatment effect of a program when treatment is endogenous and measured as a continuous variable. The design of the simulation allows comparison of the modeling approaches under a host of scenarios, including variations in the strength of endogeneity, instrument quality, instrument validity, and functional form assumptions. While the properties of the estimators are well established under particular assumptions (see Section 2.3) what is more difficult to evaluate without simulation is their performance when some of the assumptions hold and others do not. The performance of OLS, 2SLS, PSB, and CSB estimators, in both absolute and relative terms, depends on the strength of endogeneity, instrument quality, and instrument validity.

Endogeneity

If there is zero correlation between the level of treatment and the outcome error, there is by definition no endogeneity, and OLS estimation of outcome, as a function of treatment, will produce an unbiased treatment effect estimate. However, endogeneity implies that the treatment and the outcome error are correlated, thus biasing the OLS treatment effect estimates, with bias increasing as a function of endogeneity.

---

[5] When treatment is a continuous variable, the two-stage least square estimator is equivalent to the control function estimator (Garen 1987; see Section 2.3.1).

Instrument Quality

Instrument quality describes the strength of the relationship between the endogenous treatment variables and the instrument. The quality of the instrument is measured as its correlation with the endogenous treatment variable. In finite samples the magnitude of bias of the instrumental variable estimator approaches the magnitude of bias of the OLS estimator when the correlation between instrumented endogenous variable and its instrument (or the R^2 from regressing the instrumented endogenous variable on a set of instruments) approaches 0 (Bound et al. 1995). This implies that the performance (in terms of statistical bias) of the instrumental variable estimator, as compared to the OLS estimator, depends on instrument quality.

Instrument Validity

Instrument validity is related to the correlation between the instrument and the outcome error, with a valid instrument being orthogonal to the outcome error. While the instrumental variable estimator is less biased than OLS estimator when the instrument is valid, weak instruments (low instrument quality) can lead to a less biased OLS estimator compared to the instrumental variable estimator when there is even a slight correlation between the instrument and the outcome equation error (Bound et al. 1995).

In general, the inconsistency (large sample bias) of the 2SLS (IV) estimator relative to OLS estimator tends to (Bound et al. 1995):

$$\frac{\text{plim}\,(\hat{\Delta}^{iv} - \Delta)}{\text{plim}\,(\hat{\Delta}^{ols} - \Delta)} = \frac{\sigma_{\hat{t}\varepsilon} / \sigma_{t\varepsilon}}{R_{th}^2} \tag{3.1}$$

where $\hat{\Delta}^{iv}$ is the 2SLS (IV) treatment effect estimate, $\hat{\Delta}^{ols}$ is the OLS treatment effect estimate, $\sigma_{\hat{t}\varepsilon}$ is the covariance between the fitted value of the instrumented endogenous explanatory variable *t* (produced using the first stage 2SLS estimates) and the outcome error, ε, $\sigma_{t\varepsilon}$ is the covariance between *t* and ε, and R_{th}^2 is the coefficient of determination[6], between *t* and the instrument, *h*, from regressing *t* on the instrument. There need not be only one instrument used in 2SLS[7]. However, if there is only one instrument, then the 2SLS-OLS large sample bias ratio becomes (Bound et al. 1995):

$$\frac{\text{plim}\,(\hat{\Delta}^{iv} - \Delta)}{\text{plim}\,(\hat{\Delta}^{ols} - \Delta)} = \frac{\rho_{h\varepsilon} / \rho_{t\varepsilon}}{\rho_{th}} \tag{3.2}$$

where $\rho_{t\varepsilon}$ is the correlation between treatment and the outcome error (endogeneity), $\rho_{h\varepsilon}$ is the correlation between the instrument and the outcome error (instrument validity), and ρ_{th} is the correlation between the treatment and the instrument for treatment (instrument quality). If the instrument is correlated with the error $(\rho_{h\varepsilon} > 0)$, then OLS outperforms 2SLS if

[6] In two-stage least squares estimation when construction of the instrument for treatment is a modeled as a function of instruments and common exogenous variables (exogenous variables shared across the treatment and outcome equation), then the R_{th}^2 is the partial coefficient of determination between treatment and the instruments (Bound et al. 1995).
[7] In 2SLS, bias is proportional to the number of instruments used in excess of the number of endogenous regressors, i.e., the degree of overidentification (Angrist and Krueger 2001).

$(\rho_{th} < \rho_{h\varepsilon} / \rho_{t\varepsilon})$. Thus weak, invalid instruments should be avoided.

While the relative performance of OLS and 2SLS estimators are known given varying levels of endogeneity, instrument quality, and instrument validity, the present paper goes beyond these results to compare the performance of the traditional techniques versus newer, propensity-score based, program evaluation econometric techniques. Conceptually, the simulation parameterizes the level of endogeneity, instrument quality, and instrument validity underlying the generated dataset and examines the bias and root mean squared error of the OLS, 2SLS, PSB, and CSB estimators. In addition, it confirms the previous OLS/2SLS findings, but also extends these findings by describing when PSB and CSB are the least biased estimator.

3.2 SIMULATION DESIGN

The simulation is designed to assess the absolute and relative abilities of OLS, 2SLS, PSB, and CSB estimators to measure the treatment effect of a continuously distributed, endogenous variable on outcome, while varying the magnitude of endogeneity, instrument quality, instrument validity, and information "known" at the time of the model estimation (whether the confounders are observable and the true functional form). The PSB and CSB methods are nuances of a similar technique, regression with blocking performed on a balancing score. Propensity score blocking is used when the confounders are observable; the propensity score is a balancing score. Control score blocking is used when the confounders are unobservable; thus the control score is a balancing score (see Sections 2.3.2 and 2.3.3).

The simulation is defined by two structural equations, the distributions of the randomly generated variables, and the information known at time of estimation.

Structural Equations

(*outcome*) $$y_i = 1 + t_i + z_i + \varepsilon_i \tag{3.3}$$

(*outcome error*) $$\varepsilon_i = x_i + e_i \tag{3.4}$$

where y is outcome, t is treatment, z is an exogenous variable, x is a variable that if correlated with treatment confounds the outcome-treatment relationship, and e is a normally distributed error term. Both z and e are uncorrelated with the other random variables in the system.

Data Generation and Information Set

The random variables are generated based on the following distribution:

(*data distribution*) $$\begin{bmatrix} t_i \\ h_i \\ x_i \\ z_i \\ e_i \end{bmatrix} \sim N[0, \Sigma], \quad \Sigma = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & \cdot \\ \rho_{th} & 1 & \cdot & \cdot & \cdot \\ \rho_{tx} & \rho_{hx} & 1 & \cdot & \cdot \\ 0 & 0 & 0 & 1 & \cdot \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3.5}$$

These possible information sets are considered:

$$
(\textit{information set}) \qquad \begin{aligned} \Omega_1 &= [y, t, z, h, x] \\ \Omega_2 &= [y, t, z, h, x^*] \\ \Omega_3 &= [y, t, z, h], \end{aligned} \tag{3.6}
$$

where h is an instrument for treatment, ρ_{tx} is the correlation between t and x, ρ_{th} is the correlation between t and h, ρ_{hx} is the correlation between h and x, and $x^* = e^x$. The three information sets list the variables that are "known" at the time of estimation, so that the models are estimated conditioned on the information set available. This is done to allow the simulations to vary with regard to observable confounder (Ω_1), confounder misspecification (Ω_2), and unobservable confounder (Ω_3). Confounder misspecification implies that while the confounder is observable, it is misspecified in any parametric estimation. The magnitude of the correlations are varied across the simulations and are determined by the set levels of endogeneity ($\rho_{t\varepsilon}$), instrument quality (ρ_{th}), and instrument validity ($\rho_{h\varepsilon}$). Because

$$
\rho_{t\varepsilon} = \frac{\sigma_{t\varepsilon}}{\sigma_t \sigma_\varepsilon} \qquad \text{and} \qquad \rho_{h\varepsilon} = \frac{\sigma_{h\varepsilon}}{\sigma_h \sigma_\varepsilon} \tag{3.7}
$$

it follows that,

$$
\rho_{t\varepsilon} = \frac{\sigma_{tx} + \sigma_{te}}{\sigma_t \sqrt{\sigma_x^2 + \sigma_e^2}} = \frac{\rho_{tx}\sigma_t\sigma_x + \rho_{te}\sigma_t\sigma_e}{\sigma_t \sqrt{\sigma_x^2 + \sigma_e^2}} = \frac{\rho_{tx}}{\sqrt{2}}, \tag{3.8}
$$

$$
\rho_{h\varepsilon} = \frac{\sigma_{hx} + \sigma_{he}}{\sigma_h \sqrt{\sigma_x^2 + \sigma_e^2}} = \frac{\rho_{hx}\sigma_h\sigma_x + \rho_{he}\sigma_h\sigma_e}{\sigma_h \sqrt{\sigma_x^2 + \sigma_e^2}} = \frac{\rho_{hx}}{\sqrt{2}} \tag{3.9}
$$

by (3.5), where σ_t is the standard deviation of t, σ_ε is the standard deviation of ε $(\sigma_\varepsilon^2 = \sigma_x^2 + \sigma_e^2)$, σ_h is the standard deviation of h, σ_x is the standard deviation of x, σ_e is the standard deviation of e, σ_{tx} is the covariance between t and x, $\sigma_{t\varepsilon}$ is the covariance between t and ε, σ_{te} is the covariance between t and e, σ_{hx} is the covariance between h and x, $\sigma_{h\varepsilon}$ is the covariance between h and ε, σ_{he} is the covariance between h and e, ρ_{te} is the correlation between t and e, and ρ_{he} is the correlation between h and e.

In total, 72 Monte Carlo simulations are conducted—36 simulations based on 1,000 samples with 1,000 observations each, and 36 simulations based on 1,000 samples with 100 observations each. Within each 36-simulation set, there are three information set variations (observable confounder, confounder misspecification, unobservable confounder), three variations of the treatment and error correlation (0.40, 0.10, 0), two variations of the treatment and instrument correlation (0.40 and 0.20), and two variations of the instrument and error correlation (0.10 and 0). After generating the simulated data from equations (3.3) and (3.4) based on (3.5), the specified endogeneity, instrument quality, instrument validity, and sample size, the OLS, 2SLS, PSB, and CSB models are estimated conditioned on the appropriate information set. The estimated treatment effect is recorded for each sample.

3.3 MODEL SPECIFICATION

The OLS, 2SLS, PSB, CSB estimators of the treatment effect are described below. The specification of the models is dependent on the information set observed.

Ordinary Least Squares

Three OLS specifications of (3.3) are used depending on the information set:

$$
\begin{aligned}
&y_i \mid \Omega_1 = \beta_0^{ols1} + \beta_t^{ols1} t_i + \beta_z^{ols1} z_i + \beta_x^{ols1} x_i + \varepsilon_i^{ols1} \\
&y_i \mid \Omega_2 = \beta_0^{ols2} + \beta_t^{ols2} t_i + \beta_z^{ols2} z_i + \beta_x^{ols2} x_i^* + \varepsilon_i^{ols2} \ . \qquad (3.10) \\
&y_i \mid \Omega_3 = \beta_0^{ols3} + \beta_t^{ols3} t_i + \beta_z^{ols3} z_i + \varepsilon_i^{ols3}
\end{aligned}
$$

Two Stage Least Squares

Three 2SLS specifications of (3.3) are used depending on the information set. The first stage of 2SLS is OLS estimation of treatment as a function of instruments:

$$
\begin{aligned}
&t_i \mid \Omega_1 = \alpha_0^{2sls1} + \alpha_h^{2sls1} h_i + \alpha_z^{2sls1} z_i + \alpha_x^{2sls1} x_i + v_i^{2sls1} \\
&t_i \mid \Omega_2 = \alpha_0^{2sls2} + \alpha_h^{2sls2} h_i + \alpha_z^{2sls2} z_i + \alpha_x^{2sls2} x_i^* + v_i^{2sls2} \ . \qquad (3.11) \\
&t_i \mid \Omega_3 = \alpha_0^{2sls3} + \alpha_h^{2sls3} h_i + \alpha_z^{2sls3} z_i + v_i^{2sls3}
\end{aligned}
$$

The second stage models outcome (3.3) while instrumenting for the endogenous treatment variable:

$$
\begin{aligned}
y_i \mid \Omega_1 &= \beta_0^{2sls1} + \beta_t^{2sls1}\hat{t}_i^1 + \beta_z^{2sls1} z_i + \beta_x^{2sls1} x_i + \varepsilon_i^{2sls1} \\
y_i \mid \Omega_2 &= \beta_0^{2sls2} + \beta_t^{2sls2}\hat{t}_i^2 + \beta_z^{2sls2} z_i + \beta_x^{2sls2} x_i^* + \varepsilon_i^{2sls2} \\
y_i \mid \Omega_3 &= \beta_0^{2sls3} + \beta_t^{2sls3}\hat{t}_i^3 + \beta_z^{2sls3} z_i + \varepsilon_i^{2sls3}
\end{aligned} \tag{3.12}
$$

where $\hat{t}_i^1, \hat{t}_i^2, \hat{t}_i^3$ are the instruments (fitted values) from (3.11) depending on information set $\Omega_1, \Omega_2, \Omega_3$, respectively.

Propensity Score Blocking

When the confounder is observable (Ω_1) or the confounder is believed to be observable, but is mistakenly misspecified (Ω_2), the propensity score is specified as:

$$
\begin{aligned}
t_i \mid \Omega_1 &= \alpha_0^{psb1} + \alpha_x^{psb1} x_i + v_i^{psb1} \\
t_i \mid \Omega_2 &= \alpha_0^{psb2} + \alpha_x^{psb2} x_i^* + v_i^{psb2}
\end{aligned} \; . \tag{3.13}
$$

Using OLS, equation (3.13) is estimated by blocking (3.3) on the estimated propensity score (see Section 2.3.2 for blocking details):

$$
\begin{aligned}
y_{pj} \mid \Omega_1 &= \beta_{0p}^{psb1} + \beta_{tp}^{psb1} t_{pj} + \beta_{zp}^{psb1} z_{pj} + \beta_{xp}^{psb1} x_{pj} + \varepsilon_{pj}^{psb1} \\
y_{pj} \mid \Omega_2 &= \beta_{0p}^{psb2} + \beta_{tp}^{psb2} t_{pj} + \beta_{zp}^{psb2} z_{pj} + \beta_{xp}^{psb2} x_{pj}^* + \varepsilon_{pj}^{psb2}
\end{aligned} \tag{3.14}
$$

where j references observations contained in the propensity score block p. For the large sample simulations (1,000 observation sample size), 20 block groups are used, based on propensity score centiles, whereas for the small sample simulations (100 observation sample size), two block groups are used. The numbers of blocks are chosen so that 50 observations are contained within each block, regardless of sample size. The average treatment effect is a weighted average (weighted by the number of observations within each block), calculated as:

$$\hat{\beta}_t^{psb1} = \frac{1}{N}\sum_{p=1}^{P}\hat{\beta}_{tp}^{psb1} n_p, \quad \hat{\beta}_t^{psb2} = \frac{1}{N}\sum_{p=1}^{P}\hat{\beta}_{tp}^{psb2} n_p \tag{3.15}$$

where n_p is the number of observations in block p. Note that the propensity score blocking will only produce unbiased results if the estimated propensity score is a balancing score, so if modeling treatment as a function of x^* (when information set Ω_2 is known) does not balance the true confounder x, then propensity score blocking will be biased.

Control Score Blocking

As discussed in Section 2.3.3, I offer control score blocking as a new program evaluation econometric approach that combines 2SLS and control function methods with propensity score blocking. When the confounder is unobservable (Ω_3), a propensity score cannot be estimated, but a control score can. The control score is the residual from regressing treatment on an instrument:

$$t_i \mid \Omega_3 = \alpha_0^{csb3} + \alpha_h^{csb3} h_i + v_i^{csb3} \tag{3.16}$$

The residual is a function of the unobservable *x*, so conditioning on the residuals is similar to conditioning on the unobservable. Again, outcome is modeled using OLS, with blocking on the estimated propensity score residual (20 blocks for the large sample simulations; 2 blocks for the small sample simulations):

$$y_{rj} = \beta_{0r}^{csb3} + \beta_{tr}^{csb3} t_{rj} + \beta_{zr}^{csb3} z_{rj} + \varepsilon_{rj}^{csb3} \tag{3.17}$$

where *j* references observations contained in the propensity score residual block, *r*. The average treatment effect is a weighted average, with the weighting scheme similar to that described with propensity score blocking:

$$\hat{\beta}_t^{csb3} = \frac{1}{N} \sum_{c=1}^{C} \hat{\beta}_{tc}^{csb3} n_c \tag{3.18}$$

where n_c is the number of observations in block *c*.

3.4 ESTIMATOR EVALUATION METHODS

Each of the 72 simulations is evaluated using the mean bias and root mean squared error of the treatment effect estimates produced over the 1,000 samples.

Mean Bias

The true treatment effect is unity (per (3.3)), so that

$$(\text{Mean Bias})^{q} = \frac{1}{M}\sum_{i=1}^{M}(\hat{\beta}_{ti}^{q} - 1) \qquad (3.19)$$

where M indexes the 1,000 samples and q indexes the modeling approaches (OLS, 2SLS, PSB, and CSB).

Root Mean Squared Error

Dispersion is measured by the RMSE:

$$(\text{RMSE})^{q} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(\hat{\beta}_{ti}^{q} - 1)^{2}}\ . \qquad (3.20)$$

3.5 SIMULATION RESULTS

The large sample results (1,000 observations sample size) are presented in tables 3.1 – 3.3 and the small sample results (100 observations sample size) are presented in tables 3.4 – 3.6. Note that because the true treatment effect is set to unity the mean bias equals percent mean bias divided by 100. Table 3.7 compares the expected ratios of the inconsistency of the two-stage least squares estimator relative to the ordinary least squares estimator (from equation (3.2)) with the inconsistency ratios derived when sample sizes are 1,000 and 100 and when the confounder is misspecified (Ω_2) or unobservable (Ω_3). This comparison provides insight into whether the sample sizes are large enough in the simulations to gauge the asymptotic properties of the estimators. The inconsistency ratios are not calculated for the simulations when the confounder is observable (Ω_1), there is no endogeneity ($\rho_{t\varepsilon} = 0$), or there is no instrument validity ($\rho_{h\varepsilon} = 0$) because either the OLS estimate is unbiased (when the confounder is observable or no endogeneity exists) or the inconsistency ratio is undefined (no instrument validity). The inconsistency ratios (both for the 1,000 and 100 observation simulations) for the simulated data appear to approach the expected ratios only when instrument quality (ρ_{th}) is greater than instrument validity ($\rho_{h\varepsilon}$).

3.5.1 Large Sample Results

Observable Confounder (Ω_1) & N=1,000

The results are presented in table 3.1. If the confounder is observable and the confounder is correctly specified in the outcome equation (Ω_1), then there is no endogeneity (regardless of the level of $\rho_{t\varepsilon}$ there is no endogeneity because of Ω_1) and ordinary least squares is unbiased (population mean treatment effect < 1% of the true parameter). Two-stage least squares are also generally unbiased (the population mean treatment effect is within 1% of the true parameter for simulations 1-11; simulation 12 is within 2%), but have larger RMSE's. As a matter of procedure, the 2SLS model uses the confounder, if observed, in the construction of the instrument (fitted treatment value), so even if *h* is endogenous, the fitted treatment value is not. Propensity score blocking is also unbiased, although OLS has slightly smaller RMSE.

Confounder Misspecification (Ω_2) & N=1,000

The results are presented in table 3.2. If the confounder is observable, but misspecified the OLS estimator is biased. The level of bias is seen to increase as a function of endogeneity

($\rho_{t\varepsilon}$). For low levels of endogeneity ($\rho_{t\varepsilon}$ = 0.1), the OLS estimator exhibits 6% bias and for moderate levels of endogeneity ($\rho_{t\varepsilon}$ = 0.4), 28% bias. The 2SLS estimator is unbiased when the instrument *h* is exogenous ($\rho_{h\varepsilon}$ = 0), with the level of accuracy (population mean treatment effect closer to the true parameter) increasing with instrument quality (ρ_{th}). The 2SLS estimator outperforms the OLS estimator even when the instrument is weakly correlated with the confounder ($\rho_{h\varepsilon}$ = 0.1), so long as the instrument quality is strong (ρ_{th} = 0.4) and endogeneity high ($\rho_{t\varepsilon}$ = 0.4). In this case, the 2SLS estimator exhibits 16% bias versus 28% for OLS. If instead, endogeneity is low ($\rho_{t\varepsilon}$ = 0.1), the OLS estimator outperforms the 2SLS estimator, although OLS is still biased (6% versus 14%) Ordinary least squares also outperforms 2SLS when instrument quality is weak (ρ_{th} = 0.2) and the instrument is weakly correlated with the confounder ($\rho_{h\varepsilon}$ = 0.1). In these instances, the bias of the OLS estimator ranges from 6% and 28% versus 30% and 37% for the 2SLS estimator. The relative performance of the 2SLS estimator versus the OLS is consistent with Bound et al. result described before. The PSB estimator appears robust to misspecification and is unbiased regardless of strength of endogeneity. In addition, the RMSE associated with the PSB estimator are also the smallest.

Unobservable Confounder (Ω_3) & N=1,000

The results are presented in table 3.3. If the confounder is unobservable, the OLS estimator behaves similarly to the instances when the confounder is misspecified. The OLS estimator is biased by 14% when endogeneity is weak ($\rho_{t\varepsilon}$ = 0.1) versus 56% when endogeneity is moderate ($\rho_{t\varepsilon}$ = 0.4). The 2SLS results relative to the OLS results are similar to instances where the confounder is misspecified and included in the 2SLS model. However, some information is gained by including the misspecified confounder in the 2SLS model, as the biases are lower than when it is not. While standard propensity score methods require selection on observables for identification of treatment effect, table 3.3 shows conditioning on the residuals (CSB) estimated after regressing treatment on its instrument is a balancing score if the instrument is valid. The CSB estimator performs similarly to the 2SLS estimator, the only notable exception being at moderate levels of endogeneity ($\rho_{t\varepsilon}$ = 0.4) where the CSB estimator bias is slightly larger. The difference appears to be the result of a combination of the strength of endogeneity and number of block-groups used in the CSB regressions (20 groups were used). Increasing the number of blocks reduces the bias (analysis not shown).

The number of blocks used in the score blocking regressions should affect the bias in the population mean treatment effect. This makes sense because the blocking objective is to eliminate the endogeneity by grouping 'like' observations. With fewer blocks there stands to be a greater range of scores within each block, implying greater heterogeneity of the confounder within each block. Using more, smaller (in number of observations and score range) blocks increases the homogeneity of the confounder within the block. Thus, using more blocks reduces the amount of bias in the treatment effect estimate.

3.5.2 Small Sample Results

Observable Confounder (Ω_1) & N=100

The results are presented in table 3.4. Similar to the situation with large sample sizes, the OLS and PSB estimators are unbiased (population mean treatment effect < 1% of the true parameter). While RMSE's are larger than in the large sample size results (as would be expected), they exhibit the same pattern (OLS RMSE being smaller than PSB RMSE). The 2SLS estimator is generally unbiased, but there is quite a bit of variation across simulations regarding bias and RMSE when the instrument is weak. The performance of the 2SLS estimator is erratic when there is a weak instrument. This holds across all information sets.

Confounder Misspecification (Ω_2) & N=100

The results are presented in table 3.5. The OLS results are generally similar to those when the sample size is 1000, although the RMSE are larger. The PSB estimator is unbiased unless endogeneity is moderate ($\rho_{t\varepsilon}$ = 0.4), but the RMSE is larger than in the 1,000 observation simulation. The PSB estimator is biased by 4%. This is a departure from the large sample size result and demonstrates PSB robustness to imbalanced confounder with larger sample sizes (see Section 2.2.2 and Section 2.3.2 for a discussion of confounder balance). The PSB estimator outperforms (or generally does not perform any worse than) the OLS and 2SLS estimators in all instances when the confounder is misspecified. As noted above, the 2SLS estimator performs quite erratically when there is a weak instrument. The 2SLS estimator only outperforms the OLS estimator when (1) the instrument is strong (ρ_{th} = 0.4) and valid ($\rho_{h\varepsilon}$ = 0), (2) the instrument is strong (ρ_{th} = 0.4) and endogeneity is moderate ($\rho_{t\varepsilon}$ = 0.4), and (3) the instrument is weak (ρ_{th} = 0.2) and valid ($\rho_{h\varepsilon}$ = 0), and endogeneity is moderate ($\rho_{t\varepsilon}$ = 0.4).

Unobservable Confounder (Ω_3) & N=100

The results are presented in table 3.6. Again, the OLS results are generally similar to those when the sample size is 1000, although the RMSE are larger. The CSB estimator performs rather poorly except in cases when there is no endogeneity ($\rho_{t\varepsilon} = 0$) and the instrument is valid ($\rho_{h\varepsilon} = 0$). The CSB estimator only outperforms the OLS estimator when the instrument is valid ($\rho_{h\varepsilon} = 0$), although it is still biased. The 2SLS estimator only outperforms the OLS estimator when the instrument is strong ($\rho_{th} = 0.4$) and valid ($\rho_{h\varepsilon} = 0$). The 2SLS estimator only outperforms the CSB when the instrument is strong ($\rho_{th} = 0.4$) and valid ($\rho_{h\varepsilon} = 0$), and in some cases when endogeneity is present and strong ($\rho_{t\varepsilon} > 0$).

3.6 DISCUSSION

Again, the choice of program evaluation technique depends on the data available. With a rich dataset (the covariates jointly affecting treatment decisions and outcome are observed and measured) and a reasonable belief that outcome is a known function of the covariates (including treatment), the ordinary least squares estimator is unbiased, efficient, and has a low estimation cost (not complex). The realization that rich datasets do not always exist and unobserved confounders lurk has fueled the growth in program evaluation literature. Instrumental variables, two-stage least squares, control functions, propensity score approaches, and the newly introduced control score blocking approach all offer viable estimation alternatives (to ordinary least squares), although their advantages all come at a cost.

This section examined the costs associated with two-stage least squares, propensity score blocking, and control score blocking in relation to ordinary least square estimation. The propensity score blocking estimator is unbiased and appears to be the superior modeling choice when confounders are observable and the true functional form is unknown. The two-stage least squares and control score blocking estimators are unbiased when instruments are strongly correlated with treatment and uncorrelated with the unobserved confounders. The benefits of using two-stage least squares and control score blocking over ordinary least squares are sensitive to the strength of the correlation between the instrument and treatment (instrument quality), strength of the correlation between the instrument and the unobserved confounders (instrument validity), and the strength of the correlation between the treatment and the unobserved confounders (endogeneity). In some cases, the two-stage least squares and control score blocking estimators are more biased than the ordinary least squares estimator. Finally, the majority of program evaluation techniques are sensitive to sample size, as instrumental variables, two-stage least squares, control functions, propensity score blocking, and control score blocking are only unbiased as sample size becomes large (consistent), even under the best (ideal) situations.

Two-stage least squares is especially sensitive to sample size as can be seen by some of the surprisingly large biases and RMSE's when the sample size is 100 and the instrument quality is weak (ρ_{th} = 0.20). The smaller the population instrument quality (ρ_{th}), the greater the probability that a random sample will have an estimated instrument quality close to zero. The two-stage least squares estimator is very sensitive to the observed correlation between the instrument and treatment, such that as instrument quality approaches zero, the bias of the two-stage least square estimate will approach infinity. This can be seen using Bound et al. (1995) results shown in equations (3.1) and (3.2). As sample size increases, the probability of a bad sample (unrepresentative of the population) decreases. An advantage of the control score blocking approach, which uses the same instrument information, is its robustness to bad samples as compared to two-stage least squares, as the control score blocking estimator performs better than the two-stage least squares estimator when instruments are weak.

SECTION REFERENCES

Angrist, J.D. and A.B. Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives*, 15(4): 69-85.

Bound, J., Jaeger, D.A., and R.M. Baker. 1995. "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association*, 90: 443-450.

Garen, J.E. 1987. "Relationship Among Estimators of Triangular Econometric Models." *Economic Letters*, 25: 39-41.

SECTION TABLES

Table 3.1. Simulation output for sample size 1,000 when selection is on observables (*x* is observed; information set Ω_1). Bias is mean bias and RMSE is root mean squared error of the treatment effect over 1,000 samples. Expectation denotes the expected unbiased estimators, where 1=Ordinary Least Squares, 2=Two-Stage Least Squares, 3=Propensity Score Blocking, 4=Control Score Blocking. Also, O denotes OLS bias is expected less than 2SLS bias, whereas T denotes 2SLS bias is expected less than OLS bias (per Bound et al. 1995).

Model Parameterization						***Ordinary Least Squares***		***Two-Stage Least Squares***		***Propensity Score Blocking***	
Simulation	$\rho_{t\varepsilon}$	$\rho_{h\varepsilon}$	ρ_{th}	**Info Set**	**Expectation**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
1	0.00	0.00	0.40	Ω_1	1,2,3,O	-0.0010	0.0318	-0.0022	0.0789	-0.0016	0.0336
2	0.00	0.00	0.10	Ω_1	1,2,3,O	0.0001	0.0311	0.0024	0.1662	0.0002	0.0328
3	0.00	0.10	0.40	Ω_1	1,2,3,O	-0.0006	0.0327	0.0005	0.0779	-0.0006	0.0336
4	0.00	0.10	0.10	Ω_1	1,2,3,O	0.0002	0.0323	-0.0044	0.1686	-0.0001	0.0338
5	0.10	0.00	0.40	Ω_1	1,2,3,O	-0.0009	0.0311	-0.0041	0.0833	-0.0003	0.0331
6	0.10	0.00	0.10	Ω_1	1,2,3,O	-0.0010	0.0309	-0.0064	0.1594	-0.0008	0.0324
7	0.10	0.10	0.40	Ω_1	1,2,3,O	0.0008	0.0326	0.0001	0.0805	0.0006	0.0347
8	0.10	0.10	0.10	Ω_1	1,2,3,O	0.0003	0.0316	-0.0028	0.1825	0.0005	0.0333
9	0.40	0.00	0.40	Ω_1	1,2,3,O	0.0011	0.0387	0.0013	0.0778	0.0014	0.0409
10	0.40	0.00	0.10	Ω_1	1,2,3,O	0.0004	0.0378	-0.0070	0.1615	0.0003	0.0405
11	0.40	0.10	0.40	Ω_1	1,2,3,O	-0.0010	0.0386	0.0019	0.0970	-0.0003	0.0401
12	0.40	0.10	0.10	Ω_1	1,2,3,O	-0.0016	0.0405	0.0233	0.2922	-0.0011	0.0419

Table 3.2. Simulation output for sample size 1,000 when the confounder is misspecified (x^* is observed; information set Ω_2). Bias is mean bias and RMSE is root mean squared error of the treatment effect over 1,000 samples. Expectation denotes the expected unbiased estimators, where 1=Ordinary Least Squares, 2=Two-Stage Least Squares, 3=Propensity Score Blocking, 4=Control Score Blocking. Also, O denotes OLS bias is expected less than 2SLS bias, whereas T denotes 2SLS bias is expected less than OLS bias (per Bound et al. 1995).

Model Parameterization						***Ordinary Least Squares***		***Two-Stage Least Squares***		***Propensity Score Blocking***	
Simulation	$\rho_{t\varepsilon}$	$\rho_{h\varepsilon}$	ρ_{th}	**Info Set**	**Expectation**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
13	0.00	0.00	0.40	Ω_2	1,2,3,O	0.0001	0.0368	-0.0009	0.0937	0.0005	0.0322
14	0.00	0.00	0.10	Ω_2	1,2,3,O	0.0009	0.0378	0.0050	0.1949	0.0008	0.0330
15	0.00	0.10	0.40	Ω_2	1,3,O	0.0001	0.0394	0.1395	0.1691	-0.0004	0.0337
16	0.00	0.10	0.10	Ω_2	1,3,O	-0.0015	0.0370	0.2875	0.3538	-0.0008	0.0322
17	0.10	0.00	0.40	Ω_2	2,3,T	0.0576	0.0694	-0.0011	0.0953	0.0001	0.0334
18	0.10	0.00	0.10	Ω_2	2,3,T	0.0565	0.0678	-0.0054	0.1954	0.0006	0.0322
19	0.10	0.10	0.40	Ω_2	3,O	0.0548	0.0671	0.1406	0.1726	-0.0015	0.0338
20	0.10	0.10	0.10	Ω_2	3,O	0.0553	0.0672	0.3017	0.3702	0.0003	0.0336
21	0.40	0.00	0.40	Ω_2	2,3,T	0.2783	0.2841	0.0007	0.0933	0.0004	0.0410
22	0.40	0.00	0.10	Ω_2	2,3,T	0.2806	0.2860	0.0061	0.1965	0.0021	0.0411
23	0.40	0.10	0.40	Ω_2	3,T	0.2770	0.2826	0.1600	0.1905	0.0006	0.0403
24	0.40	0.10	0.10	Ω_2	3,O	0.2774	0.2832	0.3710	0.4531	-0.0015	0.0401

Table 3.3. Simulation output for sample size 1,000 when selection is on unobservables (x is unobserved; information set Ω_3). Bias is mean bias and RMSE is root mean squared error of the treatment effect over 1,000 samples. Expectation denotes the expected unbiased estimators, where 1=Ordinary Least Squares, 2=Two-Stage Least Squares, 3=Propensity Score Blocking, 4=Control Score Blocking. Also, O denotes OLS bias is expected less than 2SLS bias, whereas T denotes 2SLS bias is expected less than OLS bias (per Bound et al. 1995).

Model Parameterization						***Ordinary Least Squares***		***Two-Stage Least Squares***		***Control Score Blocking***	
Simulation	$\rho_{t\varepsilon}$	$\rho_{h\varepsilon}$	ρ_{th}	**Info Set**	**Expectation**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
25	0.00	0.00	0.40	Ω_3	1,2,4,O	0.0000	0.0436	-0.0027	0.1123	-0.0025	0.1137
26	0.00	0.00	0.10	Ω_3	1,2,4,O	0.0007	0.0444	0.0103	0.2384	0.0100	0.2279
27	0.00	0.10	0.40	Ω_3	1,O	-0.0007	0.0453	0.3556	0.3733	0.3307	0.3496
28	0.00	0.10	0.10	Ω_3	1,O	-0.0029	0.0445	0.7230	0.7712	0.6102	0.6554
29	0.10	0.00	0.40	Ω_3	2,4,T	0.1387	0.1452	-0.0085	0.1125	0.0017	0.1130
30	0.10	0.00	0.10	Ω_3	2,4,T	0.1392	0.1459	-0.0135	0.2341	0.0083	0.2241
31	0.10	0.10	0.40	Ω_3	O	0.1404	0.1480	0.3571	0.3744	0.3434	0.3610
32	0.10	0.10	0.10	Ω_3	O	0.1415	0.1481	0.7238	0.7693	0.6308	0.6756
33	0.40	0.00	0.40	Ω_3	2,4,T	0.5638	0.5653	-0.0009	0.1084	0.0397	0.1155
34	0.40	0.00	0.10	Ω_3	2,4,T	0.5674	0.5689	-0.0240	0.2328	0.0673	0.2228
35	0.40	0.10	0.40	Ω_3	T	0.5628	0.5643	0.3536	0.3684	0.3684	0.3831
36	0.40	0.10	0.10	Ω_3	O	0.5663	0.5678	0.7237	0.7563	0.7021	0.7322

Table 3.4. Simulation output for sample size 100 when selection is on observables (*x* is observed; information set Ω_1). Bias is mean bias and RMSE is root mean squared error of the treatment effect over 1,000 samples. Expectation denotes the expected unbiased estimators, where 1=Ordinary Least Squares, 2=Two-Stage Least Squares, 3=Propensity Score Blocking, 4=Control Score Blocking. Also, O denotes OLS bias is expected less than 2SLS bias, whereas T denotes 2SLS bias is expected less than OLS bias (per Bound et al. 1995).

Model Parameterization						***Ordinary Least Squares***		***Two-Stage Least Squares***		***Propensity Score Blocking***	
Simulation	$\rho_{t\varepsilon}$	$\rho_{h\varepsilon}$	ρ_{th}	**Info Set**	**Expectation**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
37	0.00	0.00	0.40	Ω_1	1,2,3,O	-0.0009	0.1017	-0.0020	0.2826	-0.0011	0.1049
38	0.00	0.00	0.10	Ω_1	1,2,3,O	0.0031	0.1040	0.1577	6.3727	0.0029	0.1081
39	0.00	0.10	0.40	Ω_1	1,2,3,O	0.0007	0.1031	0.0035	0.2862	0.0004	0.1065
40	0.00	0.10	0.10	Ω_1	1,2,3,O	-0.0026	0.1018	0.0502	2.5684	-0.0020	0.1055
41	0.10	0.00	0.40	Ω_1	1,2,3,O	-0.0011	0.1009	-0.0010	0.2761	-0.0016	0.1030
42	0.10	0.00	0.10	Ω_1	1,2,3,O	0.0012	0.1075	0.3395	11.2571	0.0015	0.1105
43	0.10	0.10	0.40	Ω_1	1,2,3,O	0.0016	0.1010	-0.0082	0.3040	0.0023	0.1036
44	0.10	0.10	0.10	Ω_1	1,2,3,O	-0.0043	0.1024	0.0387	2.6714	-0.0054	0.1047
45	0.40	0.00	0.40	Ω_1	1,2,3,O	-0.0067	0.1259	-0.0193	0.2682	-0.0060	0.1288
46	0.40	0.00	0.10	Ω_1	1,2,3,O	0.0051	0.1232	-0.0045	1.2453	0.0080	0.1270
47	0.40	0.10	0.40	Ω_1	1,2,3,O	0.0016	0.1207	-0.0067	0.3521	0.0010	0.1238
48	0.40	0.10	0.10	Ω_1	1,2,3,O	0.0024	0.1218	-4.2728	138.0009	0.0027	0.1248

Table 3.5. Simulation output for sample size 100 when the confounder is misspecified (x^* is observed; information set Ω_2). Bias is mean bias and RMSE is root mean squared error of the treatment effect over 1,000 samples. Expectation denotes the expected unbiased estimators, where 1=Ordinary Least Squares, 2=Two-Stage Least Squares, 3=Propensity Score Blocking, 4=Control Score Blocking. Also, O denotes OLS bias is expected less than 2SLS bias, whereas T denotes 2SLS bias is expected less than OLS bias (per Bound et al. 1995).

Model Parameterization						***Ordinary Least Squares***		***Two-Stage Least Squares***		***Propensity Score Blocking***	
Simulation	$\rho_{t\varepsilon}$	$\rho_{h\varepsilon}$	ρ_{th}	**Info Set**	**Expectation**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
49	0.00	0.00	0.40	Ω_2	1,2,3,O	-0.0009	0.1136	0.0077	0.3374	-0.0009	0.1028
50	0.00	0.00	0.10	Ω_2	1,2,3,O	0.0044	0.1141	0.2958	9.1816	0.0060	0.1059
51	0.00	0.10	0.40	Ω_2	1,3,O	0.0049	0.1190	0.1462	0.3624	0.0085	0.1062
52	0.00	0.10	0.10	Ω_2	1,3,O	0.0009	0.1197	0.2615	13.6631	0.0006	0.1081
53	0.10	0.00	0.40	Ω_2	2,3,T	0.0552	0.1337	0.0026	0.3107	0.0122	0.1092
54	0.10	0.00	0.10	Ω_2	2,3,T	0.0493	0.1297	0.0688	5.2226	0.0049	0.1134
55	0.10	0.10	0.40	Ω_2	3,O	0.0457	0.1261	0.1332	0.3659	0.0031	0.1068
56	0.10	0.10	0.10	Ω_2	3,O	0.0463	0.1292	0.6207	4.6361	0.0054	0.1095
57	0.40	0.00	0.40	Ω_2	2,3,T	0.2397	0.2813	-0.0279	0.3254	0.0340	0.1329
58	0.40	0.00	0.10	Ω_2	2,3,T	0.2444	0.2822	0.1576	8.2148	0.0404	0.1356
59	0.40	0.10	0.40	Ω_2	3,T	0.2410	0.2815	0.1382	0.4128	0.0401	0.1388
60	0.40	0.10	0.10	Ω_2	3,O	0.2400	0.2794	1.0242	18.1264	0.0343	0.1344

Table 3.6. Simulation output for sample size 100 when selection is on unobservables (*x* is unobserved; information set Ω_3). Bias is mean bias and RMSE is root mean squared error of the treatment effect over 1,000 samples. Expectation denotes the expected unbiased estimators, where 1=Ordinary Least Squares, 2=Two-Stage Least Squares, 3=Propensity Score Blocking, 4=Control Score Blocking. Also, O denotes OLS bias is expected less than 2SLS bias, whereas T denotes 2SLS bias is expected less than OLS bias (per Bound et al. 1995).

Model Parameterization						***Ordinary Least Squares***		***Two-Stage Least Squares***		***Control Score Blocking***	
Simulation	$\rho_{t\varepsilon}$	$\rho_{h\varepsilon}$	ρ_{th}	**Info Set**	**Expectation**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
61	0.00	0.00	0.40	Ω_3	2,4,O	-0.0023	0.1446	0.0005	0.3910	0.0025	0.2155
62	0.00	0.00	0.10	Ω_3	2,4,O	0.0096	0.1456	-0.0587	7.1334	0.0025	0.2311
63	0.00	0.10	0.40	Ω_3	O	0.0017	0.1374	0.3855	0.7560	0.0747	0.2279
64	0.00	0.10	0.10	Ω_3	O	-0.0053	0.1459	0.7808	4.3768	0.0461	0.2422
65	0.10	0.00	0.40	Ω_3	2,4,T	0.1380	0.1974	-0.0240	0.3862	0.1144	0.2408
66	0.10	0.00	0.10	Ω_3	2,4,T	0.1433	0.2076	-0.3837	9.3678	0.1381	0.2795
67	0.10	0.10	0.40	Ω_3	O	0.1432	0.2036	0.3713	0.5488	0.1967	0.2927
68	0.10	0.10	0.10	Ω_3	O	0.1319	0.1958	0.3897	11.3248	0.1699	0.2916
69	0.40	0.00	0.40	Ω_3	2,4,T	0.5612	0.5762	-0.0684	0.4337	0.4334	0.4745
70	0.40	0.00	0.10	Ω_3	2,4,T	0.5712	0.5853	-1.3238	36.7061	0.5471	0.5863
71	0.40	0.10	0.40	Ω_3	T	0.5715	0.5870	0.3386	0.4991	0.5230	0.5587
72	0.40	0.10	0.10	Ω_3	O	0.5655	0.5813	0.6470	9.5649	0.5863	0.6265

Table 3.7. Ratio of the inconsistency from the two-stage least squares estimation to the inconsistency from the ordinary least squares estimation. The expected inconsistency ratio is defined by Bound et al. 1995, while the inconsistency ratio from the simulations are calculated from the bias reported in tables 3.1-3.6. Simulations with Ω_1 (observable confounders), $\rho_{t\varepsilon} = 0$, and $\rho_{h\varepsilon} = 0$ are excluded as the OLS estimate is unbiased, the inconsistency ratio is undefined, or the inconsistency ratio is zero, respectively.

	Model Parameterization				***Expected***	***Simulated; N=1,000***	***Simulated; N=100***
Simulation	$\rho_{t\varepsilon}$	$\rho_{h\varepsilon}$	ρ_{th}	**Info Set**	**Inconsistency Ratio**	**Inconsistency Ratio**	**Inconsistency Ratio**
19	0.10	0.10	0.40	Ω_2	2.500	2.568	2.913
20	0.10	0.10	0.10	Ω_2	10.000	5.459	13.400
23	0.40	0.10	0.40	Ω_2	0.625	0.578	0.573
24	0.40	0.10	0.10	Ω_2	2.500	1.338	4.267
31	0.10	0.10	0.40	Ω_3	2.500	2.544	2.593
32	0.10	0.10	0.10	Ω_3	10.000	5.116	2.955
35	0.40	0.10	0.40	Ω_3	0.625	0.628	0.593
36	0.40	0.10	0.10	Ω_3	2.500	1.278	1.144

4. EMPIRICAL EFFICACY OF SUPPRESSION AND FUELS MANAGEMENT ON WILDFIRE BEHAVIOR

4.1 INTRODUCTION

Wildfires burn over 4 million acres annually nationwide (1960-2004), and cost Federal agencies over $830 million per year (1994-2004) in suppression costs (National Interagency Fire Center 2005). Additionally, hundreds of millions more are spent on fuels management programs. Fuels management methods include the use of prescribed fire treatments to reduce the amount of ground level and ladder fuels, which connect ground fuels to the tree crowns. Prescribed fires are intentionally set, low intensity fires administered by trained specialists and under ideal weather conditions.

I am interested in modeling the effect of suppression and fuels management (wildfire treatment variables) on wildfire behavior (the outcome of wildfire treatment). Fire crew response time, the time it takes a fire crew to respond to a fire, is used as a proxy measure for suppression effort (initial attack). Data constraints limit other measures of suppression; however, the use of response time is meaningful because it is related to available suppression resources. When fire fighting resources become constrained, for instance due to responding to multiple fires, longer response times are observed. Thus, rapid response times correspond to the availability of greater resources across space, while longer response times correspond to stretched resources. Fuels management is used to minimize the risk of a damaging wildfire through elimination of understory and ladder fuels, with prescribed fire being dominant in Florida. In this study, it is the only fuels management variable used (due to data limitations). Prescribed burning is applied in advance of the fire season and has been shown to have lasting effects up to about 3 years on very fine scale experimental plots (Outcalt and Wade 2004). I examine that effect fire crew response time and prescribed fire have on wildfire intensity-weighted size, because solely focusing on either fire size or intensity alone may be misleading. Wildfires may be big, but not intense, thus not very damaging. For instance, prescribed fire, by eliminating ladder fuels, may allow for larger wildfires (they may now move quicker through less dense forest), but ones of lower intensity. If wildfire management minimizes the chance of a crown fire at the expense of surface area burned, this might be seen as a success.

In modeling wildfire there are two sources of potential endogeneity—simultaneity bias and selection bias. Simultaneity occurs when two or more variables are co-determined with each other. Statistical bias results, in a least-squares framework, if an endogenous variable is regressed on another. It is easy to imagine why wildfire behavior and suppression effort may be simultaneously determined—initial wildfire behavior influences the fire crew effort and response, but in turn, fire crew effort affects wildfire behavior.

Selection bias occurs in least-squares regression when observations are selected (either through self-selection or by selection by the program administrator) into a group and those factors causing group participation also directly affect the outcome, but are ignored. Prescribed burning typically occurs in winter and early spring in advance of the wildfire season, so problems of endogeneity due to simultaneity are largely absent. However, because prescribed burning occurs in areas chosen by wildland managers, the selection of prescribed burn sites are not determined by a random process. In fact, it might be expected that prescribed burning is applied to areas with higher wildfire risk, implying that areas with prescribed burning may be different from those without.

4.2 DATA AND STUDY SITE

I examine the effectiveness of wildfire management on wildfires occurring in the St. Johns River Water Management District (SJRWMD) in northeast Florida, from 1996-2001. The SJRWMD area comprises portions of the 18 northeast counties in Florida. The SJRWMD is an ideal study area given its abundance of wildfire (in size and occurrence), use of prescribed fire, values at risk (fires are potentially very damaging and are actively managed), and quality of available data. Florida, unlike many areas of the western United States, is heavily populated (it is fourth in state population and ninth in population density; United States Census Bureau 2004) and averages 218,638 wildfire acres a year and 763,205 acres of silvicultural-based prescribed fire a year (476,590 is for hazard reduction). Between wildfire and silvicultural prescribed fire, approximately 3% of Florida burns annually (not including agricultural burns). The SJRWMD averages over 48,596 acres a year in wildfire and 133,833 acres of prescribed burning (73,099 for hazard reduction) resulting in about 2.3% of the SJRWMD burning annually.

Wildfire records

Data on individual wildfire occurrences were obtained from the Florida Division of Forestry (FDOF). FDOF's wildfire data contains detailed information on fires found on private and state-owned lands including, but not limited to, the date and time of ignition, location (township, range, and cadastral section), size (acres), associated weather conditions (wind speed and direction and humidity), rate of spread, flame length, and cause (arson, campfires, cigarettes, children, debris burning, equipment, lightning, miscellaneous, railroad, and unknown) from 1981-2001. Both the wildfire data and the prescribed burning data (below) are geo-located to a Public Land Survey section (township, range, section), which is approximately a one-square mile rectangle.

From 1981-2002 there were 31,603 wildfires in the SJRWMD, with almost half of them (48%) coming from non-incendiary human-caused sources (accidental ignitions), followed by arson (29%), and lightning (23%). The SJRWMD (and Florida in general) fire season appears to begin late-winter to early-spring and last until the middle of the summer, with burned area peaking around May and June.

Wildfire management

The FDOF provided a second dataset that details all prescribed fire activities within the state (in order to conduct a prescribed burn in Florida, a permit must be obtained from the FDOF). Permit data include information on the location (located by the township, range, and cadastral section), reason (hazard reduction, prior to seeding, site preparation, disease control, wildlife, ecological, or other), and total size (in acres). The dataset includes permits issued between 1989 and 2001, although full state-wide reporting did not occur until 1993.

In addition, wildfire start time, wildfire report time (time between ignition and report to fire department), and fire crew arrival time, given in the FDOF database, allow for the creation of a measure of initial attack (fire crew response time—the time between report and arrival).

Climate and weather

The El Niño Southern Oscillation (ENSO) measure used in this analysis is the Niño-3 sea surface temperature (SST) anomaly, which was obtained from the National Oceanic and Atmospheric Administration (National Oceanic and Atmospheric Administration 2002). The Niño-3 SST anomaly is measured as the positive (El Niño) or negative (La Niña) deviation, in degrees centigrade, of the Pacific sea surface temperature (at a specific location). The Keetch-Byram Drought Index (KBDI) was calculated for two weather stations in the SJRWMD region using daily data collected by the National Climate Data Center and provided by EarthInfo (2002).

Landscape characteristics

Section-level road and census data (population, income, and education) were created from US Census Bureau TIGER/Line GIS data. The National Land Cover Data, based on the Multi-Resolution Land Characteristics Consortium's land cover map (30-meter resolution grid) was used to determine landcover composition within and surrounding each section.

Five landcover classes were assembled—grass (grassland/herbaceous), upland forest (deciduous, evergreen, and mixed forest), urban (low intensity residential, high intensity residential, and commercial/industrial/transportation), water (open water), and wetland (woody wetland).

4.3 EMPIRICAL MODELS

I examine the effectiveness of wildfire management on wildfires over the period 1996-2001. In this study, wildfire management is composed of two components—fire suppression and fuels management. The fire suppression (fire fighting) variable is defined as the amount of time (in hours) it takes fire crews to respond to a fire call (the time from when the fire is reported until the fire crew arrives at the fire site). The fuels management variable is defined as the number of acres of prescribed fire (for hazard reduction), over the last three years, permitted to the section, and those contiguous surrounding sections, in which the fire occurred. Full coverage of the prescribed fire data begins in 1993, therefore the analysis begins with wildfire occurring in 1996 (to include for the three year lag). Since suppression and fuels management may affect both wildfire size (extent) and intensity, I use an intensity-weighted acres measure of wildfire behavior (as described in Mercer, et al. forthcoming), measured in kW-acre/meters. The wildfire intensity-weighted acre measure is defined as:

$$w_i = a_i * 259.833 * l_i^{2.174} \qquad (4.1)$$

where w_i is wildfire intensity-weighted acres (kW-acre/meter), a_i is acres burned, l_i is fire flame length (meters). Wildfire intensity is approximated as a function of flame length (= $259.833*l_i^{2.174}$; see Kennard [2004]), which is reported by FDOF.

I specify four models—ordinary least squares, two-stage least squares, propensity score blocking and control score blocking models—using wildfire as the unit of observation. Table 4.1 provides descriptive statistics for all variables in the models. Estimating the effectiveness of wildfire management using wildfire as the unit of observation is certainly appropriate in the case of suppression because suppression affects wildfire behavior, as oppose to the ignition process. Suppression does not prevent wildfire ignition, so sections without wildfire would also be without suppression (at least as suppression is defined here). If fuels management has a significant impact on ignition success then the current modeling framework introduces its own selection bias. The benefits of fuels management are often couched in terms of reduction in fire intensity and size rather than in terms of ignition (Outcalt and Wade 2004; Brose and Wade 2002; Wade et al. 2000). However, there is evidence that arson ignitions are correlated with fuels management activities (Prestemon and Butry 2005). This phenomenon may have less to do with the physical relation between fuel levels and ignition and more to with the realization by the arsonist that the wildfire has a lower probability of achieving a devastating size or intensity in actively managed areas. In addition, actively managed forests may pose a greater threat of apprehension to the arsonist. Therefore, the focus is the effectiveness of wildfire management in wildfire prone areas.

4.3.1 Ordinary Least Squares

The OLS models are specified as:

$$\ln(w) = \beta_0^{ols} + \mathbf{X}_{\mathrm{M}}\boldsymbol{\beta}_M^{ols} + \mathbf{X}_{\mathrm{C}}\boldsymbol{\beta}_C^{ols} + \mathbf{X}_{\mathrm{F}}\boldsymbol{\beta}_F^{ols} + \mathbf{X}_{\mathrm{I}}\boldsymbol{\beta}_I^{ols} + \mathbf{X}_{\mathrm{O}}\boldsymbol{\beta}_O^{ols} + \varepsilon^{ols} \qquad (4.2)$$

where w is wildfire intensity-weighted acres, ε^{ols} is the error term, $\boldsymbol{X}_{\mathrm{M}}$ (management variables) includes the natural log of the suppression and prescribed fire variables, described above, along with an indicator variable denoting let burn fires (fires that are allowed to burn themselves out), $\boldsymbol{X}_{\mathrm{C}}$ (climate and weather variables) includes KBDI, whether the fire occurred in a El Niño or La Niña phase, and its magnitude, humidity, spread potential (a function of wind and fuel moisture), actual fire spread in miles per hour (recorded as an ordinal variable), the natural log of wind speed, and wind direction, $\boldsymbol{X}_{\mathrm{F}}$ (fuel variables) includes fuel type, build-up index (a measure of available fuel), latitude, longitude, elevation, slope, landscape composition (percent of landscape in upland forest, agriculture, rangeland, residential, wetland, and water), forest density, previous 12-year fire history (number of ignitions), previous 12-year fire history in neighboring sections, the natural log of previous 12-year fire intensity-weighted acres (a measure of previous fire behavior that may be related to available fuels), the natural log of previous 12-year fire intensity-weighted acres in neighboring sections, natural log of road density, $\boldsymbol{X}_{\mathrm{I}}$ (ignition variables) includes ignition cause, natural log of population, natural log of family income (income is found to be correlated with wildfire in Florida; Butry et al. [2002]), and natural log of fire report time (time from fire ignition to call to fire department), $\boldsymbol{X}_{\mathrm{O}}$ (other variables) includes indicator variables denoting county, fire year (begins September of the preceding calendar year and ends in October of the current calendar year), month, fire district, and ownership type, and the β's are parameters.

While a rich dataset for modeling wildfire in Florida is available, strengthening the justification for use of ordinary least squares estimation, the potential for model misspecification is high. For instance, a number of explanatory variables, such as fire spread and previous fire behavior (the natural log of previous 12-year fire intensity-weighted acres) are likely endogenous with fire suppression and previous prescribed fire, respectively. In addition, elements of fire behavior, such as fire spread and intensity, appear to be highly nonlinear functions of weather and landscape characteristics (Rothermel 1972).

4.3.2 Two-Stage Least Squares

The fitted values from the first stage 2SLS models, for both prescribed fire and response time, are used as the instruments in the second stage wildfire model. Due to the prescribed fire variable having an excess of zeros—some fires had prior prescribed fire treatments, varying in acres burned, others had none—the prescribed fire instrument is estimated using a Tobit regression. Of the 7480 observations, only 1830 experienced previous prescribed burning. Prescribed fire acres is censored at zero, however because I am modeling the natural log of prescribed fire, and the natural log of zero is undefined, the lower censor point for the log transformed model is set to ln(0.001) or -6.91.

The model is defined as:

$$\begin{aligned} \ln(p_i) &= \mathbf{Z}_{pi}\boldsymbol{\beta}_p^{p2sls} + \varepsilon_i^{p2sls} && \text{if } p_i > 0 \\ \ln(p_i) &= -6.91 && \text{otherwise} \end{aligned} \quad . \tag{4.3}$$

The predicted value of prescribed fire (which becomes the instrument) is:

$$\ln(p) = \Phi L + (1 - \Phi)\mathbf{Z}_p\boldsymbol{\beta}_p^{p2sls} + \sigma_\varepsilon^{p2sls}\phi \tag{4.4}$$

where

$$\phi = \phi\left(\frac{L - \mathbf{Z}_p\boldsymbol{\beta}_p^{p2sls}}{\sigma_\varepsilon^{p2sls}}\right), \qquad \Phi = \Phi\left(\frac{L - \mathbf{Z}_p\boldsymbol{\beta}_p^{p2sls}}{\sigma_\varepsilon^{p2sls}}\right) \tag{4.5}$$

where p is the amount of prescribed fire preceding each wildfire, L is the lower censor limit, ϕ and Φ are the standard normal pdf and cdf, respectively, $\sigma_\varepsilon^{p2sls}$ is the standard deviation of ε_i^{p2sls}, $\boldsymbol{Z}_{\text{P}}$ are the instruments for prescribed fire, consisting of the percent of days over the past three years that relative humidity was between 30% and 50%, percent of days over the past three years that wind speed was between 6mph to 20 mph, average maximum temperature over the past three years, and average precipitation over the past three years, and the β's are parameters. These weather variables are used as instruments because they account for daily conditions when prescribed fire application is recommended that are correlated with prescribed fire, but should not be correlated with future wildfire behavior (Florida Department of Forestry, Prescribed Fire Manual 2005). The model is estimated using maximum-likelihood (see Maddala 1983).

Due to the simultaneity between suppression and wildfire, the fire crew response time instrument is specified as:

$$\ln(r) = \beta_0^{r2sls} + \mathbf{X}_M \boldsymbol{\beta}_M^{r2sls} + \mathbf{X}_{\mathrm{C}} \boldsymbol{\beta}_C^{r2sls} + \mathbf{X}_{\mathrm{F}} \boldsymbol{\beta}_F^{r2sls} + \mathbf{X}_{\mathrm{I}} \boldsymbol{\beta}_I^{r2sls} + \mathbf{X}_{\mathrm{O}} \boldsymbol{\beta}_O^{r2sls} + z_R \beta_R^{r2sls} + \varepsilon^{r2sls} \quad (4.6)$$

where r is firecrew response time and ε^{r2sls} is the error term, the other **X**'s as above, except $\boldsymbol{X}_{\mathrm{M}}$ excludes response time, and z_R is an instrument for response time, the distance to nearest fire department, and the β's are parameters. Fire crew response time is modeled using OLS.

The second stage wildfire model is estimated as:

$$\ln(w) = \beta_0^{sls} + \hat{\mathbf{X}}_{\mathrm{M}} \boldsymbol{\beta}_M^{sls} + \mathbf{X}_{\mathrm{C}} \boldsymbol{\beta}_C^{sls} + \mathbf{X}_{\mathrm{F}} \boldsymbol{\beta}_F^{sls} + \mathbf{X}_{\mathrm{I}} \boldsymbol{\beta}_I^{sls} + \mathbf{X}_{\mathrm{O}} \boldsymbol{\beta}_O^{sls} + \varepsilon^{sls} \quad (4.7)$$

where $\hat{\mathbf{X}}_{\mathrm{M}}$ consists of the instrumented response time and prescribed fire variables (the let burn variables are unchanged), ε^{2sls} is the error term, and the β's are parameters. The model is estimated using OLS.

4.3.3 Propensity Score Blocking

The fitted values from propensity score estimator models, for both prescribed fire and response time, are the propensity scores. The response time and prescribed fire propensity scores are estimated separately, but are both used in a single model of wildfire behavior. The propensity score estimator models include all the variables in the wildfire equation that are either known at the time of treatment or unaffected by treatment (e.g.: climate variables). This is done to ensure that all the covariates that directly affect wildfire behavior and the treatments (response time and prescribed fire) are included in the score estimator models.

The propensity score estimator for fire crew response time is estimated as:

$$\ln(r) = \beta_0^{rpsb} + \mathbf{X}_M \boldsymbol{\beta}_M^{rpsb} + \mathbf{X}_\mathrm{C} \boldsymbol{\beta}_C^{rpsb} + \mathbf{X}_\mathrm{F} \boldsymbol{\beta}_F^{rpsb} + \mathbf{X}_\mathrm{I} \boldsymbol{\beta}_I^{rpsb} + \mathbf{X}_\mathrm{O} \boldsymbol{\beta}_O^{rpsb} + \varepsilon^{rpsb} \qquad (4.8)$$

where r is firecrew response time and ε^{rpsb} is the error term, the **X**'s as above, except $\boldsymbol{X}_\mathrm{M}$ excludes response time, $\boldsymbol{X}_\mathrm{C}$ which excludes fire spread (it may be affected by suppression), and the β's are parameters. Fire crew response time is modeled using OLS.

The propensity score estimator for prescribed fire is estimated as:

$$\ln(p) = \beta_0^{ppsb} + \mathbf{X}_M \boldsymbol{\beta}_M^{ppsb} + \mathbf{X}_\mathrm{C} \boldsymbol{\beta}_C^{ppsb} + \mathbf{X}_\mathrm{F} \boldsymbol{\beta}_F^{ppsb} + \mathbf{X}_\mathrm{I} \boldsymbol{\beta}_I^{ppsb} + \mathbf{X}_\mathrm{O} \boldsymbol{\beta}_O^{ppsb} + \varepsilon^{ppsb} \qquad (4.9)$$

where all variables have been defined before, except $\boldsymbol{X}_{\mathrm{M}}$ excludes response time and prescribed fire, $\boldsymbol{X}_{\mathrm{C}}$ includes only El Niño and La Niña, $\boldsymbol{X}_{\mathrm{F}}$ excludes vegetation build-up, $\boldsymbol{X}_{\mathrm{I}}$ only includes population and income, ε^{ppsb} is the error term, and the β's are parameters. While the Tobit model could have been applied here, the goal of the propensity score estimator is not to consistently estimate the parameters. Rather it is used to create a balancing score.

To test the covariate balance in the response time propensity score estimator model (prescribed fire propensity score estimator model), I square each continuous variable and regress it on the natural log of response time (natural log of prescribed fire) and the estimated response time propensity score (estimated prescribed fire propensity score). If the coefficient on the natural log of response time (natural log of prescribed fire) is statistically indistinguishable from zero, then the covariate is deemed balanced, if not, then the squared term is added to the propensity score estimator model. The squared-term of the natural log of previous wildfire intensity-weighted acres, natural log of previous wildfire intensity-weighted acres in neighboring areas, and natural log of fire report time are included in the response time propensity score estimator model due to lack of initial balance. The squared-term of the natural log of neighboring non-hazard reducing prescribed fire is included in the prescribed fire propensity score model due to lack of initial balance.

Figure 4.1 shows the balancing ability of the propensity scores. If the propensity score balances the covariates and if the covariates include all the variables that confound the treatment-outcome relationship, then the propensity score blocking estimator will be unbiased. The balancing test is described in Sections 2.2.2 and 2.3.2. If the propensity score is a balancing score at a 5% significance level, then the t-statistic associated with treatment (response time and prescribed fire) when regressing a nonlinear transformation of the covariate on treatment and the propensity score should be less than 1.96. Figure 4.1 presents the t-statistics associated with regressing a nonlinear transformation of the covariates on (a) the natural log of response time, (b) the natural log of response time and the propensity score of the natural log of response time, (c) the natural log of prescribed fire, (d) the natural log of prescribed fire and the propensity score of the natural log of prescribed fire. The relevant comparisons to make are between 4.1a and 4.1b and between 4.1c and 4.1d. Figure 4.1a shows that some of the covariates are related to response time, whereas figure 4.1b shows that the covariates are not statistically related (at the 5% significance level) to response time when conditioning on the response time propensity score. This implies that conditional on the propensity score, response time is not related to the covariates—this is the condition that the propensity score must achieve to be a balancing score—so that conditioning on the response time propensity score ensures that differences in wildfire behavior are due to differences in response time and not to any underlying factors. Likewise, figures 4.1c and 4.1d show that the prescribed fire propensity score is also a balancing score.

The PSM wildfire models are estimated using the above described OLS functional form, except that separate regressions are performed on groups of the data (defined by the propensity scores). Thus, the regressions are conditioned on the propensity scores—

regressions are performed on ‘like’ observations. Because there are two arrays of propensity scores (one for response time, another for prescribed fire), placement into groups is directed by the relative rank of the two scores (low-low, low-mid, low-high, mid-low, mid-mid, mid-high, high-low, high-mid, high-high). Using the group regression framework allows modeling of the conditional wildfire intensity-acres distribution, with conditioning performed on the propensity scores. Again, by conditioning on the propensity scores, the estimated relationship between response time and prescribed fire on wildfire intensity-acres represents the actual causal effect of management on fire behavior.

4.3.4 Control Score Regression

Response time control scores are generated using the residuals from the 1st stage 2SLS models for response time (4.6) and the prescribed fire control scores are generated using the residuals from OLS estimation of (4.3):

$$\ln(p_i) = \mathbf{z}_{pi}\boldsymbol{\beta}_p^{pcsb} + e_i^{pcsb} \tag{4.10}$$

where $\ln(p_i)$ is set to -6.91 if $p_i = 0$. Because there are two arrays of control scores (one for response time, another for prescribed fire), placement into groups is directed by the relative rank of the two scores (low-low, low-mid, low-high, mid-low, mid-mid, mid-high, high-low, high-mid, high-high). The control score blocking method was estimated, but replaced with a single equation model (instead of nine used in PSB) similar to (4.2) with group effects (control score groups) and group interaction treatment effects (the response time and prescribed fire variables are interacted with the control score groups) due to estimation difficulties (not all block regression would estimate the treatment effects due to lack variation of the prescribed fire variable within the block group—no prescribed fire occurred). Instead of estimating nine separate regressions based on group rank, group rank dummy variables are included into a single equation model, along with group rank treatment variable interactions. The single equation model is specified as:

$$\begin{aligned}\ln(w) = {} & \beta_0^{cf} + \mathbf{X}_{\mathrm{M}}\boldsymbol{\beta}_M^{cf} + \mathbf{X}_{\mathrm{C}}\boldsymbol{\beta}_C^{cf} + \mathbf{X}_{\mathrm{F}}\boldsymbol{\beta}_F^{cf} + \mathbf{X}_{\mathrm{I}}\boldsymbol{\beta}_I^{cf} + \mathbf{X}_{\mathrm{O}}\boldsymbol{\beta}_O^{cf} + \mathbf{D}\boldsymbol{\beta}_D^{cf} + \mathbf{D}r\boldsymbol{\beta}_{DR}^{cf} \\ & + \mathbf{D}p\boldsymbol{\beta}_{DP}^{cf} + \varepsilon^{cf}\end{aligned} \tag{4.11}$$

where the **X**'s as before except now **D** denotes the group dummy variables and **D***r* and **D***p* denote the group dummy variables interacted with response time and prescribed fire, respectively. For block groups with no previous prescribed fire, the interaction term is not estimable and is automatically dropped from the model.

4.4 RESULTS

Results of the OLS, 2SLS, PSB, and CSB wildfire models are presented in tables 4.2 – 4.12. Table 4.2 presents the OLS results. Tables 4.3 – 4.5 present the 2SLS models (tables 4.3 and 4.4 showing the 1st stage models). Tables 4.6 – 4.8 presents the PSB models (tables 4.6 and 4.7 showing the 1st stage propensity score estimator models). Table 4.8 summarizes the influence response time and prescribed fire have on wildfire intensity-acres, rather than presenting all nine subgroup regressions (available upon request). Tables 4.9 – 4.11 present the CSB models (table 4.3 showing the 1st stage 2SLS estimates used to construct the control score for response time and table 4.9 showing the model used to construct the control score for prescribed fire when prescribed fire is modeled using OLS). Table 4.10 presents the control score model, including only the main treatment effects. Table 4.11 summarizes the influence response time and prescribed fire have on wildfire intensity-acres, focusing on the group treatment effects (main effects are shown in table 4.10). Table 4.12 summarizes the wildfire management treatment effects by estimator.

All 1st stage instrumental variable, propensity score model, and control score models are significant (based on the F-tests), but explain a limited amount of the variation of the respective management variable (11-18%), excluding the 1st stage prescribed fire instrumental variable model (only a pseudo-R^2 is calculated; psuedo-R^2 is <0.01) and control core prescribed fire model (2%). Because all models are estimated using the natural log of wildfire intensity-acres, response time, and prescribed fire variables, the reported management coefficients are elasticities, and therefore directly comparable across models. The OLS model estimates that a 1% increase in prescribed fire acres yields a 0.0159% (0.0071 standard error) reduction in wildfire intensity-acres, and that a 1% decrease in response time yields a 0.4151% (0.0393 standard error) reduction in wildfire intensity-acres. Both parameters are found to be statistically significant at the 5% level. The 2SLS model, on the other hand, finds no statistical relationship between wildfire management and wildfire intensity-acres.

The results of the propensity score model are a little more complex, as they allow for nonlinearity between wildfire management and wildfire intensity-acres. Table 4.8 presents the propensity score blocking treatment effects. While the control score model also allows for nonlinearity, there is no evidence of treatment effect variation by control score group (tables 4.10 and 4.11 present the control score regression treatment effects). Because the propensity score blocking and control score regression models are estimated differently, their results are presented in a slightly different fashion. For the control score regression model, the group treatment effects (comparable to the propensity score results shown in table 4.8) are the base case treatment effect (all group dummies set to zero; base case treatment effects shown in the “Lower Third”/“Lower Third” box in table 4.11)) *plus* the individual group treatment effects. For example, the response time treatment effect for observations associated with the middle third prescribed fire control score and the lower third response time control score is (0.4705 - 0.0804 = 0.3901).

For the propensity score blocking model, a 1% decrease in response time yields a 0.3175% decrease (0.0420 standard error) in wildfire intensity-acres, based on the average response time elasticity (averaged over all subgroups). For the control score model, a 1% decrease in response time yields a 0.4314% decrease (averaged over all subgroups; 0.0366 standard error) in wildfire intensity-acres. For the propensity score blocking model, the average prescribed fire elasticity (again averaged over all subgroups) implies a 1% increase in prescribed fire acres yields a 0.0138% decrease (0.0085 standard error) in wildfire intensity-acres, which is similar to the OLS model. For the control score model, a 1% decrease in prescribed fire acres yields a 0.0356% decrease (averaged over all subgroups; 0.0003 standard error) in wildfire intensity-acres.

However, upon further examination of the subgroup results, there appear some interesting findings. In this analysis, the propensity scores can be thought of as the expected levels of treatment. The response time propensity scores in the lower third subgroup correspond to those fires with quicker expected fire crew response rates, and those in the upper third correspond with fires with slower expected fire crew response rates. The lower third subgroup prescribed fire propensity scores correspond with fires with less expected prior prescribed burn acres, while those fires in the upper third subgroup correspond with more expected area burned by prescribed fire.

Across the nine subgroups, prescribed fire is only significant in one group—those with median levels of expected prescribed fire and with lower expected response times. The prescribed fire elasticity is -0.0436, which is larger than the OLS estimate, although not statistically different. Quick fire crew response times are generally significant across all subgroups (except for fires with the median and higher levels of expected prescribed fire and with the quicker expected fire crew response times), although the effect of quick response differ by subgroup. Focusing only on the subgroups with significant results, the weighted average elasticity is 0.3754, with subgroup variation ranging from 0.2787 to 0.5097. Moderate treatment effects are found for fires with expected median response times, with higher treatment effects found in the lower and upper third response time groups. These effects are modified as fires move from the lower third level of prescribed fire propensity scores to upper levels, with smaller treatment effects reported in areas with larger expected levels of prior prescribed fire. Fires with moderate expected response times exhibit fairly consistent treatment effects (of response time) across the prescribed fire groups. Overall, findings might suggest that prescribed fire and suppression are used more as substitutes than as complements.

The meaning of the control score is perhaps less intuitive, but it is of little consequence because the individual group effects are insignificant. In fact, many of the control score groups did not have any prescribed fire; it appears that the propensity score and control score groupings are fairly different. The main response time treatment effect elasticity is significantly different than zero at the 5% level, but is not significantly different than the OLS or propensity score estimates (using a t-test with a 5% significance level). The main prescribed fire treatment effect elasticity is not only significantly different than zero at the 5% level, but also significantly different than the OLS and propensity score estimates (using a t-test with a 5% significance level) and appear to be over two times as large.

4.5 DISCUSSION

The ordinary least squares, propensity score blocking, and control score regression models all yield evidence that wildfire management (fire crew response time and prescribed fire) is effective at limiting wildfire size and intensity. The two-stage least squares model finds no statistical relationship between wildfire management and fire behavior. The propensity score blocking treatment effect estimates (the weighted average estimates) are in absolute terms smaller than the OLS estimates, implying that OLS over-estimates the effect of treatment, whereas the control score regression treatment effect estimates (the weighted average estimates) are in absolute terms larger than the OLS, implying that OLS under-estimates the effect of treatment. However, the weighted propensity score blocking prescribed fire treatment effect is not statistically different than the OLS estimate, whereas the propensity score blocking group effects and weighted control score estimates are. The control score regression response time treatment effect is not statistically different from the OLS estimate, although the propensity score blocking estimate is statistically different from the OLS estimate. The propensity score blocking regression model finds evidence of nonlinearity between wildfire behavior and wildfire management, although the same result is not found in the control score model. If the propensity score model results hold, it appears that fires respond differently to prescribed fire and suppression depending landscape, weather and climate, and socioeconomic attributes. This result implies that the returns to management may not be homogeneous across the landscape and that these models could be used by policy makers to target (locate) areas with the highest returns to wildfire management investment.

Prestemon et al. (2002) and Mercer et al. (forthcoming) both evaluate the effectiveness of prescribed fire on wildfire behavior. I am not aware of any other research that has quantified the effectiveness of fire crew response time, initial attack, or suppression in general. Overall, Prestemon et al. (2002) finds very little statistical evidence that hazard mitigating prescribed fire reduces wildfire size. In fact, they find some significant positive correlations between prescribed fire and fire size. They state a possible explanation is an omitted variable problem (unobserved confounder). Mercer et al. (forthcoming) find significant correlations between hazard mitigating prescribed fire and wildfire intensity-weight acres. The average prescribed fire treatment effect elasticity (the average prescribed fire effect over three years) is 0.27. Because the prescribed fire variable used in Mercer et al. (forthcoming) is defined as the natural log of the ratio of prescribed fire to forested area (in acres), a 1% increase in the ratio of prescribed fire to forested area is associated with a 0.27% reduction in wildfire intensity-weighted acres. Based on the prescribed fire and forest acre data present in their table 1 (Volusia County, Florida, 1994-2001), a 1% increase in prescribed fire acres is associated with a 0.0110% reduction in wildfire intensity-weighted acres. The OLS and 2SLS estimates presented above are not statistical different from the prescribed fire elasticity given in Mercer et al. (forthcoming). However, those significant block group treatment effects in the propensity score and main effects in the control score models are significantly different. This research coupled with the findings in Prestemon et al. (1995) and Mercer et al. (forthcoming) suggest that prescribed fire does appear to limit wildfire intensity, but may be less effective at limiting fire size. The control score prescribed fire results may imply that endogeneity exists and is not accounted for with the observable data, thus undervaluing the returns to prescribed fire.

SECTION REFERENCES

Brose, P. and D. Wade. 2002. “Potential Fire Behavior in Pine Flatwood Forests Following Three Different Fuel Reduction Techniques.” *Forest Ecology and Management* 163: 71-84.

Butry, D.T., Pye, J.M., and J.P. Prestemon. 2002. “Prescribed Fire in the Interface: Separating the People from the Trees.” In K.W. Outcalt, ed. *Proceedings of the Eleventh Biennial Southern Silvicultural Research Conference*. Asheville NC: U.S. Department of Agriculture, Forest Service, Southern Research Station, pp. 132-136. General Technical Report SRS-48.

Butry, D.T. Mercer, D.E., Prestemon, J.P., Pye, J.M., and T.P. Holmes. 2001. “What is the Price of Catastrophic Wildfire?” Journal of Forestry, 99(11):9-17.

EarthInfo, Inc. (2002) NCDC First Order Summary of the Day, Data on CD.

Florida Division of Forestry. (2005) Prescribed Fire Training Manual, Available at http://www.fl-dof.com/wildfire/rx_training.html.

Kennard, D.K. (2004) “Depth of Burn.” Forest Encyclopedia, Available at http://www.forestencyclopedia.net.

Maddala, G.S. (1983) Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge.

Mercer, D.E., Prestemon, J.P., Butry, D.T., and Pye, J.M. (Forthcoming). “Evaluating Alternative Prescribed Burning Policies to Reduce Net Economic Damages from Wildfire.” *American Journal of Agricultural Economics*.

National Interagency Fire Center. (2005) Fire Statistics available at <http://www.nifc.gov/stats/index.html>. Accessed by author on September 03, 2005.

National Oceanic and Atmospheric Administration. (2002) El Niño-Southern Oscillation Sea Surface Temperature Measures. Available at <ftp://ftp.ncep.noaa.gov/pub/cpc/wd52dg/data/indices/sstoi.indices>. Accessed by author on October, 2002.

Outcalt, K.W. and Wade, D.D. (2004) Fuels Management Reduces Tree Mortality from Wildfires in Southeastern United States. *Southern Journal of Applied Forestry*, 28, 28-34.

Prestemon, J.P. and D.T. Butry. 2005. “Time to Burn: Modeling Wildland Arson as an Autoregressive Crime Function.” *American Journal of Agricultural Economics* 87(3):756-770.

Prestemon, J.P., J.M. Pye, D.T. Butry, T.P. Holmes, and D.E. Mercer. 2002. “Understanding Broadscale Wildfire Risks in a Human-Dominated Landscape.” *Forest Science* 48(4): 685-693.

Rothermel, R.C. 1972. “A Mathematical Model for Predicting Fire Spread in Wildland Fuels.” U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. Research Paper INT-115.

Smith, J. and P. Todd. 2005. “Does Matching Overcome LaLonde’s Critique of Nonexperimental Estimators?” *Journal of Econometrics* 125: 305-353.

United States Census Bureau (2004) http://www.census.gov.

Wade, D.D., B.L. Brock, P.H. Brose, J.B. Grace, G.A. Hoch, G.A. Patterson, and W.A. Patterson, III. 2000. Chapter 4: Fire in Eastern Ecosystems. In: Brown, J.B. and J. K. Smith, eds. Wildland Fire in Ecosystems: Effects of Fire on Flora. General Technical Report RMRS-GTR-42-vol2. Ogden, UT: USDA Forest Service, Rocky Mountain Research Station. 257 pgs.

SECTION TABLES

Table 4.1. Descriptive statistics. Mean and standard deviation shown for continuous variables and sum shown for indicator variables.

VARIABLES	MEAN OR SUM	STD. DEV.	VARIABLES	MEAN OR SUM	STD. DEV
ln(Wildfire Intensity-Weighted Acres)	5.8749	(3.3285)	**Ignition**		
Management			Ignition Cause (base = lightning)		
ln(Response Time)	-5814.8091		*Campfire*	111	
ln(Prescribed Fire)	-31814.1645		*Cigarette*	71	
Let Burn	230		*Debris Burning*	900	
Climate & Weather			*Arson*	1908	
KBDI	422.3254	(174.0372)	*Equipment*	248	
La Nina	-0.2980	(0.2873)	*Railroad*	76	
El Nino	0.2545	(0.7401)	*Children*	472	
Humidity	48.7856	(13.8323)	*Unknown*	818	
Spread Potential	20.3448	(13.3531)	*Misc.*	545	
Spread Index (base = not observed)			ln(Population)	8.7548	(0.4648)
0 to 1 mph	3883		ln(Income)	10.8791	(0.2301)
2 mph	880		ln(Report Time)	-0.5074	(1.5463)
3 mph	311		Private Ownership	6851	
4 mph	78		Federal Ownership	81	
³5 mph	117		**County (base = Alachua)**		
ln(Wind Speed)	1.8924	(1.3735)	*Baker*	411	
Wind Direction (base = calm)			*Bradford*	8	
East	1071		*Brevard*	506	
North	541		*Clay*	526	
Northeast	878		*Columbia*	1	
Northwest	924		*Duval*	556	
South	522		*Flagler*	455	
Southeast	757		*Indian River*	235	
Southwest	975		*Lake*	507	
Variable	714		*Marion*	499	
West	961		*Nassau*	297	
Fuel			*Okeechobee*	85	
Fuel Type (base = palmetto-gallberry)			*Orange*	368	
Dense Pine	914		*Osceola*	182	
Swamp	493		*Polk*	57	
Blowy Leaf	359		*Putnam*	698	
Grassy Fuels	1446		*St. Johns*	530	
Muck	99		*St. Lucie*	6	
Other	392		*Seminole*	143	
Vegetation Build-Up	47.0861	(33.7113)	*Sumter*	18	
Longitude	642.0426	(47.4262)	*Union*	1	
Latitude	591.0584	(85.7503)	*Volusia*	1118	
Elevation	17.4051	(13.0145)	**Fire District (base = District 6)**		
Slope	0.2034	(0.2016)	*Fire District 07*	1375	
Upland Forest	0.3427	(0.2500)	*Fire District 08*	1453	
Agricultural Lands	0.1041	(0.1739)	*Fire District 10*	2033	
Rangelands	0.0716	(0.1123)	*Fire District 11*	549	
Residential Area	0.1586	(0.2092)	*Fire District 12*	1219	
Water	0.0171	(0.0382)	*Fire District 14*	67	
Wetland Forest	0.2178	(0.1827)	*Fire District 15*	1	
Forest Density	48.6891	(21.5516)	*Fire District 16*	332	
Previous Fire Ignitions	3.6152	(5.5115)	**Fire Year (base = 1996)**		
Previous Neighboring Fire Ignitions	4.0126	(4.0288)	*Fire Year 1997*	869	
ln(Previous Fire Intensity-Acres)	5.6996	(6.5500)	*Fire Year 1998*	1466	
ln(Previous Neighboring Fire Intensity-Acres)	7.6474	(5.4685)	*Fire Year 1999*	1425	
ln(Road Density)	1.2167	(2.0996)	*Fire Year 2000*	1592	
ln(Previous Non-Hazard Prescribed Fire)	-5.8968	(3.1590)	*Fire Year 2001*	1136	
ln(Previous Neighboring Non-Hazard Prescribed Fire)	-3.9795	(4.9330)	**Month (base = October)**		
Instruments			*November*	217	
Relative Humidity	121.0192	(28.2925)	*December*	254	
Wind Speed	339.4722	(33.0258)	*January*	616	
Average Maximum Temperature	322.6086	(6.6409)	*February*	765	
Precipitation	208.2099	(20.6224)	*March*	784	
ln(Distance to Fire Department)	2.1888	(1.0244)	*April*	815	
			May	1052	
			June	1386	
			July	752	
			August	529	
			September	187	

Table 4.2. Ordinary least squares wildfire model. ***, **, * Denotes significance at the 0.01, 0.05, 0.10 level, respectively.

VARIABLES	COEFFICIENT	STD. ERROR		VARIABLES	COEFFICIENT	STD. ERROR	
Intercept	7.4736	(3.2923)	**	**Ignition**			
Management				Ignition Cause (base = lightning)			
ln(Response Time)	0.4151	(0.0393)	***	*Campfire*	-0.0733	(0.2693)	
ln(Prescribed Fire)	-0.0159	(0.0071)	**	*Cigarette*	-0.2457	(0.3302)	
Let Burn	1.2262	(0.1842)	***	*Debris Burning*	-0.0884	(0.1329)	
Climate & Weather				*Arson*	0.2672	(0.1064)	**
KBDI	0.0000	(0.0003)		*Equipment*	-0.1697	(0.1915)	
La Nina	-0.4763	(0.2056)	**	*Railroad*	-0.5456	(0.3194)	*
El Nino	0.0213	(0.0618)		*Children*	-0.0891	(0.1562)	
Humidity	-0.0075	(0.0026)	***	*Unknown*	0.2671	(0.1294)	**
Spread Potential	0.0130	(0.0040)	***	*Misc.*	0.1963	(0.1389)	
Spread Index (base = not observed)				ln(Population)	-0.0695	(0.0753)	
0 to 1 mph	0.3712	(0.0737)	***	ln(Income)	0.2651	(0.1631)	
2 mph	3.6467	(0.1088)	***	ln(Report Time)	0.1557	(0.0214)	***
3 mph	4.8884	(0.1622)	***	Private Ownership	-0.2456	(0.1263)	*
4 mph	5.8721	(0.3071)	***	Federal Ownership	-0.0005	(0.3211)	
^{3}5 mph	6.2610	(0.2534)	***	**County (base = Alachua)**			
ln(Wind Speed)	-0.0730	(0.0678)		*Baker*	0.7412	(0.7031)	
Wind Direction (base = calm)				*Bradford*	0.4269	(1.1571)	
East	0.6351	(0.6272)		*Brevard*	-1.1293	(0.6934)	
North	0.6094	(0.6329)		*Clay*	0.9284	(0.5409)	*
Northeast	0.8885	(0.6289)		*Columbia*	5.5032	(2.7336)	**
Northwest	0.3277	(0.6355)		*Duval*	1.0631	(0.5560)	*
South	0.4547	(0.6321)		*Flagler*	-0.0368	(0.5992)	
Southeast	0.6464	(0.6295)		*Indian River*	-0.4196	(1.0938)	
Southwest	0.5768	(0.6310)		*Lake*	-1.4049	(0.6284)	**
Variable	0.5940	(0.6233)		*Marion*	-0.8089	(0.2354)	***
West	0.6936	(0.6314)		*Nassau*	1.4237	(0.5793)	**
Fuel				*Okeechobee*	-0.2490	(1.1025)	
Fuel Type (base = palmetto-gallberry)				*Orange*	-1.5964	(0.6464)	**
Dense Pine	-0.0905	(0.1037)		*Osceola*	0.4628	(0.7091)	
Swamp	-0.3283	(0.1329)	**	*Polk*	0.6621	(1.1135)	
Blowy Leaf	-1.2636	(0.1576)	***	*Putnam*	-0.2473	(0.2263)	
Grassy Fuels	-0.6873	(0.0987)	***	*St. Johns*	1.0112	(0.5834)	*
Muck	-0.4028	(0.2741)		*St. Lucie*	1.7922	(1.5489)	
Other	-1.3136	(0.1510)	***	*Seminole*	-1.1189	(0.6598)	*
Vegetation Build-Up	-0.0006	(0.0012)		*Sumter*	-1.8464	(0.8786)	**
Longitude	0.0037	(0.0031)		*Union*	-3.7227	(2.7304)	
Latitude	-0.0118	(0.0021)	***	*Volusia*	-0.3088	(0.6095)	
Elevation	0.0007	(0.0051)		**Fire District (base = District 6)**			
Slope	-0.0824	(0.1953)		*Fire District 07*	-0.6793	(0.4441)	
Upland Forest	0.5452	(0.2587)	**	*Fire District 08*	0.3961	(0.6691)	
Agricultural Lands	0.2082	(0.2794)		*Fire District 10*	-0.7455	(0.6351)	
Rangelands	1.8492	(0.3651)	***	*Fire District 11*	0.6374	(0.7543)	
Residential Area	-0.3001	(0.2661)		*Fire District 12*	-0.5092	(0.7049)	
Water	-0.1029	(0.8754)		*Fire District 14*	-1.4292	(1.0660)	
Wetland Forest	0.3883	(0.2930)		*Fire District 15*	-7.8229	(2.8671)	***
Forest Density	-0.0018	(0.0021)		*Fire District 16*	-2.6715	(1.0822)	**
Previous Fire Ignitions	-0.0108	(0.0065)	*	**Fire Year (base = 1996)**			
Previous Neighboring Fire Ignitions	0.0178	(0.0092)	*	*Fire Year 1997*	-0.1349	(0.1514)	
ln(Previous Fire Intensity-Acres)	0.0189	(0.0054)	***	*Fire Year 1998*	0.0217	(0.1556)	
ln(Previous Neighboring Fire Intensity-Acres)	0.0036	(0.0067)		*Fire Year 1999*	-0.1573	(0.1478)	
ln(Road Density)	-0.1396	(0.0180)	***	*Fire Year 2000*	-0.5599	(0.1279)	***
ln(Previous Non-Hazard Prescribed Fire)	0.0176	(0.0103)	*	*Fire Year 2001*	-0.0830	(0.1305)	
ln(Previous Neighboring Non-Hazard	0.0056	(0.0089)		**Month (base = October)**			
Prescribed Fire)				*November*	0.1722	(0.2976)	
				December	0.2320	(0.2934)	
				January	0.6537	(0.2642)	**
				February	0.6479	(0.2682)	**
				March	0.6636	(0.2723)	**
				April	0.6591	(0.2662)	**
				May	0.3182	(0.2634)	
				June	0.6134	(0.2678)	**
N	7480			*July*	0.7019	(0.2692)	***
F	42.5400			*August*	0.5751	(0.2721)	**
R-Squared	0.3840			*September*	0.6568	(0.3083)	**

Table 4.3. 1st stage two-stage least squares model, natural log of response time estimator. ***, **, * Denotes significance at the 0.01, 0.05, 0.10 level, respectively.

VARIABLES	COEFFICIENT	STD. ERROR		VARIABLES	COEFFICIENT	STD. ERROR	
Intercept	3.0636	(0.9891)	***	**Ignition**			
Management				Ignition Cause (base = lightning)			
ln(Prescribed Fire)	-0.0013	(0.0021)		*Campfire*	-0.0334	(0.0799)	
Let Burn	0.0985			*Cigarette*	-0.2324	(0.0978)	**
Climate & Weather				*Debris Burning*	-0.1523	(0.0394)	***
KBDI	0.0000	(0.0001)		*Arson*	-0.0617	(0.0316)	*
La Nina	-0.0011	(0.0610)		*Equipment*	-0.1143	(0.0568)	**
El Nino	0.0257	(0.0183)		*Railroad*	-0.1081	(0.0947)	
Humidity	0.0003	(0.0008)		*Children*	-0.1820	(0.0463)	***
Spread Potential	-0.0015			*Unknown*	-0.0534	(0.0384)	
Spread Index (base = not observed)				*Misc.*	-0.1184	(0.0412)	***
0 to 1 mph	-0.1149	(0.0218)	***	ln(Population)	0.0087	(0.0223)	
2 mph	-0.1666	(0.0322)	***	ln(Income)	0.0662	(0.0484)	
3 mph	-0.1785	(0.0480)	***	ln(Report Time)	0.0592	(0.0063)	***
4 mph	-0.0477	(0.0910)		Private Ownership	-0.0123	(0.0374)	
³5 mph	-0.0667	(0.0751)		Federal Ownership	-0.2034	(0.0951)	**
ln(Wind Speed)	-0.0135			**County (base = Alachua)**			
Wind Direction (base = calm)				*Baker*	-0.4010	(0.2140)	*
East	0.3502	(0.1859)	*	*Bradford*	-0.3828	(0.3431)	
North	0.3769	(0.1875)	**	*Brevard*	-0.3762	(0.2059)	*
Northeast	0.3361	(0.1864)	*	*Clay*	-0.2432	(0.1620)	
Northwest	0.3131	(0.1884)	*	*Columbia*	0.4164	(0.8117)	
South	0.2915	(0.1873)		*Duval*	-0.1707	(0.1686)	
Southeast	0.3735	(0.1865)	**	*Flagler*	-0.6309	(0.1779)	***
Southwest	0.3274	(0.1870)	*	*Indian River*	-0.7654	(0.3246)	**
Variable	0.3160	(0.1847)	*	*Lake*	-0.5877	(0.1878)	***
West	0.4021	(0.1871)	**	*Marion*	-0.3735	(0.0760)	***
Fuel				*Nassau*	-0.1891	(0.1757)	
Fuel Type (base = palmetto-gallberry)				*Okeechobee*	-0.5658	(0.3270)	*
Dense Pine	-0.0055	(0.0307)		*Orange*	-0.1118	(0.1940)	
Swamp	0.0991	(0.0394)	**	*Osceola*	-0.3745	(0.2129)	*
Blowy Leaf	-0.1081	(0.0467)	**	*Polk*	-1.0694	(0.3311)	***
Grassy Fuels	-0.1959	(0.0292)	***	*Putnam*	-0.0222	(0.0682)	
Muck	0.1170	(0.0812)		*St. Johns*	-0.3015	(0.1733)	*
Other	-0.0821	(0.0447)	*	*St. Lucie*	-0.8408	(0.4597)	*
Vegetation Build-Up	0.0001	(0.0004)		*Seminole*	-0.2332	(0.1960)	
Longitude	-0.0019	(0.0009)	**	*Sumter*	-0.4536	(0.2604)	*
Latitude	-0.0046	(0.0006)	***	*Union*	0.1898	(0.8097)	
Elevation	-0.0049	(0.0015)	***	*Volusia*	-0.7129	(0.1807)	***
Slope	-0.0016	(0.0579)		**Fire District (base = District 6)**			
Upland Forest	0.0918	(0.0768)		*Fire District 07*	0.0339	(0.1326)	
Agricultural Lands	-0.0509	(0.0831)		*Fire District 08*	-0.4075	(0.1985)	**
Rangelands	0.1418	(0.1083)		*Fire District 10*	0.1866	(0.1888)	
Residential Area	-0.0126	(0.0789)		*Fire District 11*	-0.2214	(0.2241)	
Water	0.3171	(0.2597)		*Fire District 12*	-0.5887	(0.2093)	***
Wetland Forest	0.1484	(0.0874)	*	*Fire District 14*	0.1549	(0.3163)	
Forest Density	-0.0016	(0.0006)	**	*Fire District 15*	-1.8533	(0.8495)	**
Previous Fire Ignitions	-0.0021	(0.0019)		*Fire District 16*	-0.4138	(0.3208)	
Previous Neighboring Fire Ignitions	-0.0017	(0.0027)		**Fire Year (base = 1996)**			
ln(Previous Fire Intensity-Acres)	-0.0049	(0.0016)	***	*Fire Year 1997*	-0.0041	(0.0449)	
ln(Previous Neighboring Fire Intensity-Acres)	0.0040	(0.0020)	**	*Fire Year 1998*	0.0240	(0.0461)	
ln(Road Density)	-0.0183	(0.0053)	***	*Fire Year 1999*	-0.0448	(0.0438)	
ln(Previous Non-Hazard Prescribed Fire)	0.0029	(0.0031)		*Fire Year 2000*	-0.0537	(0.0379)	
ln(Previous Neighboring Non-Hazard **Prescribed Fire)**	-0.0029	(0.0026)		*Fire Year 2001*	-0.1196	(0.0386)	***
				Month (base = October)			
Instrument				*November*	-0.0294	(0.0882)	
ln(Distance to Fire Department)	-0.0056	(0.0163)		*December*	0.0008	(0.0870)	
				January	-0.0783	(0.0783)	
				February	-0.0858	(0.0795)	
				March	-0.1182	(0.0807)	
				April	-0.1175	(0.0789)	
				May	-0.1366	(0.0781)	*
				June	-0.1536	(0.0794)	*
N	7480			*July*	-0.0762	(0.0798)	
F	9.2000			*August*	0.0307	(0.0806)	
R-Squared	0.1187			*September*	0.0258	(0.0914)	

Table 4.4. 1st stage two-stage least squares model, natural log of prescribed fire estimator. ***, **, * Denotes significance at the 0.01, 0.05, 0.10 level, respectively.

VARIABLES	COEFFICIENT	STD. ERROR	
Intercept	46.0126	(12.1425)	***
Instruments			
Relative Humidity	-0.0339	(0.0091)	***
Wind Speed	-0.0645	(0.0086)	***
Average Maximum Temperature	-0.0530	(0.0407)	
Precipitation	-0.0934	(0.0118)	***
N	7480		
LR Chi-Square	165.2200		
Pseudo R-Squared	0.0082		

Table 4.5. Two-stage least squares wildfire model (asymptotic standard errors shown). ***, **, * Denotes significance at the 0.01, 0.05, 0.10 level, respectively.

VARIABLES	COEFFICIENT	STD. ERROR		VARIABLES	COEFFICIENT	STD. ERROR	
Intercept	-6.4804	(0.0517)	***	**Ignition**			
Management				Ignition Cause (base = lightning)			
ln(Instrumented Response Time)	5.0541	(7.8621)		*Campfire*	0.0801	(0.4531)	
ln(Instrumented Prescribed Fire)	-0.0014	(0.0315)		*Cigarette*	0.8391	(0.5527)	
Let Burn	0.7692	(0.3131)	**	*Debris Burning*	0.6162	(0.2048)	***
Climate & Weather		0.0000		*Arson*	0.5528	(0.1438)	***
KBDI	-0.0001	(0.0005)		*Equipment*	0.3568	(0.2953)	
La Nina	-0.4729	(0.3441)		*Railroad*	-0.0447	(0.5223)	
El Nino	-0.1019	(0.0919)		*Children*	0.7508	(0.2200)	***
Humidity	-0.0090	(0.0045)	**	*Unknown*	0.5122	(0.1762)	***
Spread Potential	0.0203	(0.0063)	***	*Misc.*	0.7402	(0.2011)	***
Spread Index (base = not observed)				ln(Population)	-0.1074	(0.1255)	
0 to 1 mph	0.9041	(0.1250)	***	ln(Income)	-0.0517	(0.2645)	
2 mph	4.4197	(0.1656)	***	ln(Report Time)	-0.1188	(0.0347)	***
3 mph	5.7152	(0.2627)	***	Private Ownership	-0.1715	(0.1971)	
4 mph	6.1010	(0.5155)	***	Federal Ownership	0.9378	(0.4999)	*
^{3}5 mph	6.5692	(0.4226)	***	**County (base = Alachua)**			
ln(Wind Speed)	-0.0103	(0.0967)		*Baker*	2.6527	(1.1515)	**
Wind Direction (base = calm)		0.0000		*Bradford*	2.1678	(1.5995)	
East	-0.9868	(0.4021)	**	*Brevard*	0.5742	(0.9293)	
North	-1.1366	(0.2363)	***	*Clay*	2.0809	(0.5272)	***
Northeast	-0.6675	(0.1889)	***	*Columbia*	3.7014	(4.4815)	
Northwest	-1.1196	(0.1769)	***	*Duval*	1.9149	(0.2717)	***
South	-0.8951	(0.2170)	***	*Flagler*	2.8659	(0.5208)	***
Southeast	-1.0823	(0.1802)	***	*Indian River*	3.1084	(1.4111)	**
Southwest	-0.9400	(0.1613)	***	*Lake*	1.3458	(0.7691)	*
Variable	-0.8676	(0.1823)	***	*Marion*	0.9648	(0.3151)	***
West	-1.1685	(0.1593)	***	*Nassau*	2.3611	(0.2941)	***
Fuel				*Okeechobee*	2.3446	(0.5654)	***
Fuel Type (base = palmetto-gallberry)				*Orange*	-1.0804	(0.2957)	***
Dense Pine	-0.0635	(0.1757)		*Osceola*	2.1866	(0.3670)	***
Swamp	-0.7867	(0.2209)	***	*Polk*	5.6295	(1.5395)	***
Blowy Leaf	-0.7644	(0.2613)	***	*Putnam*	-0.1338	(0.2313)	
Grassy Fuels	0.2281	(0.1538)		*St. Johns*	2.4098	(0.2669)	***
Muck	-0.9461	(0.4601)	**	*St. Lucie*	5.6803	(1.8528)	***
Other	-0.9342	(0.2419)	***	*Seminole*	-0.0742	(0.3962)	
Vegetation Build-Up	-0.0013	(0.0017)		*Sumter*	0.2499	(1.0738)	
Longitude	0.0124	(0.0052)	**	*Union*	-4.5357	(4.4825)	
Latitude	0.0092	(0.0034)	***	*Volusia*	2.9639	(0.1894)	***
Elevation	0.0236	(0.0072)	***	**Fire District (base = District 6)**			
Slope	-0.0725	(0.3165)		*Fire District 07*	-0.8043	(0.2418)	***
Upland Forest	0.1189	(0.4322)		*Fire District 08*	2.3113	(0.1573)	***
Agricultural Lands	0.4452	(0.3698)		*Fire District 10*	-1.5683	(0.1267)	***
Rangelands	1.1737	(0.5218)	**	*Fire District 11*	1.7000	(0.2007)	***
Residential Area	-0.2564	(0.3057)		*Fire District 12*	2.2800	(0.1418)	***
Water	-1.5902	(1.4123)		*Fire District 14*	-2.0945	(0.5492)	***
Wetland Forest	-0.2909	(0.3484)		*Fire District 15*	0.8071	(4.4746)	
Forest Density	0.0056	(0.0031)	*	*Fire District 16*	-0.7177	(0.2513)	***
Previous Fire Ignitions	-0.0013	(0.0108)		**Fire Year (base = 1996)**		0.0000	
Previous Neighboring Fire Ignitions	0.0247	(0.0156)		*Fire Year 1997*	-0.1083	(0.2102)	
ln(Previous Fire Intensity-Acres)	0.0416	(0.0083)	***	*Fire Year 1998*	-0.0740	(0.1727)	
ln(Previous Neighboring Fire Intensity-Acres)	-0.0150	(0.0097)		*Fire Year 1999*	0.0453	(0.1461)	
ln(Road Density)	-0.0546	(0.0266)	**	*Fire Year 2000*	-0.3217	(0.1310)	**
ln(Previous Non-Hazard Prescribed Fire)	0.0029	(0.0172)		*Fire Year 2001*	0.4682	(0.1476)	***
ln(Previous Neighboring Non-Hazard Prescribed Fire)	0.0171	(0.0138)		**Month (base = October)**			
				November	0.3040	(0.4944)	
				December	0.2195	(0.3696)	
				January	1.0070	(0.2572)	***
				February	1.0323	(0.2067)	***
				March	1.2016	(0.1892)	***
				April	1.1961	(0.1790)	***
				May	0.9477	(0.1578)	***
				June	1.3214	(0.1389)	***
N	7480			*July*	1.0511	(0.1739)	***
F	40.8200			*August*	0.4295	(0.2036)	**
R-Squared	0.3743			*September*	0.5368	(0.3317)	

Table 4.6. Natural log of response propensity score estimator. ***, **, * Denotes significance at the 0.01, 0.05, 0.10 level, respectively.

VARIABLES	COEFFICIENT	STD. ERROR		VARIABLES	COEFFICIENT	STD. ERROR	
Intercept	3.2206	(0.9750)	***	**Ignition**			
Management				Ignition Cause (base = lightning)			
Let Burn	0.0735	(0.0545)		*Campfire*	-0.0329	(0.0798)	
ln(Prescribed Fire)	-0.0015	(0.0021)		*Cigarette*	-0.2339	(0.0978)	**
Climate & Weather				*Debris Burning*	-0.1472	(0.0394)	***
KBDI	0.0000	(0.0001)		*Arson*	-0.0688	(0.0317)	**
La Nina	0.0054	(0.0609)		*Equipment*	-0.1210	(0.0567)	**
El Nino	0.0287	(0.0183)		*Railroad*	-0.1175	(0.0946)	
Humidity	0.0006	(0.0008)		*Children*	-0.1776	(0.0463)	***
Spread Potential	-0.0013	(0.0012)		*Unknown*	-0.0567	(0.0384)	
ln(Wind Speed)	-0.0186	(0.0200)		*Misc.*	-0.1223	(0.0412)	***
Wind Direction (base = calm)				ln(Population)	0.0100	(0.0223)	
East	0.3666	(0.1856)	**	ln(Income)	0.0567	(0.0484)	
North	0.3962	(0.1873)	**	ln(Report Time)	0.0608	(0.0063)	***
Northeast	0.3481	(0.1861)	*	Private Ownership	-0.0041	(0.0374)	
Northwest	0.3309	(0.1881)	*	Federal Ownership	-0.2052	(0.0951)	**
South	0.3082	(0.1871)	*	**County (base = Alachua)**			
Southeast	0.3927	(0.1863)	**	*Baker*	-0.4505	(0.2083)	**
Southwest	0.3447	(0.1868)	*	*Bradford*	-0.3936	(0.3426)	
Variable	0.3315	(0.1845)	*	*Brevard*	-0.4191	(0.2056)	**
West	0.4149	(0.1869)	**	*Clay*	-0.2733	(0.1603)	*
Fuel				*Columbia*	0.2635	(0.8101)	
Fuel Type (base = palmetto-gallberry)				*Duval*	-0.2204	(0.1649)	
Dense Pine	0.0028	(0.0308)		*Flagler*	-0.6736	(0.1775)	***
Swamp	0.1134	(0.0394)	***	*Indian River*	-0.7456	(0.3242)	**
Blowy Leaf	-0.0702	(0.0467)		*Lake*	-0.6212	(0.1861)	***
Grassy Fuels	-0.1717	(0.0291)	***	*Marion*	-0.3932	(0.0696)	***
Muck	0.1427	(0.0813)	*	*Nassau*	-0.2213	(0.1717)	
Other	-0.0370	(0.0445)		*Okeechobee*	-0.5677	(0.3267)	*
Vegetation Build-Up	0.0001	(0.0004)		*Orange*	-0.1469	(0.1916)	
Longitude	-0.0021	(0.0009)	**	*Osceola*	-0.4555	(0.2103)	**
Latitude	-0.0047	(0.0006)	***	*Polk*	-1.1278	(0.3298)	***
Elevation	-0.0048	(0.0015)	***	*Putnam*	-0.0266	(0.0670)	
Slope	0.0133	(0.0579)		*St. Johns*	-0.3309	(0.1730)	*
Upland Forest	0.0471	(0.0768)		*St. Lucie*	-0.8511	(0.4593)	*
Agricultural Lands	-0.0669	(0.0828)		*Seminole*	-0.2513	(0.1957)	
Rangelands	0.0728	(0.1083)		*Sumter*	-0.4931	(0.2602)	*
Residential Area	0.0016	(0.0788)		*Union*	0.0044	(0.8096)	
Water	0.3143	(0.2594)		*Volusia*	-0.7402	(0.1805)	***
Wetland Forest	0.1001	(0.0868)		**Fire District (base = District 6)**			
Forest Density	-0.0017	(0.0006)	***	*Fire District 07*	0.0152	(0.1316)	
Previous Fire Ignitions	-0.0033	(0.0020)	*	*Fire District 08*	-0.4602	(0.1983)	**
Previous Neighboring Fire Ignitions	-0.0046	(0.0028)		*Fire District 10*	0.1972	(0.1882)	
ln(Previous Fire Intensity-Acres)	-0.0092	(0.0019)	***	*Fire District 11*	-0.2641	(0.2236)	
ln(Previous Neighboring Fire Intensity-Acres)	-0.0026	(0.0025)		*Fire District 12*	-0.6047	(0.2088)	***
ln(Road Density)	-0.0160	(0.0053)	***	*Fire District 14*	0.1000	(0.3159)	
ln(Previous Non-Hazard Prescribed Fire)	0.0028	(0.0030)		*Fire District 15*	-1.7740	(0.8495)	**
ln(Previous Neighboring Non-Hazard	-0.0025	(0.0026)		*Fire District 16*	-0.4900	(0.3208)	
Prescribed Fire)				**Fire Year (base = 1996)**		0.0000	
Balance				*Fire Year 1997*	-0.0109	(0.0449)	
ln(Previous Fire Intensity-Acres)^2	0.0009	(0.0002)	***	*Fire Year 1998*	0.0176	(0.0461)	
ln(Previous Neighboring Fire Intensity-Acres)^2	0.0010	(0.0002)	***	*Fire Year 1999*	-0.0448	(0.0438)	
ln(Report Time)^2	-0.0062	(0.0024)	**	*Fire Year 2000*	-0.0567	(0.0379)	
				Fire Year 2001	-0.1197	(0.0386)	***
				Month (base = October)			
				November	-0.0267	(0.0882)	
				December	0.0086	(0.0870)	
				January	-0.0764	(0.0783)	
				February	-0.0829	(0.0795)	
				March	-0.1190	(0.0807)	
				April	-0.1151	(0.0789)	
				May	-0.1346	(0.0781)	*
				June	-0.1504	(0.0794)	*
N	7480			*July*	-0.0803	(0.0798)	
F	9.4300			*August*	0.0214	(0.0806)	
R-Squared	0.1184			*September*	0.0204	(0.0914)	

Table 4.7.Natural log of prescribed fire propensity score estimator. ***, **, * Denotes significance at the 0.01, 0.05, 0.10 level, respectively.

VARIABLES	COEFFICIENT	STD. ERROR		VARIABLES	COEFFICIENT	STD. ERROR	
Intercept	-1.6920	(5.3223)		**Ignition**			
Climate & Weather				ln(Population)	-0.0565	(0.1239)	
La Nina	0.1630	(0.3304)		ln(Income)	1.1880	(0.2679)	***
El Nino	0.4719	(0.0936)	***	Private Ownership	-1.7712	(0.2057)	***
Fuel				Federal Ownership	0.5713	(0.5280)	
Fuel Type (base = palmetto-gallberry)				**County (base = Alachua)**			
Dense Pine	-0.0378	(0.1703)		*Baker*	2.7421	(1.1555)	**
Swamp	-0.0158	(0.2176)		*Bradford*	5.5685	(1.9012)	***
Blowy Leaf	-0.0025	(0.2579)		*Brevard*	6.4124	(1.1381)	***
Grassy Fuels	-0.4219	(0.1584)	***	*Clay*	1.5324	(0.8888)	*
Muck	-0.0434	(0.4494)		*Columbia*	-4.9883	(4.5007)	
Other	0.1660	(0.2409)		*Duval*	-0.5123	(0.9132)	
Longitude	-0.0274	(0.0051)	***	*Flagler*	4.8811	(0.9833)	***
Latitude	0.0050	(0.0035)		*Indian River*	5.4507	(1.7989)	***
Elevation	-0.0130	(0.0083)		*Lake*	1.6054	(1.0324)	
Slope	-0.3241	(0.3211)		*Marion*	0.8263	(0.3854)	**
Upland Forest	0.8549	(0.4238)	**	*Nassau*	-0.2038	(0.9519)	
Agricultural Lands	1.1064	(0.4562)	**	*Okeechobee*	5.3113	(1.8134)	***
Rangelands	2.5178	(0.5985)	***	*Orange*	5.2736	(1.0620)	***
Residential Area	1.1339	(0.4355)	***	*Osceola*	6.7939	(1.1637)	***
Water	-0.2857	(1.4396)		*Polk*	4.0212	(1.8287)	**
Wetland Forest	0.6703	(0.4803)		*Putnam*	1.2965	(0.3714)	***
Forest Density	0.0115	(0.0035)	***	*St. Johns*	1.8569	(0.9596)	*
Previous Fire Ignitions	-0.0174	(0.0106)		*St. Lucie*	4.8394	(2.5469)	*
Previous Neighboring Fire Ignitions	0.0810	(0.0152)	***	*Seminole*	6.0739	(1.0831)	***
ln(Previous Fire Intensity-Acres)	-0.0016	(0.0088)		*Sumter*	1.1778	(1.4444)	
ln(Previous Neighboring Fire Intensity-Acres)	0.0081	(0.0110)		*Union*	-2.8737	(4.4971)	
ln(Road Density)	-0.0374	(0.0295)		*Volusia*	5.0361	(0.9999)	***
ln(Previous Non-Hazard Prescribed Fire)	0.1458	(0.0168)	***	**Fire District (base = District 6)**			
ln(Previous Neighboring Non-Hazard Prescribed Fire)	0.2123	(0.0144)	***	*Fire District 07*	-0.6148	(0.7284)	
				Fire District 08	-0.9484	(1.0992)	
				Fire District 10	-1.7482	(1.0416)	*
				Fire District 11	-1.1342	(1.2389)	
				Fire District 12	-3.7378	(1.1556)	***
				Fire District 14	-3.0532	(1.7506)	*
				Fire District 15	-4.3040	(4.7209)	
				Fire District 16	-2.4835	(1.7790)	
				Fire Year (base = 1996)			
				Fire Year 1997	-0.8950	(0.2245)	***
				Fire Year 1998	-1.8606	(0.2375)	
				Fire Year 1999	0.3916	(0.2110)	*
				Fire Year 2000	0.8359	(0.1866)	***
				Fire Year 2001	0.2632	(0.2013)	
				Month (base = October)			
				November	0.3991	(0.4871)	
				December	0.4495	(0.4786)	
				January	0.5458	(0.4267)	
				February	0.9517	(0.4318)	**
				March	0.6258	(0.4392)	
				April	0.5679	(0.4309)	
				May	0.1785	(0.4255)	
				June	0.1877	(0.4229)	
N	7480			*July*	0.1910	(0.4297)	
F	20.9600			*August*	0.2253	(0.4355)	
R-Squared	0.1771			*September*	-0.1388	(0.5018)	

Table 4.8. Propensity score wildfire model results showing the treatment effect and associated standard errors (in parentheses). ***, **, * Denotes significance at the 0.01, 0.05, 0.10 level, respectively. The weighted average across blocks is 0.3175 (0.0420 standard error) and -0.0138 (0.0085 standard error) for response time and prescribed fire, respectively.

		Prescribed Fire Propensity Score		
		Lower Third	***Middle Third***	***Upper Third***
Response Time Propensity Score	***Lower Third***	ln(Response Time) = 0.4176*** (0.1091) ln(Prescribed Fire) = -0.0005 (0.0279) R2 = 0.4651 n = 850	ln(Response Time) = 0.1235 (0.1109) ln(Prescribed Fire) = -0.0436** (0.0206) R2 = 0.5671 n = 731	ln(Response Time) = 0.1026 (0.1111) ln(Prescribed Fire) = -0.0268 (0.0169) R2 = 0.4560 n = 913
	Middle Third	ln(Response Time) = 0.2858** (0.1265) ln(Prescribed Fire) = -0.0246 (0.0333) R2 = 0.3966 n = 906	ln(Response Time) = 0.3649*** (0.1317) ln(Prescribed Fire) = -0.0266 (0.0225) R2 = 0.4827 n = 868	ln(Response Time) = 0.2900** (0.1470) ln(Prescribed Fire) = -0.0136 (0.0220) R2 = 0.4566 n = 719
	Upper Third	ln(Response Time) = 0.5097*** (0.1366) ln(Prescribed Fire) = -0.0216 (0.0367) R2 = 0.4470 n = 738	ln(Response Time) = 0.4874*** (0.1216) ln(Prescribed Fire) = 0.0357 (0.0224) R2 = 0.4304 n = 894	ln(Response Time) = 0.2787** (0.1363) ln(Prescribed Fire) = -0.0084 (0.0203) R2 = 0.4150 n = 861

Table 4.9. Natural log of prescribed fire control score estimator. ***, **, * Denotes significance at the 0.01, 0.05, 0.10 level, respectively.

VARIABLES	COEFFICIENT	STD. ERROR	
Intercept	8.6209	(2.9329)	***
Instruments			
Relative Humidity	-0.0080	(0.0022)	***
Wind Speed	-0.0154	(0.0021)	***
Average Maximum Temperature	-0.0073	(0.0029)	**
Precipitation	-0.0207	(0.0029)	***
N	7480		
F	35.7300		
R-Squared	0.0188		

Table 4.10. Control score wildfire model. Only main treatment effects shown. ***, **, * denotes significance at the 0.01, 0.05, 0.10 level, respectively.

VARIABLES	COEFFICIENT	STD. ERROR		VARIABLES	COEFFICIENT	STD. ERROR	
Intercept	7.0743	(3.3171)	**	**Ignition**			
Management				Ignition Cause (base = lightning)			
ln(Instrumented Response Time)	0.4705	(0.1396)	***	*Campfire*	-0.0628	(0.2691)	
ln(Instrumented Prescribed Fire)	-0.0367	(0.0212)	*	*Cigarette*	-0.2477	(0.3305)	
Let Burn	1.2007	(0.1846)	***	*Debris Burning*	-0.0997	(0.1336)	
Climate & Weather				*Arson*	0.2738	(0.1065)	**
KBDI	0.0000	(0.0003)		*Equipment*	-0.1658	(0.1916)	
La Nina	-0.4803	(0.2058)	**	*Railroad*	-0.4931	(0.3200)	
El Nino	0.0190	(0.0618)		*Children*	-0.0731	(0.1574)	
Humidity	-0.0071	(0.0026)	***	*Unknown*	0.2694	(0.1294)	**
Spread Potential	0.0132	(0.0040)	***	*Misc.*	0.2051	(0.1393)	
Spread Index (base = not observed)				ln(Population)	-0.0728	(0.0754)	
0 to 1 mph	0.3791	(0.0743)	***	ln(Income)	0.2569	(0.1633)	
2 mph	3.6541	(0.1097)	***	ln(Report Time)	0.1536	(0.0219)	***
3 mph	4.9034	(0.1627)	***	Private Ownership	-0.2295	(0.1265)	*
4 mph	5.8779	(0.3071)	***	Federal Ownership	0.0314	(0.3216)	
^{3}5 mph	6.2592	(0.2538)	***	**County (base = Alachua)**			
ln(Wind Speed)	-0.0766	(0.0678)		*Baker*	0.6915	(0.7056)	
Wind Direction (base = calm)		0.0000		*Bradford*	0.3432	(1.1580)	
East	0.6508	(0.6277)		*Brevard*	-1.3439	(0.6981)	*
North	0.6315	(0.6337)		*Clay*	0.8333	(0.5433)	
Northeast	0.9035	(0.6294)		*Columbia*	5.3773	(2.7318)	**
Northwest	0.3361	(0.6361)		*Duval*	0.9913	(0.5584)	*
South	0.4636	(0.6326)		*Flagler*	-0.0915	(0.6039)	
Southeast	0.6628	(0.6301)		*Indian River*	-0.5201	(1.1004)	
Southwest	0.5930	(0.6314)		*Lake*	-1.3798	(0.6321)	**
Variable	0.6116	(0.6239)		*Marion*	-0.8041	(0.2389)	***
West	0.6987	(0.6323)		*Nassau*	1.3300	(0.5818)	**
Fuel				*Okeechobee*	-0.3861	(1.1083)	
Fuel Type (base = palmetto-gallberry)				*Orange*	-1.6092	(0.6501)	**
Dense Pine	-0.0948	(0.1036)		*Osceola*	0.3486	(0.7127)	
Swamp	-0.3334	(0.1332)	**	*Polk*	0.6468	(1.1183)	
Blowy Leaf	-1.2614	(0.1577)	***	*Putnam*	-0.2664	(0.2275)	
Grassy Fuels	-0.6751	(0.1003)	***	*St. Johns*	0.9678	(0.5852)	*
Muck	-0.3854	(0.2741)		*St. Lucie*	1.5225	(1.5583)	
Other	-1.2945	(0.1513)	***	*Seminole*	-1.1809	(0.6625)	*
Vegetation Build-Up	-0.0006	(0.0012)		*Sumter*	-1.8320	(0.8806)	**
Longitude	0.0041	(0.0031)		*Union*	-3.7952	(2.7286)	
Latitude	-0.0114	(0.0022)	***	*Volusia*	-0.3253	(0.6143)	
Elevation	0.0005	(0.0051)		**Fire District (base = District 6)**			
Slope	-0.0566	(0.1956)		*Fire District 07*	-0.6884	(0.4446)	
Upland Forest	0.5301	(0.2590)	**	*Fire District 08*	0.3260	(0.6716)	
Agricultural Lands	0.1770	(0.2800)		*Fire District 10*	-0.7766	(0.6358)	
Rangelands	1.8221	(0.3654)	***	*Fire District 11*	0.6317	(0.7559)	
Residential Area	-0.3100	(0.2664)		*Fire District 12*	-0.4913	(0.7088)	
Water	-0.1976	(0.8769)		*Fire District 14*	-1.4200	(1.0677)	
Wetland Forest	0.3814	(0.2933)		*Fire District 15*	-7.9103	(2.8877)	***
Forest Density	-0.0018	(0.0021)		*Fire District 16*	-2.6258	(1.0854)	**
Previous Fire Ignitions	-0.0111	(0.0065)	*	**Fire Year (base = 1996)**			
Previous Neighboring Fire Ignitions	0.0185	(0.0092)	**	*Fire Year 1997*	-0.1437	(0.1517)	
ln(Previous Fire Intensity-Acres)	0.0192	(0.0054)	***	*Fire Year 1998*	0.0298	(0.1562)	
ln(Previous Neighboring Fire Intensity-Acres)	0.0031	(0.0067)		*Fire Year 1999*	-0.1508	(0.1479)	
ln(Road Density)	-0.1369	(0.0180)	***	*Fire Year 2000*	-0.5421	(0.1281)	***
ln(Previous Non-Hazard Prescribed Fire)	0.0162	(0.0103)		*Fire Year 2001*	-0.0599	(0.1308)	
ln(Previous Neighboring Non-Hazard Prescribed Fire)	0.0054	(0.0089)		**Month (base = October)**			
Propensity Score Groups				*November*	0.1559	(0.2978)	
Group2	-0.2770	(0.3316)		*December*	0.2154	(0.2935)	
Group3	-0.0578	(0.3679)		*January*	0.6384	(0.2643)	**
Group4	-0.2702	(0.3574)		*February*	0.6447	(0.2683)	**
Group5	-0.0546	(0.3381)		*March*	0.6597	(0.2724)	**
Group6	0.6352	(0.3724)		*April*	0.6478	(0.2664)	**
Group7	-0.1734	(0.2614)		*May*	0.3156	(0.2637)	
Group8	-0.3432	(0.2636)		*June*	0.6118	(0.2682)	**
Group9	0.0019	(0.3137)		*July*	0.7034	(0.2695)	***
N	7480			*August*	0.5803	(0.2721)	**
F	36.8300			*September*	0.6647	(0.3084)	**
R-Squared	0.3869						

Table 4.11. Control score wildfire model (group effects) results showing the treatment effect and associated standard errors (in parentheses). The main effect corresponds with the lower third prescribed fire and response time control scores group—the only group not represented with a dummy variable in the model (the baseline). The other effects denote the marginal group treatment effects. The total group effect for each group is the main effect plus the marginal group effect. ***, **, * Denotes significance at the 0.01, 0.05, 0.10 level, respectively. The weighted average across groups is 0.4314 (0.0366 standard error) and -0.0356 (0.0003 standard error) for response time and prescribed fire, respectively.

		Prescribed Fire Control Score		
		Lower Third	***Middle Third***	***Upper Third***
Response Time Control Score	***Lower Third***	ln(Response Time) = 0.4705 (0.1396) ln(Prescribed Fire) = -0.0367* (0.0212) **MAIN EFFECT** n = 867	ln(Response Time) = -0.0804 (0.1915) ln(Prescribed Fire) = N/A n = 765	ln(Response Time) = -0.0417 (0.1887) ln(Prescribed Fire) = 0.0238 (0.0263) n = 862
	Middle Third	ln(Response Time) = -0.3620 (0.3572) ln(Prescribed Fire) = N/A n = 877	ln(Response Time) = -0.1586 (0.3553) ln(Prescribed Fire) = N/A n = 815	ln(Response Time) = 0.2607 (0.3331) ln(Prescribed Fire) = N/A n = 801
	Upper Third	ln(Response Time) = -0.1057 (0.2339) ln(Prescribed Fire) = N/A n = 811	ln(Response Time) =0.0281 (0.2096) ln(Prescribed Fire) = N/A n = 883	ln(Response Time) = 0.1300 (0.2086) ln(Prescribed Fire) = -0.0153 (0.0268) n = 799

Table 4.12. Estimated treatment effect elasticities and standard errors (in parentheses). * denotes weighted average elasticities and standard errors reported.

Estimators	Response Time	Prescribed Fire
Ordinary Least Squares	0.4151 (0.0393)	-0.0159 (0.0071)
Two-Stage Least Squares	5.0541 (7.8621)	-0.0014 (0.0315)
Propensity Score Blocking	0.3175 (0.0420)	-0.0138 (0.0085)
Control Score Regression	0.4314 (0.0366)	-0.0356 (0.0003)

CHAPTER FIGURES

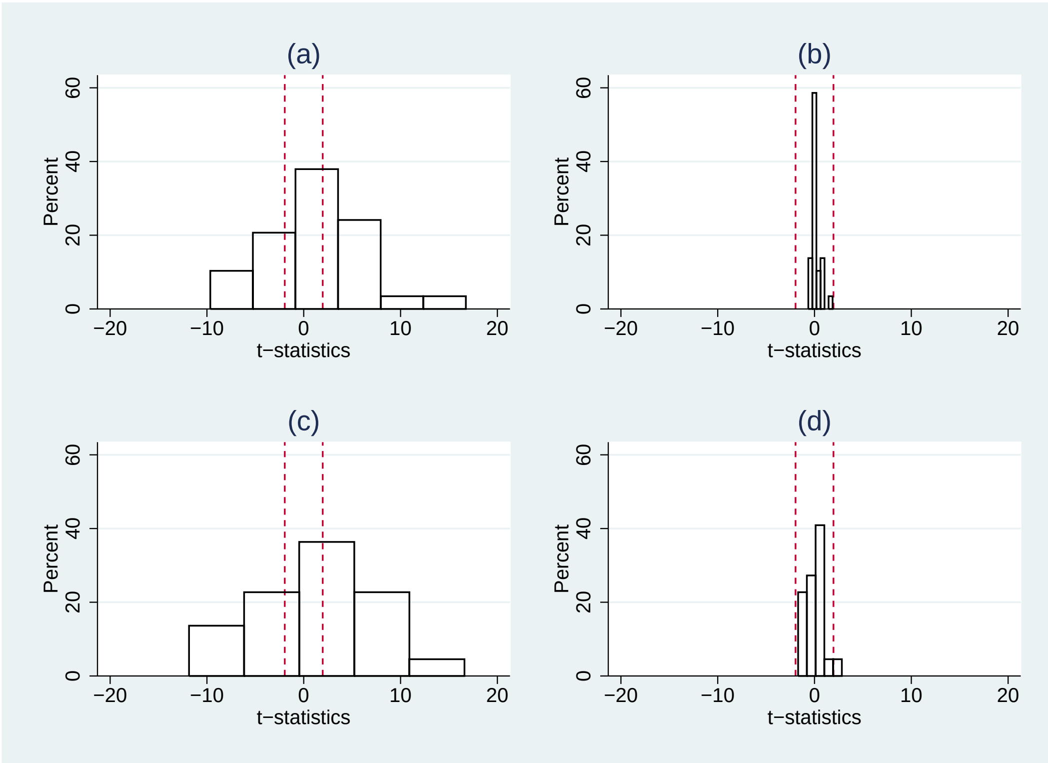

Figure 4.1. Histograms of the t-statistics associated with regressing a squared transformation of the covariates on (a) the natural log of response time, (b) the natural log of response time and the propensity score of the natural log of response time, (c) the natural log of prescribed fire, (d) the natural log of prescribed fire and the propensity score of the natural log of prescribed fire. Red dashed line denotes 5% critical values.

5. CONCLUSIONS

There are several methods to use when evaluating the effects of programs. The choice over the “best” program evaluation technique is difficult—is ordinary least squares, two-stage least squares, propensity score blocking, or control score blocking the right approach to use? Perhaps the decision should come from the assumptions the researcher is willing to make (observable confounders, potential for misspecification, unobservable confounders). The approaches are also sensitive to sample size. I summarize the findings below assuming a “large” sample size. In the simulation, a large sample is defined as 1,000 observations. The empirical results are based on a sample size of 7,480 observations.

5.1 THEORETICAL CONCLUSIONS

When confounders are observable and there is no functional form misspecification, ordinary least squares is an unbiased estimator. When misspecification occurs and valid instruments exist, two-stage least squares performs as well as propensity score blocking methods that have satisfied the balancing condition. If instrument validity is poor, meaning the instrument is correlated with the outcome error term, but the balancing condition is satisfied, then propensity score methods achieve less biased estimates of the endogenous treatment effect than ordinary least squares and two-stage least squares. In fact, it appears that the propensity score methods outperform the other methods even when the balancing condition is not perfectly satisfied—they are robust to imbalance. Even when instruments are invalid, two-stage least squares produces less biased population mean treatment effects than ordinary least squares if instrument quality (the correlation between the endogenous treatment and the instrument) is greater than the ratio of instrument invalidity (the correlation between the instrument and the outcome error term) and endogeneity.

When confounders are unobservable, ordinary least squares is outperformed (in terms of bias) by at least one other method, except when the correlation between treatment and its instrument is less than the ratio of correlation between the instrument and the outcome error term and endogeneity. In this case, the ordinary least squares population mean treatment effect is still biased, but just less so than the other methods. Thus, if only poor instruments exist and the confounding variables are unobserved, ordinary least squares will achieve the least bias of the four methods examined. If instruments are valid, however, two-stage least squares and control score methods outperform ordinary least squares. It also appears in this simulation, that when instruments are valid, two-stage least squares and control score blocking methods perform reasonably similar.

5.2 EMPIRICAL CONCLUSIONS

The question examined here is whether or not wildfire management pays off. Statistical difficulties in modeling wildfire behavior at policy relevant scales make quick answers challenging. Of the four models estimated, only the two-stage least squares model yielded insignificant wildfire management treatment effects. The response time coefficient was unbelievably large (an elasticity > 5). Again, using weak instruments in two-stage least square models has been shown to be outperformed by OLS estimation (Bound, Jaeger, and Baker 1995). The correlation between prescribed fire and its instrument is 0.14 and the correlation between response time and its instrument is 0.34, which are modest. The control score model, which requires good instruments, yielded estimates in line with the propensity score blocking approach for response time, but found larger prescribed fire treatment effects. As Smith and Todd (2005) discuss, propensity score matching will achieve low bias if there exists a rich set of confounders, if the comparison group comes from the same population as the treated group, and if the outcome variable is measured consistently across both comparison and treated groups. The wildfire data is rich and includes information related to values at risk, historic fire risk, forest and fuel conditions, climate, and historic weather. The treated and control groups do come from the same population—all wildfires in the SJRWMD—and have been measured consistently.

After accounting for potential endogeneity and nonlinearities of prescribed fire and fire crew response time with wildfire behavior, I find evidence that quicker response times limit wildfire size and intensity, and that prescribed fire may provide beneficial effects against wildfire extent and intensity up to three years of its application. In addition, the financial returns to wildfire management appear great. From 1993-2001, approximately 669,290 acres were prescribed burned in the SJRWMD for hazard mitigation purposes. Cleaves, Martinez, and Haines (2000) estimate prescribed fire to cost $26.30 an acre in the southeast US, thus roughly $17,602,327 was spent on prescribed fire in the SJRWMD. Based on the propensity score blocking prescribed fire elasticity of -0.0138, the SJRWMD experienced 30,651,129.12 kW-acre/meters less wildfire (1.38% less). Mercer et al. (forthcoming) estimate the market value of damage from wildfire to be $0.88 per kW-acre/meter, based on the findings of Butry (2001). Using this value as an approximation of the unit cost of wildfire, prescribed fire in the SJRWMD saved $26,972,993.63 in wildfire damages. Thus, for every $1 spent on prescribed fire treatments, $1.53 in wildfire damage was avoided. These numbers may be conservative given the control score model results, which find prescribed fire to have a larger impact on fire behavior than did the propensity score model.

It is not possible to calculate a similar estimate for suppression response (it is not known how large a fire would have become without any response). However, it is possible to estimate what the financial return would have been, from 1996-2001, if suppression response would have been 10% quicker. Using the propensity score blocking estimated response time elasticity of 0.3175 and the fire damage estimate of $0.88 kW-acre/meter, a 10% decrease in response time (a total reduction of 543.7 hours) would have resulted in $62,057,420 reduction in fire damages. Unfortunately, a dollar equivalent to the 543.7 hour reduction is not available. Future research might focus on evaluating the costs associated with suppression response.

SECTION REFERENCES

Bound, J., Jaeger, D.A., and R.M. Baker. 1995. “Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak.” *Journal of the American Statistical Association*, 90: 443-450.

Butry, D.T. Mercer, D.E., Prestemon, J.P., Pye, J.M., and T.P. Holmes. 2001. “What is the Price of Catastrophic Wildfire?” Journal of Forestry, 99(11):9-17.

Cleaves, D.A., Martinez, J., and Haines, T.K. 2000. Influences of Prescribed Burning Activity and Costs in the National Forest System. U.S. Department of Agriculture Forest Service, General Technical Report, SRS-37.

Mercer, D.E., Prestemon, J.P., Butry, D.T., and Pye, J.M. (Forthcoming). “Evaluating Alternative Prescribed Burning Policies to Reduce Net Economic Damages from Wildfire.” *American Journal of Agricultural Economics*.

Smith, J. and P. Todd. 2005. “Does Matching Overcome LaLonde’s Critique of Nonexperimental Estimators?” *Journal of Econometrics* 125: 305-353.