Estimating Student Responses to HOPE Scholarship Retention Rules Using Non-Parametric Matching Techniques

by

Brian Patrick Quinif

(Under the direction of Christopher M. Cornwell)

Abstract

Using data on University of Georgia students from the 1990 through 1995 entering classes, I estimate the effects of the HOPE scholarship's retention rules on a variety of variable related to the accumulation of credit hours using non-parametric matching techniques. I find that HOPE causes first-year students to complete about one credit hour less than they would otherwise and that this reduction is not accompanied by an increase in subsequent years, thus delaying graduation. I estimate that this effect varies depending on a student's race, gender, high school GPA, and family income. I also find that HOPE increases enrollment in summer courses, while there do not seem to be significant effects on the amount of credits being transferred to the university.

Index words: Matching Estimators, Non-Parametric Estimation, Merit-Based Financial Aid, Georgia's HOPE Scholarship, Withdrawal and Completion, Income, Transferred Credits

Estimating Student Responses to HOPE Scholarship Retention Rules Using Non-Parametric Matching Techniques

by

Brian Patrick Quinif

A Thesis Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Master of Arts

Athens, Georgia

2006

Estimating Student Responses to

HOPE Scholarship Retention Rules

Using Non-Parametric Matching Techniques

by

Brian Patrick Quinif

Approved:

Major Professor: Christopher M. Cornwell

Committee: William D. Lastrapes
David B. Mustard

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2006

Dedication

This thesis is dedicated to my parents, Alice and Nicholas Quinif, who have always encouraged me in all of my goals and aspirations. Their hard work has given me many opportunities; I truly am a "lucky brother."

Acknowledgments

I thank the Terry College of Business Institute for Leadership Advancement and Daniel Feldman for providing me with a grant which supported this research. The University of Georgia's Foundation Fellowship has supported me throughout my college career. In addition to inspiring me to study Economics, Christopher Cornwell offered invaluable help and guidance during the course of this project. I also thank William Lastrapes and David Mustard for their participation on my thesis committee as well as for teaching me various courses in Economics. I owe Theodore Shifrin a debt of gratitude for challenging me in all of his mathematics classes.

Table of Contents

List of Tables

Chapter 1

Introduction

Recently, there has been a sharp increase in number and size of merit-based college financial aid programs. In the past, the vast majority of monies allocated for college financial aid have been based on demonstrated financial need rather than academic merit. Of course, individual schools frequently had small merit-based scholarships intended to attract the most outstanding students, but these scholarships were generally quite small.

The largest of these relatively new merit-based programs is Georgia's HOPE (Helping Outstanding Pupils Educational) Scholarship. Founded in 1993 and funded entirely by state lottery revenues, the HOPE Scholarship pays the tuition and fees at any in-state public institution of higher education (or a flat $3,000 at in-state private institutions) for eligible legal Georgia residents. To be eligible, students must have graduated from high school with a "B" average (as defined by a 3.0 grade point average (GPA) or an 80 average on a 100-point scale). To keep the scholarship, students must maintain a GPA of 3.0 or higher. Students who do not graduate high school with a "B" average may qualify for HOPE during college if they earn a 3.0 GPA or higher.

Although the HOPE Scholarship initially had a family income cap, the cap was raised in 1994 and removed in 1995. Thus, since 1995 the HOPE Scholarship has been entirely merit-based and has become a virtual entitlement for all Georgia students. In the years since the implementation of HOPE in 1993, many other states have instituted their own merit-based programs modeled after HOPE. The interested reader can consult Cornwell, Leidner, and Mustard (2003) for a summary of these programs.

These programs are justified in a number of ways. Nearly all of them relate to the accumulation (or retention) of human capital in the state. One such justification is that merit-based aid will increase overall college enrollment in the state. Another related justification is that merit-based scholarships, by reducing the relative price of education for the most qualified students, will entice the best students to stay in-state for college and hopefully afterwards. These justifications are partially supported by the data, at least in Georgia, but the effects on enrollment are generally not as large as many proponents of merit-aid would like to believe (Cornwell, Mustard, and Sridhar 2005).

A third justification, which this study examines in detail, is that merit-based scholarships promote human capital accumulation by encouraging extra effort on the part of students in order to make a 3.0 GPA and qualify for or maintain the HOPE Scholarship. It should come as no surprise that there are numerous ways to improve one's likelihood of earning a 3.0 GPA, many of which may actually reduce human capital accumulation, such as taking easier classes or enrolling in fewer credit hours. Anecdotal evidence from the University of Georgia campus provides many examples of students using such options to increase their odds of keeping HOPE.

This study seeks to build on Lee (2004) and Cornwell, Lee, and Mustard (2005), which used difference-in-differences to estimate the effects of HOPE on a variety of outcomes, including GPA, credit hour adjustment, and course and major selection. First, I re-examine a number of effects that those studies examined but with the non-parametric matching estimator. I use the estimates in Lee (2004) as the benchmark and discuss when my estimates differ significantly. In addition to merely re-examining the HOPE effects previously studied, I also examine several new aspects of the HOPE effects. First, I examine the how the raising and eventual removal of the income cap for HOPE impacts the various HOPE effects. I also examine the differences in HOPE effects by both race and gender. Finally, I estimate the effects of HOPE on credits transferred from outside the university.

The main advantage of using matching estimators to examine and re-examine the effects of HOPE is that they are non-parametric. That is, matching does not impose a parametric form (let alone linear) on the outcomes of interest. If the parametric form of a model is indeed correct, then using a parametric estimator that takes advantage of this form will be efficient. Likewise, a non-parametric form would not be taking advantage of true assumptions and would therefore be inefficient. However, responses to HOPE might certainly be non-linear, especially binary choices like whether or not to withdraw from a course, so there are certainly gains to be had from using a non-parametric estimator.

Matching is primarily used to estimate average treatment effects. It "matches" treated and untreated individuals with similar covariates and then averages the differences in their outcomes of interest. In this case, the treatment variable is HOPE eligibility (that is, the product of the two dummies for Georgia residence and for the years in which HOPE was available), while an example of an outcome of interest is the number of credit hours students complete in an academic year. For example, a Georgia resident student in the high school class of 1995 is a treated individual, while all non-resident students and Georgia resident students from the class of 1990 are untreated individuals. Additionally, in this study, I am interested only in the effects that HOPE has on the individuals who are treated (i.e. those who were eligible for HOPE), so I estimate the average treatment effect of the treated (ATT). The average treatment effect (ATE) refers to the effect for both the treated and untreated individuals.

For most of the estimates in this study, I limit the data set to students from the high school classes of 1990 and 1995 who enrolled as typical first-time freshmen. I chose this subset of the population because the 1990 class is the last class to never have any eligibility for HOPE, and the 1995 class was the first class to have "full coverage" by HOPE (i.e. there were no income caps beginning with this class). I also use students from the 1993 and 1995 classes in order to examine the effect of removing the family income caps. As I mentioned above, HOPE was initially limited by family income, but by 1995, due to higher than expected lottery revenues, the income caps were totally eliminated

Overall, I estimate effects that are similar to those in (Lee 2004), with some important differences. I estimate that HOPE reduces the number of hours first-year students' complete by around one hour. Perhaps more important, however, is the fact that I estimate small and statistically insignificant effects on credit hour completion by second- and third-year students. Thus, students are delaying graduation because of HOPE. Lee's estimates, although imprecisely estimated, imply that students make up for their slowdown in credit accumulation by taking more hours in their second and third years, thus not delaying their graduation. I found that the HOPE effects vary throughout the GPA distribution and that students with low and mid-range high school GPAs were effected most by HOPE.

I used a variety of different specifications, including different numbers of matches and estimators adjusted and unadjusted for both bias and heteroscedasticity. I found my estimates to be very robust to different specifications, and I decided to use 12 matches and the bias-adjusted estimator for my results.

I did not find overwhelming evidence of increases in transferred credits, although students do slightly increase the number of hours they transfer in their third year. When the sample of first-year students is partitioned by race, I estimate that the effect on white students is actually stronger. That is, they reduce the number of credit hours they complete as first-year students by more than black students do. Similarly, the response among female students is stronger than for male students.

After examining the effects for the 1990 and 1995 classes, when HOPE was either nonexistent or fully in effect, I estimated the HOPE effects in 1993 and 1994, when there was an income cap that greatly reduced the number of HOPE recipients relative to 1995, when the income cap was eliminated. I find that the HOPE effect strengthens as the income cap is removed but at a rate less than proportional to the increase in the number of HOPE recipients, suggesting that wealthier students might be less affected by the HOPE retention rules.

The rest of this study is organized as follows. The second chapter provides the theoretical framework and notation for matching estimation with references to the estimation of HOPE effects. The third chapter analyzes some of these issues regarding the practical implementation of matching estimators: necessary assumptions and parameter choices. One of the main advantages of matching is its simplicity and minimal assumption set, but it is still important to address these issues. In the fourth chapter, I discuss the results from using matching estimators to measure the effects of the HOPE scholarship on UGA students. The final chapter provides concluding remarks, and the appendices contain numerous tables with the results of the estimations.

Chapter 2

Matching Estimation Theory

There exist many techniques for estimating average treatment effects, such as the effect of the HOPE scholarship on various outcomes, and estimation of average treatment effects has been the subject of increased interest by applied and theoretical econometricians in recent years. Reviews of the literature on estimating average treatment effects can be found in Angrist and Krueger (2000), Heckman, Lalonde, and Smith (2000), Blundell and Costa Dias (2002) and Imbens (2004).

Although there is considerable interest in the estimation of average treatment effects, there is no consensus regarding which estimator is "best" or even if there is an optimal estimator (Imbens 2004). The choice of which estimator to use will certainly depend on the specific situation. Moreover, each estimator has strengths and weakness, including the relative ease/difficulty of its implementation. Zhao (2004) uses Monte Carlo methods to test the properties of a variety of ATE estimators under different circumstances, and the results do not offer much guidance regarding the choice of estimator. My study focuses on matching estimators, but brief summaries of both propensity-score and regression estimators can also be found below. For a further discussion of the strengths and weaknesses of matching estimators, the reader can consult Smith and Todd (2001) and Smith and Todd (2005).

In an ideal situation, we would be able to randomly assign the treatment and estimate the average treatment effect by comparing the differences in outcomes of identical individuals. Unfortunately, the ability to randomly assign treatment is generally limited to scientific experiments, and in economic studies we are left with a problem of "unobserved outcomes."

That is, we have only individuals who are treated or not, and we cannot control who receives the treatment.

Matching estimators solve this unobserved outcome problem by matching individuals with a number of similar individuals who received the opposite treatment. The outcome of interest is first averaged over all matches. Then, the difference in this mean and the observed outcome of the individual is calculated to determine the estimated treatment effect for the individual. Averaging this estimated individual treatment effects over all of the population or sample of interest (all individuals, the treated individuals, or the control group) yields the simple matching estimator for the average treatment effect. Unfortunately, the simple matching estimator is biased when the matching is not exact (Abadie, Drukker, Herr, and Imbens 2001). A bias-corrected estimator does exist, however, and combines matching with regression in order to remove the bias. Additionally, matching estimators are generally not efficient (Abadie and Imbens 2002).

Matching estimators are often used settings with the following two characteristics: (i) the interest is in the average treatment effect for the treated (ATT), (ii), there is a large reservoir of potential controls (Imbens 2004). In studying the effect of HOPE, I am primarily interested in the effect the program has on recipients of the scholarship (i.e. the ATT). Data on University of Georgia students has many controls, and the probability of quality matches is quite high, providing an ideal setting for matching estimators.

One strength of matching estimators lies in their relative simplicity and ease of use. These estimators are based on a very minimal set of assumptions. Moreover, they only require the researcher to choose the number of matches to be used, whereas regression and propensity score estimators (see below) are far more complicated in their implementation (Imbens 2004). For example, estimating the propensity score is not very straightforward.

2.1 Notation

In general, we are interested in the average treatments effects of a binary treatment on individual units indexed by i. If we were able to perform true experiments, we would simply randomly apply the treatments to a variety of individuals. Unfortunately, such is not the case, and we are limited to "natural experiments" in which circumstances dictate who receives the treatment.

Let the observed outcome for individual i be denoted by Y_i:

$$Y_i = Y_i(W_i) = \begin{cases} Y_i(0) & \text{if} \quad W_i = 0, \\ Y_i(1) & \text{if} \quad W_i = 1, \end{cases} \tag{2.1}$$

where $W_i \in \{0, 1\}$ indicates the treatment received.

In my empirical work, I primarily use data on students in the 1990 and 1995 entering classes at UGA. In 1990, there was no HOPE, while in 1995 any Georgia resident student with a 3.0 GPA received HOPE. Clearly, all non-resident students in both years are controls, and the resident students in the 1990 are controls, as well, since they were never exposed to HOPE. Therefore, the treatment variable in this instance is a cross between the dummies for Georgia residency and the 1995 entering class. If it were possible to observe the outcomes for identical individuals with and without HOPE, then we could, in principle, calculate the average treatment effects, which we will call τ. Both a population average treatment effect (PATE) and a sample average treatment effect (SATE) exist and are given below:

$$\tau^{pop} = E[Y(1) - Y(0)] \quad \text{and} \quad \tau^{sample} = \frac{1}{N} \sum_{i=1}^{N} (Y_i(1) - Y_i(0)). \tag{2.2}$$

The distinction between the population effect and sample effects is subtle. Fortunately, the same matching estimator will estimate both, although the variance will differ. Generally, the variance of the SATE is smaller than the variance of the PATE. For the purposes of this paper, we will not prove this, but the interested reader can find a full discussion of this topic in Abadie and Imbens (2002) and Imbens (2004).

It should be clear that the ATEs above are calculated over both the treated and control units. The effect of interest, however, could be how the treated units were effected by the treatment or how the control units would respond to the treatment. This effect, known as the population ATT (PATT), and its sample counterpart (SATT), are defined as follows:

$$\tau^{pop,t} = E[Y(1) - Y(0)|W = 1] \quad \text{and} \quad \tau^{sample,t} = \frac{1}{N_1} \sum_{i|W_i=1} (Y_i(1) - Y_i(0)). \tag{2.3}$$

Additionally, one might be interested in the population or sample ATE for the individuals in the control group, known as the ATC:

$$\tau^{pop,c} = E[Y(1) - Y(0)|W = 0] \quad \text{and} \quad \tau^{sample,c} = \frac{1}{N_2} \sum_{i|W_i=0} (Y_i(1) - Y_i(0)), \tag{2.4}$$

where N_1 and N_2 are the number of treated and control units, respectively.

In this study, I am only interested in the response of the treated individuals, since it is primarily this effect on which the HOPE Scholarship is being justified. Therefore, all of the estimates reported are SATTs. In all of my tables, I report the number of treated observations, in addition to the number of observations, because it is upon the number of treated observations (N_1) that the SATTs are formed. For the sake of being complete, however, below I review all of the matching estimators (ATE, ATT, and ATC).

2.2 Basic Assumptions

In general, the consistent estimation of the average treatment effects requires two assumptions. The first is known as "unconfoundedness" (Rosenbaum and Rubin 1983), "selection on observables" (Barnow, Cain, and Golberger 1980) or the "conditional independence assumption" (Lechner 1999).

Assumption 2.2.1 (Unconfoundedness)

$$(Y(0), Y(1)) \quad \perp \quad W \mid X \tag{2.5}$$

Intuitively, this assumption is similar to the types of exogeneity assumptions made in linear regression models; that is, the errors are uncorrelated with the covariates. In fact, if the treatment effect is assumed to be constant, then these assumptions are identical. However, without this assumption, unconfoundedness is not equivalent to the regression assumption, but exogeneity is not a bad way to begin to think about the meaning of unconfoundedness (Imbens 2004).

Assumption 2.2.2 (Overlap)

$$0 < Pr(W = 1|X) < 1 \tag{2.6}$$

This assumption states that the probability of receiving the treatment conditional on X, know as the propensity score, is bounded away from 0 and 1. Intuitively, this means that the treatment and control groups "overlap." That is, for individuals with covariates in the neighborhood of X, there are both treated and control individuals. This assumption allows individuals to be "matched" with a similar individual that received the opposite treatment. In this study, the overlap assumption requires that for students with similar covariates (high school GPA, SAT score, etc), there will be a mixture of treated and untreated individuals.

2.3 Population v. Sample Effects

The distinction between the population and sample average treatment effects is somewhat subtle and is often ignored in recent papers. This distinction, however, has a long history in the analysis of randomized experiments (Neyman 1923). Somewhat unfortunately, given only the overlap and unconfoundedness assumptions, the sample effects contain no information about the effects on the broader population. However, a good estimator for one effect is necessarily a good estimator for the other. More precisely, the best estimator for the population effect is the sample effect (Imbens 2004). Fortunately, the estimator for the sample ATE will be consistent for the population ATE under a weaker assumption of mean independence, defined below (Imbens 2004).

Assumption 2.3.1 (Mean Independence)

$$E[Y(w)|W, X] = E[Y(w)|X] \tag{2.7}$$

Typically, research in economics has focused on the population effects, but these effects generally cannot be estimated as precisely. The reader should be aware that it is certainly possible to have an estimate of the population effect that is not statistically significant, while the sample effect is (Imbens 2004). In these cases, one must be careful about extrapolating results from the sample to the broader population.

2.4 Definitions

As was mentioned above, a primary problem in the estimation of average treatment effects is that only one of the potential outcomes is observed. All of the estimators we discuss seek to remedy this problem by imputing estimated values for the unobservable missing outcomes. This section and the following section concerning the estimation of variance follow the setup and discussion found in Abadie, Drukker, Herr, and Imbens (2001) very closely.

In order to match "similar" individuals, we must concern ourselves with the "distance" between the individuals in terms of their covariates. Let $\| x||_V = (x'Vx)^{1/2}$ be the vector norm, where V is a positive definite matrix of weights. The distance between two vectors x and z is then defined as $\| z - x \|_V$. It should be clear to the reader that if $V = I_n$, then this distance is merely the geometric distance as found using the n-dimensional generalization of the Pythagorean theorem (Shifrin 2004). However, we certainly might wish to place more weight on the similarity of certain variables as opposed to others, and the V matrix allows us to do so.

Moreover, we are not only interested in the distance between individuals but also the number and distance to nearby individuals in the control group. Let $d_M(i)$ be the distance from the covariates for unit i, X_i, to the Mth nearest match with the opposite treatment. Allowing for the possibility of ties, this is the distance such that strictly fewer than M units are closer to unit i than $d_M(i)$, and at least M units are as close as $d_M(i)$. Formally, $d_M(i)$ is the positive real number satisfying:

$$\sum_{l:W_l=1-W_i} 1\Big\{ \parallel X_l - X_i \parallel_V < d_M(i)\Big\} < M \quad \text{and} \sum_{l:W_l=1-W_i} 1\Big\{ \parallel X_l - X_i \parallel_V \leq d_M(i)\Big\} \geq M, \tag{2.8}$$

where $1\{\cdot\}$ is the indicator function, equal to one if the expression in brackets is true and zero otherwise. Let $\mathcal{J}(i)$ denote the set of indices for the matches for unit i that are at least as close as the Mth match:

$$\mathcal{J}_M(i) = \big\{l = 1, \ldots, N | W_l = 1 - W_i, \parallel X_l - X_i \parallel_V \leq d_M(i)\big\}. \tag{2.9}$$

Let the number of elements of $\mathcal{J}_M(i)$ be denoted by $\#\mathcal{J}_M(i)$. Note that if there are no ties, then $\#\mathcal{J}_M(i) = M$. In general, $\#\mathcal{J}_M(i) \geq M$. Now, let $K_M(i)$ denote the sum of the weights unit i has as a match for other units, and $K'_M(i)$ the sum of the squared weights of the matches:

$$K_M(i) = \sum_{l=1}^{N} 1\{i \in \mathcal{J}_M(l)\} \cdot \frac{1}{\#\mathcal{J}_M(l)} \tag{2.10}$$

$$K'_M(i) = \sum_{l=1}^{N} 1\{i \in \mathcal{J}_M(l)\} \cdot \left(\frac{1}{\#\mathcal{J}_M(l)}\right)^2. \tag{2.11}$$

Note that $\sum_i K_M(i) = N$, $\sum_{i:W_i=1} K_M(i) = N_0$, and $\sum_{i:W_i=0} K_M(i) = N_1$.

2.5 Identification

Unconfoundedness and overlap imply that the systematic differences in the outcomes between the treated and control units with the same covariate values are attributable to the treatment (Abadie and Imbens 2006). Therefore, under these assumptions, the average treatment effect is identified. Below I will discuss testing the validity of the identifying restrictions, as well as potential remedies for their violation.

2.6 Simple Matching Estimators

Now that we have completed the setup, defining at a simple matching estimator is very straightforward. Once the researcher has chosen the desired number of matches M, she simply estimates the unobserved outcome for individual i by averaging the outcomes of the M closest individuals who received the opposite treatment:

$$\hat{Y}_i(0) = \begin{cases} Y_i & \text{if} \quad W_i = 0, \\ \frac{1}{\#\mathcal{J}_M(i)} \sum\limits_{l \in \mathcal{J}_M(i)} Y_l & \text{if} \quad W_i = 1, \end{cases} \tag{2.12}$$

and

$$\hat{Y}_i(1) = \begin{cases} \frac{1}{\#\mathcal{J}_M(i)} \sum\limits_{l \in \mathcal{J}_M(i)} Y_l & \text{if} \quad W_i = 0, \\ Y_i & \text{if} \quad W_i = 1. \end{cases} \tag{2.13}$$

Note that at this point, we have no criterion on which to choose M. Clearly, the "quality" of the matches, as measured by the distance between the covariates of the individuals, should be taken into consideration. The issue of number and quality of matches will be discussed below. In my empirical work, I examine the sensitivity of the estimates to changes in the number of matches. The simple matching estimator for the ATE is defined as follows:

$$\hat{\tau}_M^{sm} = \frac{1}{N} \sum_{i=1}^{N} \left[\hat{Y}_i(1) - \hat{Y}_i(0) \right] = \frac{1}{N} \sum_{i=1}^{N} (2W_i - 1) \cdot [1 + K_M(i)] \cdot Y_i \tag{2.14}$$

We can modify this estimator in order to estimate the ATT or ATC:

$$\hat{\tau}_M^{sm,t} = \frac{1}{N_1} \sum_{i:W_i=1} \left[Y_i - \hat{Y}_i(0) \right] = \frac{1}{N_1} \sum_{i=1}^{N} \left[W_i - (1 - W_i) \cdot (1 + K_M(i)) \right] \cdot Y_i \tag{2.15}$$

$$\hat{\tau}_M^{sm,c} = \frac{1}{N_0} \sum_{i:W_i=0} \left[\hat{Y}_i(1) - Y_i \right] = \frac{1}{N_0} \sum_{i=1}^{N} \left[W_i \cdot K_M(i) - (1 - W_i) \right] \cdot Y_i \tag{2.16}$$

2.7 Bias Corrected Matching Estimator

Unfortunately, in small samples when matching is not exact, the simple matching estimator will be biased. When the covariates are continuous, the matches will certainly not be exact.

The bias is a result of matching discrepancies (the differences in covariates between matched units and their matches). In order to remove the bias, the bias corrected estimator adjusts the difference within the matches for the differences in their covariates, based on an estimate of the two regression functions $\mu_w(x) = E[Y(w)|X = x]$.

If the estimand of interest is the ATE, then we estimate the regression functions using only the data in the matched sample:

$$\hat{\mu}_w(x) = \hat{\beta}_{w0} + \hat{\beta}'_{w1}x, \tag{2.17}$$

for $w = 0, 1$, where

$$(\hat{\beta}_{w0}, \hat{\beta}_{w1}) = \text{argmin} \sum_{i:W_i=w} K_M(i) \cdot (Y_i - \beta_{w0} - \beta'_{w1}X_i)^2. \tag{2.18}$$

Similarly, for the ATT, it is only necessary to estimate the regression function for the controls, $\mu_0(x)$. For the ATC, only estimation of the regression function for the treated, $\mu_1(x)$ is needed. Once the regression functions have been estimated, then they are used to create "better" imputed values for the unobserved outcomes:

$$\tilde{Y}_i(0) = \begin{cases} Y_i & \text{if} \quad W_i = 0, \\ \frac{1}{\#\mathcal{J}_M(i)} \sum_{l \in \mathcal{J}_M(i)} [Y_l + \hat{\mu}_0(X_i) - \hat{\mu}_0(X_l)] & \text{if} \quad W_i = 1, \end{cases} \tag{2.19}$$

$$\tilde{Y}_i(1) = \begin{cases} \frac{1}{\#\mathcal{J}_M(i)} \sum_{l \in \mathcal{J}_M(i)} [Y_l + \hat{\mu}_0(X_i) - \hat{\mu}_0(X_l)] & \text{if} \quad W_i = 0, \\ Y_i & \text{if} \quad W_i = 1. \end{cases} \tag{2.20}$$

These imputed values are then substituted for $\hat{Y}_i(0)$ and $\hat{Y}_i(1)$ in the simple (biased) matching estimators to create an unbiased estimator. Using $\tilde{Y}_i$ in place of $\hat{Y}_i$ will generate an unbiased matching estimator. Accordingly, the bias-corrected matching estimator is the following:

$$\hat{\tau}_M^{bcm} = \frac{1}{N} \sum_{i=1}^{N} \left[\tilde{Y}_i(1) - \tilde{Y}_i(0) \right], \tag{2.21}$$

and the bias-corrected matching estimators for the ATT and ATC are, respectively:

$$\hat{\tau}_M^{bcm,t} = \frac{1}{N} \sum_{i:W_i=1} \left[Y_i - \tilde{Y}_i(0) \right] \quad \text{and} \quad \hat{\tau}_M^{bcm,c} = \frac{1}{N_1} \sum_{i:W_i=0} \left[\tilde{Y}_i(1) - Y_i \right]. \tag{2.22}$$

In the estimations below, I compare the bias-corrected and non-corrected matching estimators and do not generally find large differences in their magnitudes for the ATTs of HOPE, but this is certainly not always the case. Given the ease with which the bias-correction can be made, it is a good idea to compute both estimators.

2.8 Alternatives: Propensity Score and Regression Estimators

Common alternatives to matching include Propensity Score and Regression Estimators. In this section, I briefly review both in relation to matching. The interested reader can find a more thorough treatment in Becker and Ichino (2002), where the basic setup is outlined and implementation using STATA is thoroughly explained.

Rosenbaum and Rubin (1983) define the propensity score as the conditional probability of receiving a treatment given pre-treatment characteristics:

$$p(X) \equiv Pr(W = 1|X) = E(W|X). \tag{2.23}$$

If the propensity score is known, then the average treatment effect of the treated $\tau^{pop,t}$ is given by

$$\begin{aligned} \tau^{pop,t} &\equiv E(Y_i(1) - Y_i(0)|W_i = 1) && (2.24)\\ &= E[E(Y_i(1) - Y_i(0)|W_i = 1, p(X_i))] && (2.25)\\ &= E\{E(Y_i(1)|W_i = 1, p(X_i)) - E[Y_i(0)|W_i = 0, p(X_i)]|W_i = 1\}. && (2.26) \end{aligned}$$

The propensity score estimator of the average treatment effect is based on the two following assumptions.

Assumption 2.8.1 (Balancing given the propensity score)

$$W \perp X \quad | \quad p(X) \tag{2.27}$$

Assumption 2.8.2 (Unconfoundedness given the Propensity Score)

$$Y_1, Y_0 \perp W \quad | \quad X \quad \Longrightarrow \quad Y_1, Y_0 \perp W \quad | \quad p(X) \tag{2.28}$$

If assumptions 3.5.1 and 3.5.2 are true, then observations with the same propensity score must have the same distribution of observable (and unobservable) characteristics independent of the treatment received. That is, for a given propensity score, exposure to treatment is random; therefore, treated and control units should be on average observationally identical (Becker and Ichino 2002). Then, the propensity scores can be estimated (via a probit or logit model) and then used to calculate the propensity score estimator. Becker and Ichino (2002) outline this estimator and its implementation in STATA.

It is worth noting here that a major drawback to the use of propensity score estimators is that the true propensity score is rarely, if ever, known. Thus, must be estimated. When matching is done on an estimated propensity score, there are no principled standard errors. Little is known about the asymptotic theory of propensity score estimators when an estimated propensity score is used. On the other hand, nearest-neighbor matching estimators, which is a more descriptive name for the "matching estimators" I use in this paper, do have principled standard errors and well-derived asymptotic theory (Abadie and Imbens 2006).

Propensity score estimators are useful when we are interested in the probability that an individual selects the treatment as part of an optimization process. In the case of HOPE, students do not select their treatment status, so the propensity score estimator is of little interest for this study.

As I briefly outlined above, regression estimators rely on consistent estimation of conditional regression functions (Imbens 2004). Let these regression functions be denoted by $\mu_w(x)$ for $w = 0, 1$. Given consistent estimates, $\hat{\mu}_w(x)$, the regression estimators for the PATE and SATE are found by averaging their difference over the empirical distribution of the covariates:

$$\hat{\tau}_{reg}^{pop} = E\big[\hat{\mu}_1(X) - \hat{\mu}_0(X)\big] \quad \text{and} \quad \hat{\tau}_{reg}^{sample} = \frac{1}{N}\sum_{i=1}^{N}\Big[\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)\Big]. \tag{2.29}$$

Since we are not primarily concerned with regression estimators, we will not give the forms of the estimators of the ATT and ATC, but the interested reader can find a full treatment in Imbens (2004). Imbens also outlines the various parametric and nonparametric estimators of $\mu_w(x)$ that have been employed by Rubin (1977), Hahn (1998), and Heckman, Ichimura, and Todd (1997). Regression estimators require decisions about how to estimate $\mu_w(x)$ and about both the smoothing and bandwidth parameters (Imbens 2004). In other words, they are much more complicated to implement than matching estimators. However, it is certainly worth noting that the bias-corrected matching estimator mentioned above and used below corrects for the bias using a regression correction (Imbens 2004).

2.9 Matching Estimator Variance Estimation

The variance of the matching estimator will clearly depend on which estimator is implemented; that is, the variance has a different form, dependent on whether the ATE, ATT, or ATC is estimated and also whether the effect is for the population or the sample. However, the bias-correction does not affect the form of the variance of the estimator, although it may affect its numerical value (Abadie, Drukker, Herr, and Imbens 2001).

Below are the estimators for the variance of the sample average treatment effect (SATE), followed by the estimators for the SATT and SATC, respectively (Abadie, Drukker, Herr, and Imbens 2001).

$$\hat{V}^{sample} = \frac{1}{N^2}\sum_{i=1}^{N}\left[1+K_M(i)\right]^2 \cdot \hat{\sigma}^2_{W_i}(X_i) \tag{2.30}$$

$$\hat{V}^{sample,t} = \frac{1}{N_1^2}\sum_{i=1}^{N}\left[W_i-(1-W_i)\cdot K_M(i)\right]^2 \cdot \hat{\sigma}^2_{W_i}(X_i) \tag{2.31}$$

$$\hat{V}^{sample,c} = \frac{1}{N_2^2}\sum_{i=1}^{N}\Big[W_i\cdot K_M(i)-(1-W_i)\Big]^2 \cdot \hat{\sigma}^2_{W_i}(X_i), \tag{2.32}$$

where $\sigma^2_{W_i}(X_i)$ is the conditional error variance. The form of this variance will depend on whether there is a constant treatment effect and whether homoscedasticity or heteroscedasticity is assumed. I treat both cases below.

Correspondingly, the estimators for the variances of the population effects are:.

$$\hat{V}^{POP} = \frac{1}{N^2}\sum_{i=1}^{N}\left\{\left[\hat{Y}_i(1) - \hat{Y}_i(0) - \hat{\tau}\right]^2 + \left[K_M^2(i) + 2K_M(i) - K_M'(i)\right]\cdot\hat{\sigma}^2_{W_i}(X_i)\right\} \quad (2.33)$$

$$\hat{V}^{POP,t} = \frac{1}{N_1^2}\sum_{i=1}^{N}\left\{W_i\cdot\left[Y_i(1) - \hat{Y}_i(0) - \hat{\tau}^t\right]^2 + (1-W_i)\cdot\left[K_M^2(i) - K_M'(i)\right]\cdot\hat{\sigma}^2_{W_i}(X_i)\right\} \quad (2.34)$$

$$\hat{V}^{POP,c} = \frac{1}{N_2^2}\sum_{i=1}^{N}\left\{(1-W_i)\cdot\left[\hat{Y}_i(1) - Y_i(0) - \hat{\tau}^c\right]^2 + W_i\cdot\left[K_M^2(i) - K_M'(i)\right]\cdot\hat{\sigma}^2_{W_i}(X_i)\right\} \quad (2.35)$$

Abadie and Imbens (2002) note that in large samples $\hat{V}^{POP}$ will be at least as large as $\hat{V}^{sample}$. In small samples, however, it may be smaller, so they recommend using the larger of $\hat{V}^{POP}$ and $\hat{V}^{sample}$ as the estimator for the variance of τ^{pop}. Clearly, the form of $\hat{\sigma}^2_{w_i}$ will depend on the scedastic function. Our estimate of $\hat{\sigma}^2_{wi}$ will differ based upon homoscedasticity or heteroscedasticity is assumed (Abadie, Drukker, Herr, and Imbens 2001). In the following sections, I treat both cases.

2.9.1 Homoscedasticity and Constant Treatment Effect

In this section, estimates for $\sigma^2_{W_i}$ will be based on two assumptions. First, the treatment effect for individual i, $\tau_i = Y_i(1) - Y_i(0)$ is assumed to be constant for all individuals. Additionally, the conditional variance of $Y_i(w)$ given X_i is assumed to be independent of the covariates and the treatment. Under these two assumptions, the variance of the outcome $\hat{\sigma}^2_{w_i} = \sigma^2 \forall i$ and w and is estimated as

$$\hat{\sigma}^2 = \frac{1}{2N}\sum_{i=1}^{N}\left\{\frac{1}{\mathcal{J}_M(i)}\sum_{l\in\mathcal{J}_M(i)}\left[W_i\cdot(Y_i - Y_l - \hat{\tau}) + (1-W_i)\cdot(Y_l - Y_i - \hat{\tau})\right]^2\right\} \quad (2.36)$$

Therefore, $\hat{\sigma}^2$ can be substituted for $\hat{\sigma}^2_{w_i}$ in all of the variance expressions above and the variances for the ATT and ATC are given by

$$\hat{\sigma}^2_t = \frac{1}{2N_1}\sum_{i:W_i=1}\left[\frac{1}{\mathcal{J}_M(i)}\sum_{l\in\mathcal{J}_M(i)}(Y_i - Y_l - \hat{\tau})^2\right] \quad (2.37)$$

and

$$\hat{\sigma}_c^2 = \frac{1}{2N_2} \sum_{i:W_i=0} \left[\frac{1}{\mathcal{J}_M(i)} \sum_{l \in \mathcal{J}_M(i)} (Y_i - Y_l - \hat{\tau})^2 \right] \tag{2.38}$$

2.9.2 Heteroscedasticity

Under heteroscedasticity, $\hat{\sigma}^2_{w_i}$ will vary over both x and w, so it is necessary to form estimates for all x and w. These estimates are formed by matching individuals with the *same* treatments (treated matched with treated and controls matched with controls), and the number of matches used in the correction for heteroscedasticity does not have to be equal to the number of matches used in estimating the treatment effect. The notation below is somewhat tedious and follows Abadie, Drukker, Herr, and Imbens (2001) closely, but it should not be difficult for the reader to understand it once she has interpreted the similar notation in the section above providing the basic setup for matching estimators. Let us define $d'_M(i)$ as the distance to the Mth closest individual receiving the same treatment. Formally, $d'_M(i)$ is defined as the positive real number satisfying

$$\sum_{l:W_l=W_i, l\neq i} 1\Big\{ \parallel X_l - X_i \parallel_V < d'_M(i) \Big\} < M \quad \text{and} \sum_{l:W_l=W_i, l\neq i} 1\Big\{ \parallel X_l - X_i \parallel_V \leq d'_M(i) \Big\} \geq M. \tag{2.39}$$

Similarly to above, $\mathcal{J}'_M(i)$ is defined as the set of indices for the first M matches for individual i:

$$\mathcal{J}'_M(i) = \big\{ l = 1, \ldots, N | W_j = W_i, j \neq i \parallel X_j - X_i \parallel_V \leq d'_M(i) \big\}. \tag{2.40}$$

Again, consistent with previous notation, $\#\mathcal{J}'_M(i)$ denotes the number of elements in $\mathcal{J}'_M(i)$. The conditional variance is then estimated as the sample variance for the set that is the union of $\mathcal{J}'_M(i)$ and individual i: $\mathcal{J}'_M(i) \cup \{i\}$. This sample variance is given by

$$\tilde{\sigma}^2_{W_i}(X_i) = \frac{1}{\#\mathcal{J}'_M(i)} \sum_{j \in (\mathcal{J}'_M(i) \cup \{i\})} \left(Y_j - \bar{Y}_{\mathcal{J}'_M(i) \cup \{i\}} \right)^2, \tag{2.41}$$

where

$$\bar{Y}_{\mathcal{J}'_M(i) \cup \{i\}} = \frac{1}{\#\mathcal{J}'_M(i) + 1} \sum_{j \in (\mathcal{J}'_M(i) \cup \{i\})} Y_j. \tag{2.42}$$

Both STATA and R have options in their matching routines which allow for the user to correct for heteroscedasticity using a selected number of matches which all but removes the complication involved in correcting for heteroscedasticity. Unfortunately, however, computing the heteroscedasticity consistent standard errors greatly increases the computing time necessary, so it might be wise to initially compute estimates using the homoscedasticity assumption. Generally, if one does not have good reason to believe in homoscedasticity, it is wise to use the robust option. At a minimum, the robust and normal standard errors could both be computed to see if homoscedasticity/heteroscedasticity margin has any effect on the statistical significance of the estimates.

Chapter 3

Practical Implementation Issues

Among the three sets of estimators outlined in this paper, matching procedures generally require the least effort on the part of the researcher. The researcher must only choose the covariates to use for matching and the number of matches to use. The only assumptions to be satisfied are unconfoundedness and overlap. Below I discuss some of the primary practical issues with regard to the implementation of matching estimators.

3.1 Specification and Assumptions

In any model, whether it is parametric or non-parametric, specification and assumptions are critically important. While the non-parametric nature of matching estimation does allow us to have a very loosely specified model, we still must choose which covariates are important for a match. It should be immediately clear that certain assumptions must be met for matching estimation to be appropriate.

3.1.1 Choice of Covariates

Clearly, when the researcher decides which covariates will be used to match individuals, she should use economic theory to determine what characteristics should influence the outcomes of interest. It should be no surprise that the matching estimates may be sensitive to the specification used. Some variables should not be adjusted for, and their inclusion as covariates may lead to efficiency loss. Additionally, the mean squared error (MSE) of the estimator can potentially be reduced by ignoring some of the covariates that have only weak correlation with the treatment indicator and the outcomes (Imbens 2004).

It is known that in large samples including a covariate will not lower asymptotic precision of the estimator if the assumptions are correct. In small samples this is not the case, and there are limited procedures for optimally choosing the set of covariates (Imbens 2004). As I have repeatedly emphasized, it is wise to try various specifications to test for robustness. If results are not robust to different specifications, then perhaps one should be extra diligent in her theoretical justification for the specification used.

3.1.2 Overlap Assumption

If there is no "overlap," then there will be no similar individuals in both the treated and control groups. Therefore, this assumption is crucial for being able to make good matches. In principle, this assumption is directly testable, since it restricts the joint distribution of the observable variables. However, direct tests are generally of interest (Imbens 2004). Several methods exist for detecting if there is a lack of overlap, and there are also methods for correcting this violation.

One method that is of primary interest when individuals are matched over only one or two covariates is to plot distributions of the covariates by treatment groups (Imbens 2004). One can then compare these distributions and check that they do overlap sufficiently. Figures 2-4 of Lee (2004) show that there is substantial overlap in the distributions of SAT scores of resident and non-resident students. It is safe to assume that the overlap assumption is satisfied in this situation.

Another method that is easier to implement when more covariates are used for matching is to inspect the distribution of the propensity score for both treatment groups (Imbens 2004). This requires nonparametric estimation of the propensity scores, after which it is possible to check directly for overlap. A third, and simpler method, is to check the quality of the worst matches. The researcher can examine the difference, component-by-component, in the vectors of covariates of matched individuals. If these differences are sufficiently large (relative to the sample standard deviation of the components of the elements of the covariate vectors) there may be reason to worry that overlap has been violated (Imbens 2004).

Once it has been determined that the overlap assumption is suspect, there are a few ways to correct the problem. One solution is to drop observations with propensity scores very close to 0 and 1. What determines "close" is a function of the sample size. However, it should be noted that dropping observations will change the average treatment effect that is being estimated (Imbens 2004). Another option is to discard observations with poor match quality. Rosenbaum and Rubin (1983) suggest accepting only matches where the difference in propensity score values between matches is small. Another similar method is to drop matches where other covariates are very poorly matched.

3.1.3 Unconfoundedness

The assumption of unconfoundedness, also known as "selection on observables" is crucial to identification when using matching estimators. Violations of this assumption can jeopardize the validity of the estimates generated by matching (or the other methods, as well). If the individuals choose their own treatment, unconfoundedness may very well be violated. In other words, if selection into the treatment is the result of optimizing behavior, it may be quite difficult to justify that the treatment is exogenous. However, unconfoundedness can possibly be maintained if the treatment is chosen optimally. This can happen if the choice of treatment is driven by unobserved characteristics that are unrelated to the outcomes that interest the researcher (Imbens 2004).

Ichino, Mealli, and Nannicini (2005) warn that the increasing ease with which matching estimators can be implemented may lure researchers into using matching estimators without properly justifying this and other assumptions. They outline a process to conduct a sensitivity analysis to aid in identifying cases in which the estimated effect should not be considered robust to violations of the unconfoundedness assumption.

3.1.4 Number/Quality of Matches

Unfortunately, little is currently known about the optimal number of matches or about data-dependent ways of choosing the number of matches (Imbens 2004). In some ways, this problem is similar to issues that arise in instrumental variables applications. We know that instruments must be orthogonal to the errors and partially correlated with the covariates for which they are used as instruments, but their is no "optimal" level of correlation. We do know, however, that within the class of matching estimators, using only a single match yields the most credible inference with the least bias, although using only a single match can lead to some loss of efficiency (Imbens 2004). This clearly does not imply, that using a single match is the best way to go.

Abadie and Imbens (2002) point out that matching estimators are not generally efficient. To attain efficiency, one would need to increase the number of matches with the sample size. But even here, the notion of "efficiency" is relative. With a given data set, if one can use M matches and calculates the variance as if M would increase at the appropriate rate with the sample size, then the estimator would be efficient. However, if one calculates the variance conditional on the number of matches, then the estimator would not be efficient (Imbens 2004). Therefore, even thinking about what efficiency means in the context of matching estimators is complex.

Additionally, as I mention above, inexact matches will lead the simple matching estimator to be biased, so anytime continuous covariates are used, the bias-corrected estimator should certainly be considered. In this project, I compared the results using both the standard and bias-adjusted estimators. The results of these comparisons did not find the bias adjustment to have a large effect on the estimated treatment effect. However, because high school GPA and SAT scores are close to continuous, exact matching is unlikely. Therefore, I did chose to use the bias-adjusted one for all of my estimates.

Because there is no clear rule for choosing of the number of matches, the researcher would certainly be wise to compute the matching estimators using various numbers of matches, in order to check the estimates and variance for robustness (Abadie, Drukker, Herr, and Imbens 2001). In my empirical work, I estimate the ATT using various numbers of matches and generally find that although the ATT estimates are relatively stable, the standard errors usually decrease as the number of matches increases. This is most likely because there a many good matches.

Another factor to consider regarding the matches is how to define which individuals are close. In Section 2.4 I mentioned that any positive definite matrix V can be used in the norm which determines the distance between two individuals. In this study I use the default weighting matrix for both R and STATA, which is the diagonal matrix with the inverse of the covariate variances on the diagonal. Another common distance metric is the Mahalanobis metric, which uses the inverse of the covariance of the pretreatment variables (Rosenbaum and Rubin 1985). In Table A.3 I compare estimates using these two distance metrics.

3.2 Software Options for Matching

At present, there are two options for researchers who want to perform matching estimation with a relatively "canned" package: STATA and R. Both packages offer simple syntax and can be used to compute any one of the three estimands (ATE, ATT, or ATC) and have numerous options regarding the number of matches, bias-correction, heteroscedasticity correction, etc. STATA has the advantage of being much more well known, but the matching routine in R is significantly faster (easily ten times faster in my experience) and is not difficult to implement (Sekhon 2005).

Chapter 4

Empirical Strategy

In this study, I estimate the treatment effect of the HOPE scholarship for a variety of human capital related outcomes using matching estimators. Many of these effects were also estimated in Lee (2004) using simple difference-in-differences regressions.[1] While I do revisit some of the effects that KHL analyzed with a different estimator, I also extend the analysis and look at a number of previously unexamined aspects of the effects of HOPE.

4.1 Model

Because the matching estimator is a non-parametric procedure, the model for this study does not involve any particular functional form. However, the matching estimator involves making decisions about what variables are important controls. To match individuals, one must chose which covariates will be used in determining the matches.

The implicit model is that the outcomes of interest (number of course hours completed, for example) are functions of the matching covariates and the treatment variable:

$$Y_i = f\left(X_i, W_i\right) + \epsilon_i \tag{4.1}$$

Unless specified otherwise, X_i always contains SAT Verbal Score, SAT Math Score, Weighted High School Grade Point Average (HSGPA), AP Credit Hours, as well as race and gender dummies. In general, the treatment variable W_i is the interaction between GA, a dummy for Georgia residency, and $HOPE$, a dummy equal to 1 beginning in 1993 (i.e. when HOPE was implemented). Conditional on these covariates, the students are assumed to be on average equal so that the difference in matched students' outcomes are entirely attributable to HOPE.

[1] From this point forward, I will refer to Lee (2004) as KHL.

4.2 Data

I use data on UGA undergraduate students compiled by KHL from the Office of the Registrar, the Office of Admissions, and the Office of Student Financial Aid. The KHL dataset includes every UGA undergraduate who was enrolled between 1989 to 1997 and matriculated as typical first-time freshmen (TFTF).[2] However, I generally use only students from the 1990 and 1995 entering classes. The 1990 entering class was the last class to never have any interaction with HOPE, and the 1995 entering class was the first class to be "fully covered" by HOPE.[3] Therefore, using these two classes allows me to focus solely on the HOPE effect and clearly identify the treated individuals and the controls. Because the dataset ends in 1997, I cannot follow the 1995 class into their fourth year, so none of the estimates involving class-year distinctions include students in their fourth year. TFTF make up around 95% of the undergraduate student body (Lee 2004). The other 5% is comprised of transfers, etc. who interact with HOPE differently, so I exclude them from this study.

Table A.1 contains the means and standard deviations of the variables I use in my estimations. By looking at the high school GPA, SAT, and AP data, one can see that the quality of resident students clearly increased relative to non-resident students after the implementation of HOPE. For example, in 1990 the average resident TFTF had an SAT Verbal score of 557, while the average for non-residents was 574. By 1995, residents had totally closed this gap, and both groups had an average of 597. In terms of credit accumulation, however, residents did not see such improvements. In 1990 the two groups completed a full load (45 quarter hours per year) at similar rates: 62% for residents and 60% for non-residents. By 1995, the full-load completion rate for residents had fallen to 53% while the rate for non-residents actually rose to 61%.

[2]Typical first-time freshmen (TFTF) are students who matriculate in the fall term of the same year as they graduate from high school and have not enrolled in another post-secondary institution before entering the university (Lee 2004).

[3]In 1995 the income cap for HOPE was removed, so any Georgia resident student could qualify for HOPE, in principle.

With regards to summer enrollment, the percentage of residents enrolling in summer courses does not change much, but the mean number of hours taken in the summer by residents does increase from 2.9 to 3.19 from 1990 to 1995. The percentage of non-residents taking summer courses actually falls by 4 percentage points, and the mean number of summer hours completed by non-residents falls from 2.98 to 2.25 hours. Relative to non-residents, residents increased their participation in summer courses significantly. Interesting, the percentages of students transferring credits and the mean number of hours transferred fell for both resident and non-resident students from 1990 to 1995.

These data provide a good basis for matching. Because there are many "similar" students (where similar is defined by having similar covariate values), there are many good matches. As I mentioned in Section 2.7, when matching is not exact, the estimator will be biased. Although exact matching is unlikely in this situation, the matches are very good, so the precision of the estimates generally increases with the number of matches used. Therefore, the bias-adjustment does not change the estimates by much.

4.3 Weaknesses of Matching

While the matching estimator does free us from having to parameterize our model, this freedom does have a cost. First of all, if we do not take advantage of valid assumptions (for example, linearity), we cannot achieve efficiency. However, even ignoring the question of efficiency, the inherent nature of the matching estimator carries some limitations.

One limitation is the inability to easily calculate partial effects. When one computes a matching estimator, only *one* estimate is produced: the average treatment effect (be it the ATE, ATC, or ATT). The matching estimator controls for any number of covariates, but it does not give partial effects for any of the covariates. When KHL does difference-in-differences based on linear regression, she reports not only the HOPE effect but also the partial effects of race, gender, etc. I cannot do that, or at least not as simply. For example, with matching it is not possible to calculate the partial effect of gender, but I can divide the sample into males and females and calculate two different ATT's to determine the *difference* in the effects by gender. This is an important distinction to understand. Below I discuss the differences in the treatment effects using various partitions of the data.

Another is that matching estimators (and ATE estimators more generally) cannot be used for models with multiple outcomes. For example, one could not use matching to estimate the effect of HOPE on the choice of major, where there are more than 2 major choices. However, one could use matching to estimate the effects on the probability of students choosing a *particular* major or course.

Chapter 5

More Evidence on Student Responses to HOPE Retention Rules

The primary objective of this study is to estimate the effect of HOPE on the accumulation of credit hours as captured by course completion, course enrollment, and course withdrawal. I use variables for the extensive and intensive margins for each of these three outcomes. The extensive margin involves whether or not the action was completed, whereas the intensive margin involves the number of hours. For example, the extensive margin variable, "Full Load Completion," is a dummy indicating whether or not the student completed a "full load" (45 quarter hours), while "Credit Hours Completed" is the intensive margin variable. It is important to understand that "Credit Hours Completed" = "Credit Hours Enrolled" − "Credit Hours Withdrawn."

As I mentioned in the introduction, many proponents of merit-based scholarships claim that the merit component will induce students to work harder in order to qualify for and maintain their scholarships. For nearly all Georgia residents at UGA, qualifying for HOPE is a non-issue; admission to the university almost always requires a high school GPA above a 3.0. Since 1995 nearly all resident students have begun their careers at UGA as HOPE Scholars. For them, the issue is maintaining at least a 3.0 GPA in order to "keep HOPE."

There are several ways in which students can increase their probability of keeping HOPE. In addition to increasing studying and effort, students can change majors, course selection, or the number/timing of hours taken. In this chapter I focus on the latter.

5.1 Reassessing KHL

In this section I use matching estimators to estimate effects that KHL estimated using linear difference-in-differences.

5.1.1 Intertemporal Substitution or Delay

When we think about the effect that the HOPE retention rules may have on the accumulation of credit hours, the number of hours required to graduate is fixed and HOPE will not pay for more than the required number. That said, if HOPE influences the number of hours completed, then the effect must either be an intertemporal substitution of course completion or a delay.

For example, if a student completes fewer hours in her first year but compensates that loss in credit hours by completing more hours in subsequent years, she could still graduate on time. In this case, she has merely substituted her course taking across time. On the other hand, if she does not compensate for her lost hours by completing more hours after her first year, then she will delay her graduation.

Both KHL and I find strong evidence that students complete fewer hours in their first year due to HOPE. My results are the first column of Table A.2. I find that first-year students complete on average .887 fewer credit hours because of HOPE. This loss of credit hours is due to an approximately one hour increase in withdrawn credit hours. That is, students adjust by withdrawing from more courses rather than enrolling in fewer courses. KHL estimates a larger effect of 1.745 fewer hours completed, although this estimate is less precisely estimated than mine. KHL estimates the reduction of 1.745 hours when she only uses the 1990 and 1995 classes. When she uses the 1989-1997 classes, she finds a reduction of about one hour, closer to my estimates. That said, one would expect the estimates to be lower when including students from years when hope was only available to students whose family's income did not exceed the HOPE income cap. Later I will discuss my results when I estimate the effects for those classes.

When we look beyond the first year, my results diverge from KHL. In the second year, I find a small, statistically insignificant positive effect on credit hours completed, and in the third year the estimated effect is even smaller and less precisely estimated. KHL, on the other hand, estimates an increase of 1.732 credit hours completed in the second year, which virtually erases the loss of progress in the first year. In the second and third years, I estimate that students do enroll in more hours because of HOPE, but these gains in enrollment are counteracted equivalent increases in withdrawals. Whereas KHL estimates that students are merely moving around their course load, I estimate that they are actually delaying their progress towards graduation.

If a student chooses to complete fewer hours, why might the effect be much stronger in the first year of college? First, almost by definition, the variance of students' GPAs is much higher in the first year. They have taken fewer courses, so their GPAs will be more vulnerable to bad grades. Therefore, it is more costly to make bad grades as a first-year student.

Additionally, until recently, a student's GPA was only reevaluated for HOPE eligibility every 45 credit hours. It is important to note that for HOPE purposes, all "attempted hours" count towards the HOPE checkpoints.[1] Therefore, if a student enrolled in fewer than 45 hours during her first year, then she would be guaranteed HOPE during the fall of her second year, regardless of her GPA. Because all attempted hours count towards the HOPE checkpoint, one might assume that students would be more tempted to adjust by enrolling in fewer hours rather than withdrawing from courses.

My estimates, however, do not support this hypothesis. Course enrollment does not seem to be affected much, yet students withdraw from many more courses. Why might students choose to act in such a way? One idea is that students can more accurately influence their GPAs by withdrawing from courses rather than initially enrolling from fewer courses. Assume that a student wants to complete 12 hours in a quarter rather than 15 hours. If she enrolls in 15 hours, she then has until the midpoint of the semester to determine which course will

[1] "Attempted hours" include all enrolled hours, even courses from which the student withdraws.

likely yield the lowest grade and then withdraw from it. This strategy might be the most effective (and easiest) way to maximize one's probability of keeping HOPE.

Recently, in response to so many students "gaming the system," the rules for evaluating HOPE eligibility changed. Students' GPAs are no longer evaluated every 45 quarter hours (or the equivalent in semester hours), but they are evaluated every spring. Students can no longer delay evaluation of their HOPE eligibility by enrolling in fewer hours. Although I do not find a significant effect on enrollment, as on might expect due to the old rule, this is a step in the right direction. Policy makers would be wise to consider further reforms to minimize the negative responses to HOPE retention rules.

5.1.2 Effects Throughout High School GPA Distribution

Until this point I have been partitioning the students by their year in school. However, it seems logical that the effects of HOPE would vary throughout the GPA distribution. In particular, one would expect that students near or below the HOPE eligibility point would be most affected by the retention rules. Students with high GPAs are very unlikely to lose their scholarship, so why would they alter their behavior significantly in response to the retention rules?

To examine how the effects differ based on students' grades, I partitioned the students based on their weighted high school GPAs (HSGPA). I used three HSGPA ranges: <3.0, 3.0-3.5, and > 3.5. Table A.6 shows the effects throughout the GPA distribution for typical first year students. Somewhat surprisingly, first-year students with the highest HSGPAs ($>$ 3.5) actually reduced their number of credit hours completed by the most, with a reduction of -1.075 hours, on average. The effect for students with mid-range GPAs reduced their credit hour accumulation by 0.921 hours because of HOPE, while students with HSGPAs reduced their loads by the least, with a reduction of -0.559 hours. It certainly seems a bit odd that the best students, who are least likely to be in danger of losing their scholarships would show the strongest response to the HOPE retention rules.

If we extend the analysis to the second and third years, however, the estimates start to make more sense. Tables A.7 and A.8 show the HOPE effects by HSGPA range for second- and third-year students. Students with in the high HSGPA range actually increase their the number of credit hours they complete by .540 and .586 hours in their second and third years, respectively. Summing the effects for these students over the first three years of college yields a net effect of essentially zero. Therefore, it seems that students with high HSGPAs are merely engaging in intertemporal substitution. For students with mid-range HSGPAs, however, there is a small statistically insignificant effect in the second year followed by a large ($-.823$) but statistically insignificant effect in the third year. The effects for students with low HSGPAs in their second and third years are statistically insignificant, although the effect in the second year is quite large in magnitude (-1.150). Therefore, for students with both low and mid-range HSGPAs, the net effects over the first three years are strongly negative. Although the effects in the first year do not agree with the expectation that students with the highest HSGPAs would be least effected by HOPE, once we extend the analysis to the second and third years, the net effect over time does seem to support that idea.

5.1.3 Summer Course Taking

Another strategy to increase one's expected GPA and thus the probability of keeping HOPE is to enroll in courses with higher expected GPAs. Figure 9 of KHL shows that the average grades in summer classes are much higher than their normal academic year equivalents. Because HOPE places no restrictions on when students may take courses, an optimizing student might shift some credit hours to the summer, not only to decrease workload during the fall, winter, and spring, but also to take advantage of the higher average GPAs of summer courses.

Table A.9 shows the results of my estimation of this effect. Because the academic year begins in the fall quarter and ends in the following summer, no TFTF will have any summer credits from her first year. As I mentioned above, the dataset does not have observations for the fourth year of the 1995 entering class, so I could only measure the effects for second- and third-year students. For both years, I measured small but statistically significant increases in summer credit hours due to HOPE. Second-year students increased their summer credit hours by .210, while third-year students did so by .581 hours.

5.2 Specification and Estimation Options

5.2.1 Number of Matches

Since the literature on matching does not provide much guidance regarding the number of matches to use, I compare the estimation results using various numbers of matches. These comparisons can be seen in Table A.4. Generally, the HOPE effect estimates were relatively invariant to changes in the number of matches used, at least qualitatively. The magnitude of the effects tended to stabilize for $M > 4$, but the precision of the estimates usually continued to increase with M. The standard errors generally decreased with more matches but did not change much beyond $M = 12$, so I chose to use 12 matches for all of my estimates. For example, SATT for the percentage of first-year students completing a full load with $M = 1$ is -0.124 but for $M = 4, 8, 12, 16$ the SATT only ranged between -0.132 and -0.132. Similarly, the effect on full-load completion for second year students only ranged between 0.020 and 0.025 for M greater than or equal to 4. As I mentioned above, the high quality of matches is probably the reason why a relatively large number of matches seems to work well here.

5.2.2 Bias-Adjustment and Heteroscedasticity

Given that exact matching is highly unlikely for variables like high school GPAs and SAT scores, I used the bias-adjusted estimator discussed in Section 2.7. However, due to the quality of the matches, the bias adjustment did not affect the magnitude of the estimates by much. Table A.5 shows side-by-side comparisons of estimates that have been bias-adjusted with those that have not as well as standard errors computed under the assumption of homoscedasticity with those consistent under heteroscedasticity. For the most part, the bias-adjustment does not impact the size of the estimates significantly. For example, the unadjusted estimate of the SATT on credit hour completion for first year students is $-.867$, while the bias-adjusted estimate is $-.887$. The heteroscedasticity adjustment makes almost no difference for my standard errors. For the SATT estimate just mentioned, the unadjusted and the heteroscedasticity corrected standard errors are exactly the same (rounded to three decimal places). For the SATT on credit hours withdrawn, using the heteroscedasticity corrected standard error only changes the value of the standard error from .100 to .092, which is a very small change for an estimate of .992. Since the differences between the unadjusted or heteroscedasticity adjusted standard errors are very small, this margin does not affect the significance of any estimates. Because the heteroscedasticity consisted standard errors take a very long time to compute, I did not use them in any tables other than those mentioned above. In the I did use the bias-adjusted estimator in all of the other tables, though.

5.3 Transferred Credits

An additional way that students might seek to increase their chances of maintaining HOPE eligibility is through taking courses at institutions other than UGA where the courses may be easier, such as community colleges. Anecdotal evidence suggests that many students now take difficult classes such as calculus at a local community college in Athens or at other similar institutions in their hometowns.

My estimates, however, do not support the hypothesis that students transfer more credits because of HOPE. Table A.10 shows the estimates of the HOPE effect on transferred credits. The only large and significant effect I find is in the third year, with third year students transferring an average of .366 more credit hours because of HOPE.

5.4 Effects by Race and Gender

KHL finds that the HOPE retention rate at UGA varies greatly by race and gender. At the first HOPE checkpoint, 63.85% of white students keep remain eligible for hope whereas 38.78 % of black students do. The gender difference is not as great, with 63.57 % of females keeping their scholarships compared to 57.69% of males. In Tables A.11 and A.12 I show the HOPE effects by race and gender for first-year students. Interestingly, the negative effect on credit hours completed is stronger for whites and for females. This suggests that part of the success of whites and females in maintaining there HOPE scholarships may be due to their effective response to the retention rules. I am not claiming that this is necessarily the case, but these estimates do suggest that it is a possibility.

5.5 Effects of Removing the Income Cap

The family income cap for HOPE recipients was set at $66,000 in 1993, raised to $100,000 in 1994, and eliminated altogether in 1995. Previously I used only the 1990 and 1995 classes so that I would have a treated population with full coverage and a control population with no HOPE. I did this to avoid having the estimates influenced by the effects of removing the income cap. However, it is certainly of interest to see how students coming from different socioeconomic backgrounds respond to the HOPE incentives.

To estimate these differences, I also estimated the HOPE effects using 1993 and 1994 students. In these estimations, the treatment variable is the interaction of the dummy for the Georgia residency with a dummy for the students' class. For example, I estimated the effects of HOPE for the 1993 entering class, using the resident 1993 first-year students as treated individuals and nonresidents from 1993 along with all first-year students from 1990. I estimated the effects for the 1994 class similarly, as well. One should note that defining the treatment variable as I did actually labels some students as treated when they were actually not eligible for HOPE. For example, only a fraction of resident students were eligible for

HOPE in 1993 and 1994, but my treatment variable labels all of those students as being treated.

Table A.13 shows these estimates. The HOPE effects on credit hours complete for 1993, 1994, and 1995 were $-.462$, $-.851$, and $-.887$, respectively.As one would expect, the HOPE effects are weakest in 1993 when only 35% of resident first-year students were HOPE scholars and strongest in 1995 when nearly all (95% of resident first-year students received HOPE. However, the magnitude of these HOPE effects grows less than proportionally with the expansion of HOPE. For example, from 1994 to 1995 around 25% more students received HOPE, but the effect on credit hours completed only increased in magnitude by 4%. This implies that students from wealthier backgrounds might not be effected as much by the HOPE retention rules.

One way to give more meaning to these estimates is to divide them by the percentage of resident students actually receiving HOPE in from their respective years. Although the percentage of HOPE scholars is a crude measure of the portion of students who were eligible for hope under the income cap, it is the best I can do given the data. For example if I had estimated an effect of $-.5$ but only 50% of the resident students were eligible for HOPE due to the income cap, I might think of the effect as being close to -1. Dividing by the percentage of HOPE scholars yields -1.31 for 1993 and -1.12 for 1994. The 1995 estimate does not need this "adjustment," since there was no income cap in that year. My crude adjustment does give some idea of the difference in the responses to HOPE as the income cap was phased out. The fact that this "adjusted" estimate decreases as the income cap was phased out does support the idea that students from wealthier backgrounds might not be effected as much by the HOPE retention rules.

Chapter 6

Conclusions

In this paper I used non-parametric matching estimators both to re-examine the effects of the HOPE scholarship on a number of outcomes related to students' accumulation of credit hours examined in Lee (2004) and Cornwell, Lee, and Mustard (2005) and to examine some outcomes not previously studied. Because they are relatively unknown, I provided a review of matching estimators which provided the basic framework for matching estimation and discussion various issues and options related to the use of matching techniques, including number/quality of matches, bias correction and heteroscedasticity. In my empirical project, I used a variety of different specifications to test the robustness of the model and did not find the number of matches, bias adjustment, or heteroscedasticity correction to significantly effect the estimates.

The data on UGA undergraduate students provided a good basis for using matching techniques, and I used matching estimators to estimate some of the effects of HOPE estimated in Lee (2004). Generally, the results I obtained from matching were similar to Lee's results, but I found evidence that students reduce their credit hour accumulation in the first-year of college by about one hour but do not make up for this reduction by increasing credit hours in subsequent years. Lee's study found evidence that students might merely be engaging in intertemporal substitution by reducing their loads as first-year students but later increasing them to prevent delaying time to graduation.

I found the effect of HOPE to differ throughout the high school GPA distribution. While students in all high school GPA ranges significantly reduce the number of hours the complete in their first year of college, there is evidence that students with higher high school GPAs make up for this reduction by increasing their loads during the second and third years. students with mid-range and low high school GPAs, on the other hand, do not increase their course loads later in college and thus delay their progress to graduation. The final area of Lee's study which I re-examined was the HOPE effect on summer course enrollment. Again, consistent with Lee, I found an increase in summer course enrollment attributable to HOPE.

I did not merely repeat Lee's study with a different estimator, however. I extended the analysis to examine the effect of HOPE on the transferring of credits to UGA. For the most part, I did not find significant evidence that HOPE affected the amount of credit transferring by UGA students, although I estimated that HOPE led to an increase of .366 transferred credits by the average third-year student. I also partitioned the dataset to see if the HOPE effects differed significantly by race and gender. While the effects were all in the same direction, I found that the effects in the first year were greater in magnitude for women and whites than for men and blacks.

Finally, I examined the how the HOPE effects differed while the HOPE income cap was sequential raised and then eliminated. While I found stronger HOPE effects as more students received HOPE due to the the removal of the income cap, the effects increased less than proportionally, suggesting that students from wealthier backgrounds might not be effected as strongly by the HOPE retention rules.

Although these results vary in some important ways from previous studies of this topic, the key message is the same. We should be critical of claims that merit-aid programs are an effective way to increase human capital. At a very minimum, we should also seek to make good retention and eligibility rules that minimize any of the potential negative behavior induced by merit-aid programs. Recent changes in HOPE retention rules have started a small movement in this direction.

Bibliography

Abadie, A., D. Drukker, H. Herr, and G. W. Imbens (2001): "Implementing Matching Estimators for Average Treatment Effects in STATA," *The Stata Journal*, 1(1), 1–18.

Abadie, A., and G. W. Imbens (2002): "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects," NBER Working Paper No. T0283.

——— (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74(1), 235–267.

Angrist, J., and A. Krueger (2000): "Empirical Strategies in Labor Economics," in A. Ashenfelter and D. Card (eds.) Handbook of Labor Economics, vol. 3. New York: Elsevier Science.

Ashenfelter, O., and D. Card (1985): "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *Review of Economics and Statistics*, 67, 648–660.

Barnow, B., G. Cain, and A. Golberger (1980): "Issues in the Analysis of Selectivity Bias," in E. Stromsdorfer and G. Farkas (eds). Evaluation Studes, vol 5. San Francisco: Sage.

Becker, S., and A. Ichino (2002): "Estimation of Average Treatment Effects Based on Propensity Scores," *The Stata Journal*, 2(4), 358–377.

Blundell, R., and M. Costa Dias (2002): "Alternative Approaches to Evaluation in Empirical Microeconomics," Institute for Fiscal Studies, Cemmap working paper cqp10/02.

Card, D., and Sullivan (1988): "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment," *Econometrica*, 56(3), 497–530.

Cornwell, C., K. Lee, and D. Mustard (2005): "Student Responses to Merit Scholarship Rules," *Journal of Human Resources*, 40, 895–917.

Cornwell, C., M. Leidner, and D. Mustard (2003): "Rules, Incentives, and Policy Implications of Large-Scale Merit-Based Financial Aid Programs," Univesrity of Georgia Department of Economics Working Paper, Athens, GA.

Cornwell, C., D. Mustard, and D. Sridhar (2005): "The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's HOPE Scholarship," *Journal of Labor Economics*, forthcoming.

Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66(2), 315–331.

Heckman, J., H. Ichimura, and P. Todd (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Econoimc Studies*, 64, 605–654.

Heckman, J., R. Lalonde, and J. Smith (2000): "The Economics and Econometrics of Active Labor Markets Programs," in A. Ashenfelter and D. Card eds. Handbook of Labor Economics, vol. 3. New York: Elsevier Science.

Heckman, J., and R. Robb (1984): "Alternative Methods for Evaluating the Impact of Intervenctions," in Heckman and Singer (eds.), Longitudinal Analysis of Labor Market Data, Cambridge: Cambridge University Press.

Ichino, A., F. Mealli, and T. Nannicini (2005): "Sensitivity of Matching Estimators to Unconfoundedness: An Application to the Effect of Temporary Work on Future Employment," Working Paper.

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

Lalonde, R. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76, 604–620.

Lechner, M. (1999): "Earnings and Employment Effects of Continuous Off-the-job Training in East Germany After Unification," *Journal of Business and Economics Statistics*, 17(1), 74–90.

Lee, K. H. (2004): "The Effects of Merit-Based Financial Aid on Academic Choices in College: Evidence from Georgia's HOPE Scholarship Program," Ph.D. thesis, University of Georgia.

Neyman, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," translated in Statistical Science (with Discussion), Vol 5, No 4, 465-480, 1990.

Rosenbaum, P., and D. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Casual Effects," *Biometrika*, 70, 41–55.

——— (1985): "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *American Statistician*, 39, 33–38.

Rubin, D. (1977): "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Education Statistics*, 2(1), 1–26.

Sekhon, J. (2005): Multivariate and Propensity Score Matching Estimator for Causal Inference. http://sekhon.berkeley.edu/matching/Match.html.

Shifrin, T. (2004): *Multivariable Mathematics : Linear Algebra, Multivariable Calculus, and Manifolds*. New York: John Wiley & Sons.

Smith, J., and P. Todd (2001): "Reconciling Conflicting Evidence on the Performance of Matching Estimators," *American Economic Review - Papers and Proceedings.*

Smith, J., and P. Todd (2005): "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?," *Journal of Econometrics*, 125(1-2), 305–353.

Zhao, Z. (2004): "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence," *The Review of Economics and Statistics*, 86(1), 91–107.

Appendix A

Tables

Table A.1: Sample Means and Percentages
Typical Students, 1990/1995 Classes
(Standard Deviations in Parentheses)

	Pre-HOPE (1990)		Post-Hope (1995)	
	Non-Residents	Residents	Non-Residents	Residents
Full Load Completion[a]	0.60	0.62	0.61	0.53
Credit Hours Completed	41.59	42.79	42.86	42.20
	(8.14)	(6.81)	(6.22)	(6.37)
Full Load Enrollment[b]	0.80	0.80	0.84	0.80
Credit Hours Enrolled	43.37	44.31	44.91	44.62
	(7.28)	(6.09)	(5.31)	(5.15)
Course Withdrawal[c]	0.28	0.26	0.35	0.40
Credit Hours Withdrawn	1.78	1.52	2.05	2.42
	(8.14)	(6.81)	(6.22)	(6.37)
High School GPA[d]	2.95	3.14	3.23	3.45
	(0.45)	(0.52)	(0.42)	(0.41)
SAT Verbal Score	574.52	557.34	596.85	596.74
	(70.53)	(76.50)	(68.71)	(68.30)
SAT Math Score	561.08	550.90	590.97	587.13
	(65.81)	(68.49)	(62.99)	(68.58)
AP Credit Hours Earned	2.57	2.25	4.67	4.47
	(5.52)	(5.19)	(7.47)	(7.66)
Summer Course Enrollment[e]	0.21	0.22	0.18	0.23
Summer Credit Hours	1.85	1.83	1.41	2.04
	(3.89)	(3.89)	(3.89)	(3.89)
Transferred Credits[f]	0.23	0.22	0.19	0.16
Transfer Credit Hours	2.50	2.39	2.01	2.17
	(6.46)	(6.46)	(6.46)	(6.46)

[a] Portion of typical first-year students completing a full load (> 45 quarter hours)
[b] Portion of typical first-year students enrolling in a full load (> 45 quarter hours)
[c] Portion of typical first-year students withdrawing from at least one course
[d] HSGPA is weighted high school GPA.
[e] Portion of typical second- and third-year students enrolling in summer courses
[f] Portion of typical students transferring at least one course in a year

Table A.2: Estimated HOPE Effect on
Intertemporal Subsitution of Course Enrollment
Typical Students, 1990/1995 Classes
(Robust Standard Errors in Parentheses)

	First Year	Second Year	Third Year
Full Load Completion	−0.131	−0.051	−0.017
	(0.014)	(0.015)	(0.015)
Credit Hours Completed	−0.887	0.174	−0.016
	(0.186)	(0.295)	(0.346)
Full Load Enrollment	−0.024	0.025	0.022
	(0.011)	(0.013)	(0.013)
Credit Hours Enrolled	0.105	1.054	0.884
	(0.159)	(0.273)	(0.324)
Course Withdrawal	0.145	0.121	0.082
	(0.013)	(0.014)	(0.015)
Credit Hours Withdrawn	0.992	0.880	0.901
	(0.092)	(0.117)	(0.135)
No. of Obs.	6918	6225	5706
No. of Treated Obs.	3110	2837	2659

Table A.3: Estimated HOPE Effect on Credit Accumulation
Typical Students, 1990/1995 Classes
(Robust Standard Errors in Parentheses)
Comparison of Matching Metrics

	Inverse Variance[a]	Mahalanobis[b]
Full Load Completion	**−0.131**	−0.132
	(0.014)	(0.014)
Credit Hours Completed	**−0.887**	−0.942
	(0.186)	(0.185)
Full Load Enrollment	**−0.024**	−0.025
	(0.011)	(0.011)
Credit Hours Enrolled	**0.105**	0.060
	(0.159)	(0.159)
Course Withdrawal	**0.145**	0.147
	(0.013)	(0.013)
Credit Hours Withdrawn	**0.992**	1.002
	(0.092)	(0.092)
No. of Obs.	**6918**	6918
No. of Treated Obs.	**3110**	3110

[a] Matches based on distance using the default weighting matrix
[b] Matches based on the Mahalanobis metric

Table A.4: Sensitivity of the Matching Estimator to the Number of Matches Used
Typical First-Year Students, 1990/1995 Classes
(Standard Errors in Parentheses)

No. of Matches[a]	1	4	8	**12**	16
Full Load Completion	−0.124	−0.132	−0.131	**−0.131**	−0.132
	(0.017)	(0.015)	(0.014)	**(0.014)**	(0.014)
Credit Hours Completed	−0.851	−0.862	−0.878	**−0.887**	−0.933
	(0.225)	(0.196)	(0.189)	**(0.186)**	(0.184)
Full Load Enrollment	−0.016	−0.023	−0.024	**−0.024**	−0.025
	(0.014)	(0.012)	(0.011)	**(0.011)**	(0.011)
Credit Hours Enrolled	0.100	0.140	0.113	**0.105**	0.066
	(0.192)	(0.165)	(0.159)	**(0.156)**	(0.154)
Course Withdrawal	0.136	0.145	0.145	**0.145**	0.146
	(0.016)	(0.014)	(0.014)	**(0.014)**	(0.013)
Credit Hours Withdrawn	0.951	1.003	0.992	**0.992**	0.999
	(0.117)	(0.103)	(0.101)	**(0.099)**	(0.099)
No. of Obs.	6918	6918	6918	**6918**	6918
No. of Treated Obs.	3110	3110	3110	**3110**	3110

[a] Estimates in all other tables use 12 matches.

Table A.5: Sensitivity of Matching Estimators to Bias-Adjustment and Heteroscedasticity
Typical First-Year Students, 1990/1995 Classes
(Standard Errors in Parentheses)

	Non-Adjusted[a]	**Adjusted**[b]
Full Load Completion	−0.125	**−0.131**
	(0.014)	**(0.014)**
Credit Hours Completed	−0.867	**−0.887**
	(0.186)	**(0.186)**
Full Load Enrollment	−0.021	**−0.024**
	(0.011)	**(0.011)**
Credit Hours Enrolled	0.101	**0.105**
	(0.156)	**(0.159)**
Course Withdrawal	0.139	**0.145**
	(0.014)	**(0.013)**
Credit Hours Withdrawn	0.968	**0.992**
	(0.100)	**(0.092)**
No. of Obs.	6918	**6918**
No. of Treated Obs.	3110	**3110**

[a] Estimates not adjusted for bias and std. errors not adjusted for heteroscedasticity

[b] Estimates adjusted for bias and std. errors adjusted for heteroscedasticity.

Table A.6: Estimated HOPE Effect on
Course Enrollment, Completion, and Withdrawal
by Weighted High School GPA Category
Typical First-Year Students, 1990/1995 Classes
(Robust Standard Errors in Parentheses)

	< 3.0	3.0-3.5	> 3.5
Full Load Completion	-0.139	-0.170	-0.100
	(0.032)	(0.023)	(0.020)
Credit Hours Completed	-0.559	-0.921	-1.075
	(0.444)	(0.312)	(0.262)
Full Load Enrollment	-0.050	-0.050	-0.003
	(0.029)	(0.019)	(0.015)
Credit Hours Enrolled	0.593	0.378	-0.366
	(0.383)	(0.263)	(0.224)
Course Withdrawal	0.150	0.192	0.107
	(0.032)	(0.023)	(0.019)
Credit Hours Withdrawn	1.152	1.299	0.709
	(0.242)	(0.161)	(0.125)
No. of Obs.	2062	2294	2562
No. of Treated Obs.	479	1072	1559

Table A.7: Estimated HOPE Effect on
Course Enrollment, Completion, and Withdrawal
by Weighted High School GPA Category
Typical Second-Year Students, 1990/1995 Classes
(Robust Standard Errors in Parentheses)

	< 3.0	3.0-3.5	> 3.5
Full Load Completion	−0.102	−0.065	−0.040
	(0.034)	(0.025)	(0.022)
Credit Hours Completed	−1.150	0.129	0.540
	(0.810)	(0.498)	(0.399)
Full Load Enrollment	−0.034	0.012	0.042
	(0.032)	(0.022)	(0.019)
Credit Hours Enrolled	−0.402	0.876	1.571
	(0.778)	(0.447)	(0.370)
Course Withdrawal	0.091	0.099	0.149
	(0.034)	(0.025)	(0.021)
Credit Hours Withdrawn	0.748	0.747	1.031
	(0.317)	(0.211)	(0.156)
No. of Obs.	1803	2052	2370
No. of Treated Obs.	418	961	1458

Table A.8: Estimated HOPE Effect on
Course Enrollment, Completion, and Withdrawal
by Weighted High School GPA Category
Typical Third-Year Students, 1990/1995 Classes
(Robust Standard Errors in Parentheses)

	< 3.0	3.0-3.5	> 3.5
Full Load Completion	−0.002	−0.038	0.004
	(0.036)	(0.025)	(0.022)
Credit Hours Completed	0.220	−0.823	0.586
	(0.989)	(0.579)	(0.473)
Full Load Enrollment	0.038	0.004	0.040
	(0.033)	(0.022)	(0.019)
Credit Hours Enrolled	1.467	−0.169	1.519
	(0.912)	(0.538)	(0.446)
Course Withdrawal	0.126	0.043	0.104
	(0.036)	(0.026)	(0.021)
Credit Hours Withdrawn	1.247	0.654	0.933
	(0.407)	(0.236)	(0.174)
No. of Obs.	1587	1885	2234
No. of Treated Obs.	373	896	1390

Table A.9: Estimated HOPE Effect on
Summer School Taking by Year in School
Typical Students, 1990/1995 Classes
(Robust Standard Errors in Parentheses)

	Second Year	Third Year
Summer Course Enrollment	0.007	0.044
	(0.013)	(0.016)
Summer Credit Hours	0.210	0.581
	(0.130)	(0.166)
No. of Obs.	6225	5706
No. of Treated Obs.	2837	2659

Table A.10: Estimated HOPE Effect on
Credit Transfers by Year in School
Typical Students, 1990/1995 Classes
(Robust Standard Errors in Parentheses)

	First Year	Second Year	Third Year
Transferred Credits (Year)	−0.001	−0.032	−0.034
	(0.003)	(0.012)	(0.012)
Transfer Credit Hours	−0.092	−0.146	0.366
	(0.058)	(0.201)	(0.203)
Transferred Credits (Cumulative)	−0.001	−0.033	−0.052
	(0.003)	(0.012)	(0.015)
Cumulative Transfer Credits Hours	−0.101	−0.502	−0.589
	(0.057)	(0.202)	(0.286)
No. of Obs.	6918	6225	5723
No. of Treated Obs.	3110	2837	2674

Table A.11: Estimated HOPE Effect on Course Enrollment
for Typical First-Year Students, 1990/1995 Classes
(Robust Standard Errors in Parentheses)

	White	Black
Full Load Completion	−0.132	−0.112
	(0.014)	(0.044)
Credit Hours Completed	−0.937	−0.637
	(0.196)	(0.524)
Full Load Enrollment	−0.012	−0.089
	(0.012)	(0.037)
Credit Hours Enrolled	0.071	0.252
	(0.170)	(0.440)
Course Withdrawal	0.153	0.094
	(0.013)	(0.043)
Credit Hours Withdrawn	1.008	0.889
	(0.097)	(0.267)
No. of Obs.	6226	692
No. of Treated Obs.	2705	405

Table A.12: Estimated HOPE Effect on Course Enrollment by Gender for Typical First-Year Students, 1990/1995 Classes (Robust Standard Errors in Parentheses)

	Males	Females
Full Load Completion	−0.127	−0.133
	(0.021)	(0.018)
Credit Hours Completed	−0.733	−0.976
	(0.295)	(0.240)
Full Load Enrollment	−0.001	−0.040
	(0.018)	(0.014)
Credit Hours Enrolled	0.402	−0.071
	(0.256)	(0.205)
Course Withdrawal	0.159	0.138
	(0.020)	(0.017)
Credit Hours Withdrawn	1.134	0.904
	(0.141)	(0.121)
No. of Obs.	2949	3969
No. of Treated Obs.	1295	1815

Table A.13: Estimated HOPE Effect on Credit Accumulation
With and Without Income Cap
Typical First-Year Students, 1990/1993/1994/1995 Classes
(Robust Standard Errors in Parentheses)

	1990/1993	1990/1994	1990/1995
Full Load Completion	−0.062	−0.112	−0.131
	(0.013)	(0.013)	(0.014)
Credit Hours Completed	−0.462	−0.851	−0.887
	(0.192)	(0.198)	(0.186)
Full Load Enrollment	−0.002	−0.004	−0.024
	(0.011)	(0.011)	(0.011)
Credit Hours Enrolled	0.061	0.125	0.105
	(0.167)	(0.172)	(0.159)
Course Withdrawal	0.083	0.128	0.145
	(0.012)	(0.013)	(0.013)
Credit Hours Withdrawn	0.523	0.976	0.992
	(0.091)	(0.097)	(0.092)
No. of Obs.	6523	6789	6918
No. of Treated Obs.	2668	2995	3110