THREE APPLICATIONS OF PROPENSITY SCORE MATCHING IN MICROECONOMICS AND CORPORATE FINANCE: US INTERNAL MIGRATION; SEASONED EQUITY OFFERINGS; ATTRITION IN A RANDOMIZED EXPERIMENT

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Xianghong Li, M.A.

The Ohio State University 2004

Dissertation Committee: Professor John C. Ham, co-adviser Professor René Stulz, co-adviser Professor Patricia B. Reagan Approved by

Co-Adviser

Co-Adviser Department of Economics

ABSTRACT

Propensity score matching is becoming an increasingly popular tool in economics for dealing with selection issues. Recently economists have been evaluating matching methods by trying to mimic the experimental evidence in social experiments where individuals were randomly assigned to job training. There is mixed evidence on whether matching can mimic the experiment results. This dissertation applies the propensity score matching on two non-experimental settings, US internal migration and seasoned equity offerings, and one social experiment outside of job training, the RAND Health Insurance Experiment. These applications generate new empirical evidence on these important issues and also shed light on the practical strengths and weaknesses of matching methods.

The first essay evaluates the effect of migration on wage growth. Using a distance-based measure of migration, the study finds a significant positive effect for college graduates and a marginally significant negative effect for high school dropouts. The study does not find any significant effect for other educational groups or for the overall sample. The results suggest that the better measure of migration matters.

There is mixed evidence on long-run stock performance after Seasoned Equity Offerings (SEOs). Some studies have found that firms issuing SEOs show substantial long-run underperformance, which poses one of the strongest challenges to the Efficient Market Hypothesis. The second essay shows that the long-run abnormal returns after SEOs found in previous studies are an artifact of the inability of traditional matching techniques to control properly for all risk factors. Propensity score matching allows matching on multiple firm characteristics simultaneously. This essay finds that SEO firms do not behave differently from similar firms that are small in size, have low book-to-market ratios and high recent returns.

Attrition has long attracted the attention of econometricians. The third essay uses data from the RAND Health Experiment to explore using matching to address attrition. Specifically, I introduce attrition into the experimental data and then use matching to deal with the problem, creating a new way of using post-baseline data in the matching procedure. I find that matching performs reasonably well, but its performance does depend on the treatment effect estimated. Dedicated to my parents, Tingying and Wenbing, for their support, advice and encouragement. They are the ones who first showed me the beauty of math.

To my husband, Jian, for being there for me through the most difficult and challenging years of my life.

ACKNOWLEDGMENTS

My true interest in economics research began after I started to work on my first project with John Ham and Patricia Reagan. I wish to thank them for introducing me to the wonderland of economics and for seeing me through this unpredictable, frustrating and ultimately rewarding process. I owe my adviser, John Ham, most of my intuition and knowledge about econometrics, and I am grateful for his unwavering support. I thank René Stulz for his guidance in finance research. I thank Randy Olsen for all his help in key moments during my Ph.D study. I am grateful to Audrey Light, Bruce Weinberg, Hajime Miyazaki, Paul Evans, James Peck, and Lixin Ye for always being there when I needed advice. I thank Steve Cosslett and Lung-fei Lee for their help in econometrics.

I also thank Xinlei Zhao for productive collaborations and Curtis Eiberwein for insightful discussions; Traci Mach for her help and support, especially during the early stage of my research; Alfonso Flores-Lagunes, Stefan Krause, and Róisín O'Sullivan for allowing me to share their experience when I was facing difficult decisions; Daching Yang, Yinglei Zhan, Yingshi Jing, Qing Liu, Tingting Ji, Xibao Li for discussions about my research, but more importantly, for their true friendship and moral support; Jo Ducey and John David Slaughter for always being ready to help; and Yong Yu and Eileen Kopchik for their help on computational issues.

VITA

June, 1966	Born, Chongqing, P. R. China
July, 1988	B.A. Economics Beijing University, Beijing, P. R. China
1988 - 1995	Researcher The Economic Reform Research Institute of China Beijing, P. R. China
1995 - 1997	Research Specialist McKinsey & Company Beijing, P. R. China
September, 1998	M.A. Economics The Ohio State University

FIELD OF STUDY

Major Field: Economics

TABLE OF CONTENTS

Abstractii
Dedicationiv
Acknowledgmentsv
Vitavi
List of Tablesx
Chapters:
1. Propensity Score Matching1
1.1 Counterfactual Framework in Evaluation Studies
1.2 Identification Conditions of Matching Methods
1.3 Propensity Score Matching
1.4 Identification Strategy of Matching
1.5 Choice of Matching Method
1.6 Choice of the Bandwidth Parameter
1.7 Common Support Constraint and Balancing Tests
2. Matching and Selection Estimates of the Effect of Migration Migration on Wages
for Young Men14
2.1 Introduction
2.2 Migration Literature Review

2.3 A Theoretical Model of Migration	20
2.4 Data Description	24
2.5 Sample Selection Models	31
2.5.1 Basic Model	31
2.5.2 Different Measues of the Effect of Migration on Wages	35
2.5.3 The Effect of Endogenous Explanatory Variables in the Probit Equation	36
2.5.4 Empirical Results from the Selection Model	37
2.6 Propensity Score Matching	46
2.6.1 Conditioning Variables and Finer Balancing	46
2.6.2 Propensity Score Models	49
2.6.3 Balancing Tests and a Specification Test	52
2.6.4 Estimates of the Migration Effects from the Baseline Model	55
2.6.5 Robustness of the Propensity Score Specification	59
2.6.6 Alternative Definitions of Migration	60
3. Are There Abnormal Returns After Seasoned Equity Offerings	69
3.1 Introduction	69
3.2 Sample Selection	73
3.3 Traditional Matching Methods	76
3.4 Propensity Score Matching	81
3.5 Illustration of the Propensity Score Matching Method Using 1993 Data	82
3.6 Results Over the Period 1986-1998	88
3.7 Conclusion	92

4.	Assessing the Ability of Matching to Address Attrition Bias Using Data from the	
	RAND Health Experiment	95

4.1 Introduction	
4.2 Rand Health Insurance Experiment	
4.2.1 Overview	
4.2.2 Sample Description and the Design of the Experiment	
4.3 Methodology	
4.3.1 A Simulation Procedure	
4.3.2 Model Specification and Matching Procedure	
4.4 Empirical Results	
4.1 Conclusion	
Appendix A: Common Support	115
Appendix B: Three-step Method for Choosing the Number of	
Bootstrap Repetitions	117

Appendix C: Other Variables Investigated in Chapter 3......119

Bibliography.....121

LIST OF TABLES

Table		Page
2.1	Sample Selection Criteria and Resulting Sample Size	27
2.2	Variable Definitions and Descriptive Statistics	29
2.3	Maximum Likelihood Estimates of the Mover-Stayer Model of Wages	
	on the Second Job - Unrestricted Model	41
2.4	Maximum Likelihood Estimates of the Mover-Stayer Model of Wages	
	on the Second Job – Restricted Models	43
2.5	Migration Effects from Restricted Selection Models	46
2.6	Propensity Score Coefficient Estimates	52
2.7	Balancing Tests (Nearest Neighbor Matching)	54
2.8	Specification Tests: "Effect" of Migration on Wage Growth on Job 1 by	
	Education Group	55
2.9	Matching Estimates of the Effect of Migration on Wage Growth from	
	Model I	58
2.10	Matching Estimates of the Effect of Migration on Wage Growth Based	
	on Three Alternative Models	62
2.11	Matching Estimates of the Effect of Migration on Wage Growth Based	
	on Three Definitions of Migration	63
2.12	Matching Estimates of the Effect of Migration on Wage Growth Based	
	on Three Definitions of Migration	66
3.1	SEO Samples	74
3.2	Ex post 36-Month Returns for SEO and Matching Firms	78
3.3	Difference in <i>ex ante</i> Variables from Traditional Matching	80
3.4	Propensity Score Model Coefficient Estimates	84
3.5	Balancing Tests of Propensity Score Matching: Difference in ex ante	
	Variables (1993 Sample)	85
3.6	Balancing Tests of Propensity Score Matching - Difference in ex ante	
	Variables (1986 - 1998)	90
3.7	Difference in 3-Year Returns (1986 - 1998)	91
4.1	Sample Selection Criteria and Resulting Sample Size	100
4.2	Variable Definitions and Descriptive Statistics	106
4.3	Simulating Attrition in the 95% Coinsurance Plan Group (Scenario I)	109
4.4	Simulating Attrition in the 95% Coinsurance Plan Group (Scenario II)	

CHAPTER 1

PROPENSITY SCORE MATCHING

Propensity score matching was first developed by two statisticians, Rosenbaum and Rubin, in 1983. The counterfactual framework, which is the basis for propensity score matching and all matching methods in general, was pioneered by Rubin (1974); its applications in medical treatment evaluation can be traced back to the 1970s. Extensive adoption of this method has been developed in economics since the early 1990s, and most matching studies in economics evaluate effects of job training.¹ This dissertation presents three applications of propensity score matching in microeconomics and corporate finance. This chapter discusses the general econometric theory behind the propensity score matching method and technical issues related to the three empirical studies in this dissertation.

1.1 Counterfactual Framework in Evaluation Studies

Propensity score matching is usually discussed under the counterfactual framework, introduced in this section. In a typical evaluation study, we observe a binary

¹ The empirical literature on matching as an alternative means of addressing selection is too extensive to review here. Some of the earlier papers are Dehejia and Wahba (1999), Heckman, Ichimura, and Todd (1997) and Lechner (1999, 2000). These studies all consider matching in the context of job training.

variable² representing treatment status. To assess the effect of a treatment, we need to know outcomes both with and without treatment for each individual in the sample. The treatment effect is the difference between the two outcomes. However, since each individual is either treated or not treated, we only observe one outcome and the other is a piece of missing data. When there is a randomized treatment assignment, the treatment effect may be estimated by comparing the results for the treated and untreated individuals directly, so the missing data problem is solved by randomization. In reality, most treatments involve voluntary participation, so the treated individuals may be systematically different from those remaining untreated, and such direct comparison may be misleading. The counterfactual framework, pioneered by Rubin (1974) and since adopted by many in both statistics and econometrics, is aimed at solving the missing data problem in a nonrandomized setting.

We first define the terminologies and notations used in this chapter. Let D = 1 if an observation is treated (in the treatment group), and D = 0 if an observation is untreated (in the comparison group). Let Y_1 be the outcome with treatment and Y_0 be the outcome without treatment. The **average treatment effect on the treated**³, which we denote **ATET** hereafter, is:

$$\Delta = E(Y_1 | D = 1) - E(Y_0 | D = 1)$$
(1.1.1)

² There are on-going development in the literature addressing multiple treatment and dynamic treatment. ³ In principle, the counterfactual framework is able to estimate both average treatment effect (ATE), the expected effect of treatment on a randomly drawn person from the population, and average treatment effect on the treated (ATET), the mean effect for those who actually participated in the program. Since all matching estimates in this dissertation are average treatment effect (ATE) can be formulized as: $\Delta = E(Y_1) - E(Y_0)$

For the treatment group, only $(Y_1 | D = 1)$, results with treatment, are observed; $(Y_0 | D = 1)$, results had they received no treatment, are not observed. The major task of evaluation studies is to construct various counterfactuals, and in this case the counterfactual under construction is $(Y_0 | D = 1)$

1.2 Identification Conditions of Matching Methods

As discussed above, directly inferring ATET by $E(Y_1 | D = 1) - E(Y_0 | D = 0)$ in a nonrandomized setting could be misleading, because the treatment group and the comparison group may be (different) non-random samples. If factors that affect the treatment participation decision also affect the outcome, using $E(Y_0 | D = 0)$ as a substitute for $E(Y_0 | D = 1)$ will introduce systematic bias. This is a typical sample selection problem also common in economic studies.

To solve the selection problem, matching methods impose the following conditional independence assumption (CIA):

$$(Y_0, Y_1) \perp D \| X$$
 (1.2.1)

where X is a vector of *ex ante* variables.

This condition is also called *ignorable treatment assignment* (Rosenbaum and Rubin 1983) or *selection on observables* (Heckman and Robb 1985). Equation (1.2.1) says that, conditional on X, had the treatment and comparison groups both chosen (or not chosen) the treatment, their outcomes would follow the same distribution. This condition does not mean that a treatment group and a comparison group with the same *ex ante* X will have the same *ex post* outcome. It rather says that, had the participants not chosen the treatment, their outcomes would have followed the same distribution as those

of non-participants with the same *ex ante* X, and vice versa for the nonparticipants. The idea underlying CIA is this: if we have enough information in X about the participation decision, we can eliminate the correlation between (Y_0, Y_1) and D by conditioning on X. In Section 1.4, we will further discuss this identification assumption and the advantages and disadvantages of matching estimators. To estimate ATET, instead of condition (1.2.1), we need a weaker mean independence assumption, which is

$$E(Y_0 || X, D) = E(Y_0 || X)$$
(1.2.2)

Matching identification also requires that

$$0 < \Pr(D = 1 \| X) < 1 \tag{1.2.3}$$

This common support condition says at each level of X, the probability of observing both the participants and nonparticipants is positive. (This condition can be enforced by adding a common support constraint, discussed in Section 1.7). Estimating ATET a weaker condition, which is

$$\Pr(D = 1 \| X) < 1 \tag{1.2.4}$$

because we only need to estimate one counterfactual to calculate ATET.

Another implicit assumption required by the matching estimators is the stable unit-treatment value assumption (SUTVA) (Rosenbaum and Rubin 1983). SUTVA states that treatment of a unit affects only the outcome of itself, and in a sample the outcome of unit i given treatment is independent of the outcome of unit j given treatment.

1.3 Propensity Score Matching

Matching on all variables in X becomes impractical as the number of variables increases. To overcome this "curse of dimensionality", Rosenbaum and Rubin (1983)

propose propensity score matching, which reduces a multidimensional matching problem to a one-dimensional problem. Specifically, instead of matching on a vector X, we match on an index function P(X). P(X) is the propensity score, where

$$P(X) = \Pr(D = 1 || X)$$
 (1.3.1)

Rosenbaum and Rubin show that if the conditions in equations 1.2.1 and 1.2.3 are satisfied, then

$$(Y_1, Y_0) \perp D \| P(X)$$
 (1.3.2)

and

$$0 < \Pr(D = 1 \| P(X)) < 1 \tag{1.3.3}$$

If CIA holds given X, it also holds conditional on P(X). The major advantage of matching on P(X) instead of X is the reduction in the number of dimensions.

1.4 Identification Strategy of Matching

The heart of the sample selection problem is the potential correlation between the error terms in the outcome equation and the selection equation. The critics of matching argue that it solves the problem by assuming away its existence. Matching assumes that the right condition variables will eliminate the problematic correlation between the two error terms.⁴ Guidance for choosing conditioning variables can only from economic theory, knowledge of the institutional setting for a particular problem, and previous empirical evidence.

⁴ Please refer to Heckman and Navarro-Lozano (2003) for details on the information requirements for matching.

Another downside of matching is that if we have perfect prediction of participation so that P(X) = 1 or 0, the method breaks down because we do not observe counterparts in the treatment and comparison groups to construct a counterfactual with The identification condition in equation (1.2.3) assumes that, given X, common X. some unspecified randomization device allocates people to treatment. Usually this unspecified randomization assumption is hard for economists to accept. Heckman and Navarro-Lozano (2003) give an excellent discussion of how to solve this identification issue. They argue if there is a vector $Z(Z ext{ is not in } X)$ such that: If P is a nontrivial function of Z (so P(X,Z) varies with Z for all X) and X can be varied independently of Z, and outcomes are defined solely in terms of X, treatment parameters can be defined for at least some values of X in its support, or all values of Xif we are willing to put forward a stronger assumption.⁵ If we can find such Z variables with economic justification and exclude them from the conditioning variables, we can argue condition (1.2.3) is satisfied by the specific economic influences introduced by Z instead of by some unknown random factors.

Given the above two major disadvantages of matching, we must also recognize three important advantages: It does not require separability of outcome or choice equations; it utilizes valuable information in exogenous as well as endogenous variables as long as they are pre-treatment variables; and there is no need for exclusion restrictions or adoption of specific functional forms of outcome equations. Such requirements are common in conventional selection models and instrument variable estimators.

⁵ The stronger assumption would be, for each X, either P(X,Z) = 1 or P(X,Z) = 0.

In summary, pre-treatment variables that influence *both* the treatment participation decision and treatment outcome should be chosen as conditioning variables for matching. If a variable affects only the participation decision but not its outcome, its exclusion will not distort the mean effect to be evaluated in equation (1.1.1). In fact, as discussed above, this type of variable should be excluded from the propensity score model to help the identification. Conversely, if a variable affects only the outcome but not the participation decisions, then this variable will be identically distributed in the treatment and comparison groups, and excluding this type of variable will not introduce selection bias into the evaluation problem. However, if we omit a variable influencing both participation decision and outcome, the final estimates will suffer from selection bias.

1.5 Choice of Matching Method

I now discuss the matching methods used in the following three chapters of this dissertation. Let N_1 be the number of observations in the treatment group and N_0 be the number of observations in the comparison group. The outcomes for the two groups can be written as $Y_1 = \{Y_{1i}\}_{i=1}^{N_1}$ and $Y_0 = \{Y_{0j}\}_{j=1}^{N_0}$ respectively. Consider member *i* from the treatment group. The nearest neighbor matching (with replacement) technique approximates $E(Y_{0i} || D = 1)$ by Y_{0j} , the outcome for member *j* from the comparison group whose propensity score $\hat{P}(X_j)$ is closest to $\hat{P}(X_i)$.

Nearest neighbor matching, although intuitively appealing, is inefficient: it uses only one observation in the comparison group to estimate the potential outcome for a treated observation. Heckman, Ichimura, and Todd (1997, 1998), and Heckman, Ichimura, Smith, and Todd (1998) incorporate local regression into matching. For each observation i ($i = 1, ..., N_1$) in the treatment group, local regression matching opens a window around $\hat{P}(X_i)$ and uses all observations in the comparison group with propensity scores within that window to construct a weighted mean, $\hat{m}(\hat{P}(X_i))$, to approximate $E(Y_{0i} | D = 1)$. Within the window, the closer the $\hat{P}(X_j)$ is to $\hat{P}(X_i)$, the greater the weight the observation j gets in estimating $\hat{m}(\hat{P}(X_i))$.

To formally define local regression, suppose we observe two paired vectors (w_j, z_j) , where j = 1 to n. At each point of interest, W_0 , local regression estimates $m(W_0)$ by solving the following minimization problem:

$$\min_{\alpha_0,\beta_{0,\dots}^l,\beta_0^M} \sum_{j=1}^n \left\{ z_j - \alpha_0 - \sum_{l=1}^M \beta_0^l \left(w_j - W_0 \right)^l \right\}^2 K(\frac{w_j - W_0}{h(W_0)})$$
(1.5.1)

where $K(\cdot)$ is a kernel function and $h(W_0)$ is the bandwidth. In our case the bandwidth varies with W_0 , as will be discussed later. This minimization problem yields $\hat{m}(W_0) = \hat{\alpha}_0$.

Applying local regression to the matching problem defined above, we let $(w_j, z_j) = (\hat{P}(X_j), Y_{0j})$. For each member $i (i = 1, N_1)$ from the treatment group, we run a local regression at its estimated propensity score $\hat{P}(X_i)$ and estimate $\hat{m}(\hat{P}(X_i))$. Of course, to implement this procedure we must choose M, the highest order of the polynomial. Generally, as M increases, the asymptotic bias will be smaller but the asymptotic variance will be larger. Fan and Gijbels (1996) prove that, asymptotically, a choice of M = q, where q is an odd number, dominates a choice of M = q-1. The

intuition is that moving from q-1 to q introduces an extra parameter, reducing the asymptotic bias (especially in boundary regions and in highly clustered regions). There is no corresponding increase, however, in the asymptotic variance. (Their result implies that kernel regression is dominated by local linear regression, at least asymptotically.) Fan and Gijbels (1996) also point out that in practice the typical optimal choice is usually M = 1 and occasionally M = 3. Thus, their work suggests that we use a local linear regression or possibly a local cubic regression.

However, the discussion above does not consider the finite sample behavior of the estimators. Frölich (2001) investigates finite-sample performance of matching estimators including kernel regression (M = 0) and local linear regression (M = 1). He concludes that kernel regression is more robust to misspecification in the bandwidth than local linear regression. Two aspects of Frölich's results are worth noting. First, his results are based on the use of a global bandwidth, and local linear estimators have a well-known problem over regions of sparse data with such a bandwidth. One solution is to use a variable or locally adaptive bandwidth (Fan and Gijbels 1996); bandwidth choice is discussed in the next section. Second, in Frölich's results, the quality of local linear regression depends on the sample size of the treatment group compared to the sample size of the comparison group. Frolich's results suggest that local linear matching performs reasonably well when the comparison group is large relative to the treatment group (say a 5 to 1 ratio of the comparison group to the treatment group). The migration study in Chapter 2 has a ratio of 4.5 to 1, so we expect that local linear regression matching estimator should perform reasonably well. However, for the health insurance study in Chapter 4, the sample size of the treatment group is close to that of the comparison group, so kernel regression is also implemented.

1.6 Choice of the Bandwidth Parameter

The choice of a bandwidth or smoothing parameter is often the most important decision a researcher makes in nonparametric regression. There is a trade-off in choosing the bandwidth - the smaller the bandwidth, the smaller the bias; the larger the bandwidth, the smaller the variance. Basically, there are two types of bandwidths: global (fixed) bandwidths and local (variable) bandwidths. The global bandwidth approach uses the same window width at each point W_0 . The variable bandwidth approach changes the bandwidth according to the data density around W_0 . In other words, it allows us to use a small bandwidth where the probability mass is dense and a larger bandwidth where the probability mass is used at different data locations."

Fan and Gijbels (1992) suggest it is advantageous to combine local regression with variable bandwidth, and this paper uses a simple adaptive variable bandwidth proposed by Fan and Gijbels (1996). In their procedure the size of the window $h_{k_n}(W_0)$ varies with point W_0 ; $h_{k_n}(W_0)$ is chosen to give the same number of data points k_n that are closest to each W_0 to fit the local regression. The number k_n is determined by the sample size n. Essentially, k_n should become larger as the sample size grows but not too quickly.⁶ In the following studies, k_n is determined by setting k_n/n equal to a fixed percentage, say 15% or 25%.⁷

1.7 Common Support Constraint and Balancing Tests

The matching parameter is identified only over the portion of X's support where identification assumption (1.2.3) is satisfied. A common practice when conducting matching is to add a common support constraint to enforce this inequality condition. As discussed in Section 1.2, to estimate ATET we only need a weak condition of equation (1.2.4). The intuition behind condition (1.2.4) is this: if at some level of X we only observe members from treatment group but none from comparison group, for the treated individuals with those X we cannot find similar untreated individuals to construct the counterfactual. In this chapter I follow the procedure proposed by Heckman, Ichimura, and Todd (1997).⁸ If the target trimming level is q, their procedure will trim between q percent and 2q percent of participants. The exact trimming level depends on the data structure. The closer the modes and shapes of the two distributions are, the closer the actual trimming is to q percent. Since the three empirical studies estimate ATET, this procedure trims participants only.

⁶Fan and Gijbels (1996, theorem 4.2) prove that if $k_n \to \infty$ such that $k_n/n \to 0$ and $k_n/\log n \to \infty$, then the adaptive variable bandwidth h_{k_n} behaves asymptotically as $k/\{nf(w)\}$, where k is the number of the nearest neighbors, f(w) is the density function of w_j , j = 1, n and n is the sample size. This bandwidth choice bears some resemblance to the k-nearest neighbor estimates of Härdle (section 3.2, 1990). However, Härdle's estimator puts equal weight on all neighbors, while in our case the weight depends on how close the neighbor is to W_0 .

⁷ Ruppert, Sheather, and Wand (1995) derive three optimal fixed (global) bandwidth selectors for local linear regression. In the migration study, I considered their preferred selector, the direct plug-in bandwidth selector (p. 1262), but it performed poorly in terms of producing matching estimates with large standard errors.

⁸ See also Smith and Todd (forthcoming) for details.

The trimming rule is illustrated in Figure 1. The treatment effect is estimated only for the treated observations between the two vertical lines. The treated at the right tail are discarded because there are no or very few untreated observations. (see the lower panel of Figure 1, and note that the propensity score density of the untreated observations falls below the horizontal line, which is determined by the cutoff level c_q from equation (A2) in Appendix A). The treated observations at the left tail are discarded because there are too few treated observations in this range (in the upper panel of Figure 1, the propensity score density of the treated because there are too few treated observations in this range (in the upper panel of Figure 1, the propensity score density of the treated falls below the horizontal line). Since there is no rule determining the appropriate common support constraints, I try different cutoff levels c_q to test the robustness of the results. Appendix A presents the algorithm for adding the common support constraints.

For the model to be correctly specified, the conditioning variables X should be distributed identically across the treatment group and the matching sample. If they are, the propensity score balances the sample. We can test whether this is satisfied for nearest neighbor matching via two types of tests, paired *t*-tests and joint *F* tests.

Paired *t*-tests examine whether the mean of each element of X for the treatment group is equal to that for the matched sample. However, these tests are not able to detect differences in the two distributions beyond the sample means. Since all matching methods require that the two distributions mimic each other at each quantile, instead of just exhibiting similar means, a joint F test is also used. The treatment group and matched sample are broken down into quartiles according to the estimated propensity scores. Each quartile is tested to determined whether all elements of X are *jointly* different across the two groups. If a model fails to pass either the *t*-tests or the F tests,

12

higher order terms or interaction terms must be added until the variables are balanced across the two groups (Smith and Todd, forthcoming).



Figure 1: Illustration of Adding Common Support Constraint

CHAPTER 2

MATCHING AND SELECTION ESTIMATES OF THE EFFECT OF MIGRATION ON WAGES FOR YOUNG MEN

2.1 Introduction

Internal migration is an important economic phenomenon in the United States. Between March 1999 and March 2000, about 43 million Americans moved, and more than 67 percent of these movers were 20 to 29 years old.⁹ Labor economists typically model migration as an investment in human capital, and a natural question to ask is what the return to this investment is in terms of higher wages. While this issue has received some attention, previous empirical research has focused more on the causes of moving rather than on the consequences. Most migration studies find that factors such as age, education, job tenure, wage on the current job, skills, family composition, length of residence in the current location, local amenities, and the local cost of living affect the migration decision. However, evidence on whether moving increases wages is mixed. By using data on young men from the 1979-1996 waves of the National Longitudinal Surveys of Youth 1979 (NLSY79), this study attempts to quantify the average individual wage gain from U.S. internal migration.

⁹ See http://www.census.gov/prod/2001pubs/p20-538.pdf.

This study contributes to the migration research in several aspects. First, I allow migration effects to differ across education groups and find that this distinction is important. Previous studies have pooled different education groups to estimate the average return to all migrants. If returns to migration are positive for some education group(s), such as college graduates, and negative or zero for other groups, then the overall sample average may be statistically indistinguishable from zero. From matching estimators, I find a significant positive migration effect for college graduates of around 10 percent and a marginally significant negative effect for high school dropouts of about -12 percent. For the overall sample and the other educational groups, I do not find a significant migration effect. From selection models, we cannot have a precise estimate of the moving effect for any specific educational group or for the sample as a whole.

Second, this study uses a distance-based measure of migration instead of a measure based on moving across a state or county line. The confidential geocode for the NLSY79 at the Center for Human Resource Research at the Ohio State University includes the exact latitude and longitude of the respondent's residence at the time of each interview. This allows calculation of a distance-based measure of migration. Compared to a measure based on moving across a state or county line, the measure commonly used in the literature, the distance-based measure of migration corresponds more closely to the theoretical notion of changing local labor markets described by Hanushek (1973). I find that measuring migration by changing state underestimates migration by about 36%, and measuring migration by changing county overestimates migration by about 43%. Further, the significant migration effect for college graduates or high school dropouts

from matching estimators disappears when the alternative measures of moving across state or county are used.

Selection bias can be a severe problem in migration studies. This paper uses both selection models and matching estimators to correct the sample selection problem. A theoretical model is developed to guide the choice of exclusion restrictions in selection models and conditioning variables in matching estimators. Using selection estimators, I find that different model specifications do indeed produce different unconditional and conditional estimates of the effect of migration, although this sensitivity arises from the use of models that are rejected by the data using a simple specification test. For the maintained model I cannot estimate a precise unconditional or conditional estimate of the effect of the sample as a whole.

In constrast, matching estimators produce stable and sensible results. By matching movers and stayers *within* each education group (based on the estimated propensity score) rather than matching individuals across the entire sample, we can implement what Rosenbaum and Rubin (1985) call "finer balancing." The finer balancing matching estimators produce results that are not sensitive to the propensity score model specification, bandwidth choice and trimming level, and they pass balancing tests and a specification test.

Potentially of use to other applied researchers are the following statistical findings. First, there is no advantage in using higher-order polynomials than the local linear regression in the matching procedure, and there may well be a cost in terms of overparamaterizing the model. Second, a variable bandwidth works well in this study, and the results are not sensitive to the size of the bandwidth. Third, the Andrews-

16

Buchinsky procedure for choosing the number of bootstrap repetitions is quite helpful, and for this study it suggests a higher number of repetitions than the number often used by applied researchers. In the case of one estimator, using too small a number of bootstrap repetitions gives misleading results.

The remainder of the chapter is organized as follows. Section 2.2 reviews the migration literature. Section 2.3 develops a theoretical model used to guide the choice of exclusion restrictions for the selection models and the conditioning variables for the matching estimators. Section 2.4 describes the data and the next two sections present the main empirical results from the selection models and matching methods respectively. The final section draws conclusions.

2.2 Migration Literature Review

The most common theoretical model of migration treats the decision to migrate as an investment in human capital: individuals migrate if the present value of real income in a destination minus the costs of moving exceeds what could be earned at the place of origin (Sjaastad 1962).¹⁰ The empirical studies based on this model can be classified into two broad areas for our purposes, those on the determinants of migration and those on the consequences of migration for wages and earnings.¹¹ While the determinants of migration are not the focus of this paper, they play a crucial role in the estimation of the propensity score. Polachek and Horvath (1977) and Plane (1993) find that migration propensities vary over a person's lifecycle. Geographic mobility peaks during the early to mid-twenties and declines thereafter with age, because the time horizon over which gains

¹⁰ See also McCall and McCall (1987). They develop a "multi-armed bandit" approach to the migration decision. Workers rank locations by their pecuniary and nonpecuniary attributes, and then sample locations sequentially until a suitable match is found. Search costs limit the number of markets sampled.

¹¹ Greenwood (1997) provides an excellent review of the literature.

from migration can be realized grows shorter. These studies also find that the propensity to migrate increases with education. Highly educated workers operate in labor markets that compete across broad geographic areas, whereas workers with low levels of education operate in more geographically isolated labor markets. Workers with more education also may be better informed about opportunities outside their local labor market and better able to evaluate that information.

In addition, the migration propensity is affected by migration costs, both pecuniary and nonpecuniary. Goss and Schoening (1984) provide some indirect evidence that households with fewer assets are less mobile, since they find that the probability of migration declines with the duration of unemployment. Lansing and Mueller (1967) report that many moves are attributable to family-related issues, such as proximity to family members or health considerations. Thus, psychic costs entailed in moving away from friends and family play a role in deterring migration.

Of course, in the human capital model of migration, expected wage gains, local demand shocks, and inter-regional differences in returns to skill play an important role in the migration decision. Shaw (1991), Borjas, Bronars, and Trejo (1992b), Dahl (2002), and Kennan and Walker (2003) use a Roy model of comparative advantage to explain migration. Although the human capital model of migration clearly predicts a higher present value of lifetime earnings for those who migrate, the literature on the consequences of migration reaches no consensus on the returns to migration. Estimates of the average contemporaneous returns can be negative, zero, or positive. Positive contemporaneous returns are found by Bartel (1979) for younger workers, Hunt and Kau (1985) for repeat migrants, and Gabriel and Schmitz (1995) and Yankow (2003) for less-

18

educated workers. Negative contemporaneous returns are found by Polachek and Horvath (1977), Borjas, Bronars, and Trejo (1992a), and Tunali (2000).¹² Studies that find statistically insignificant contemporaneous returns include Bartel (1979) for older workers, Hunt and Kau (1985) for one-time migrants, and Yankow (2003) for workers with more than a high school degree.

The sign and significance of the migration effect depend on the sample chosen and on how researchers address three critical questions. First, what is the definition of migration being used? Although all authors have in mind a migration as a change of labor market, most define migration as occurring if a geographic boundary is traversed. The majority of authors, including most of those cited above, focus on interstate migration. A few, such as Hunt and Kau (1985) and Gabriel and Schmitz (1995), define migration as a change of Metropolitan Statistical Area (MSA). Falaris (1987) defines it as a change of Census region. Finally, some authors, such as Linneman and Graves (1983), study inter-county migration. By comparing alternate definitions of migration, I find that migration counts and the estimated returns to migration are sensitive to the definition of migration.¹³

The second question affecting the estimated effect of migration concerns the choice of the comparison group. Most authors use all workers who do not migrate as the comparison group. But it is well known that there is wage growth associated with voluntary job turnover (Topel and Ward 1992). Since most migrants change jobs, the

¹² A negative return is not necessarily inconsistent with utility maximization, since a high growth rate can overcome a negative contemporaneous effect. Alternatively, Tunali (2000) views migration as a lottery and finds that while a substantial proportion of migrants experience wage reductions after moving, a minority realizes very high returns. Individuals are willing to invest in an activity that has a high probability of yielding negative returns because of the potential for a very large payoff.

¹³ The distance-based measure in this study is also used by Baumann and Reagan (2002) to study mobility in Appalachia.

"return to migration" may confound returns to job changing with a return to geographic mobility. Bartel (1979) was the first to focus on the relationship between types of job separation and migration. Others, such as Yankow (1999), condition on job changing but do not differentiate between types of job turnover. Finally, Raphael and Riker (1999) consider only workers who were laid off.

Third, what is the treatment of sample selection? Because migration is a choice variable and not randomly assigned, there is no reason to presume that migrants constitute a random sample of all workers. Nakosteen and Zimmer (1980, 1982) were among the first to provide evidence of positive self-selection into migration. Robinson and Tomes (1982) and Gabriel and Schmitz (1995) also find favorable self-selection. However, Hunt and Kau (1985) and Borjas, Bronars, and Trejo (1992a) find no evidence of self-selection. This study deals with the selection issue using both selection models and propensity score matching.

2.3 A Theoretical Model of Migration

I modify the Willis and Rosen (1979) model of education to apply it to the problem of migration choice. At the beginning of the period, assume all workers have quit their first job. They face a choice between accepting another job locally or moving to another labor market and accepting a job there. Moving involves time costs and pecuniary costs. Switching jobs either locally or in the other market involves search costs. Let X_{ki} denote a vector of observed individual human capital variables and local labor market to which an individual migrates. Let ε_{1ki} and ε_{2ki} represent unobserved individual components (such as motivation and unmeasured ability) of the initial wage

and growth rate respectively in earnings capabilities in location k. At the beginning of the period, expected future earnings in each location k are given by

$$V_{ki} = v \left(X_{ki}, \varepsilon_{1ki}, \varepsilon_{2ki} \right), \quad k = c, m \tag{2.3.1}$$

The cost of changing jobs is given by

$$C_{ki} = C_k (W_i, u_{ki}), \ k = c, m \tag{2.3.2}$$

where W_i is a vector of variables that affects the relative costs of changing jobs locally and across locations. The vector W_i includes tastes for migration measured by family background variables, such as whether the individual was living in his county of birth at age 14. Further, u_{ci} and u_{mi} are unobserved error terms. Net expected future earnings from changing jobs locally and across markets are

$$V_{ki} - C_{ki}, \ k = c, m \tag{2.3.3}$$

Workers choose to migrate if

$$V_{mi} - C_{mi} > V_{ci} - C_{ci} \tag{2.3.4}$$

The starting wage on the second job, y_{ki}^{s} , is

$$\ln y_{ki}^{s} = X_{1ki} \gamma_{1k} + \varepsilon_{1ki}, \quad k = c, m$$
(2.3.5)

where X_{1ki} is a vector consisting of a subset of the elements in X_{ki} , and γ_{1k} is a vector of returns to the X_{1ki} variables. The labor market specific growth rates of wages on the second job are

$$g_{ki} = X_{2ki}\gamma_{2k} + \varepsilon_{2ki}, \quad k = c, m \tag{2.3.6}$$

where X_{2ki} is a vector consisting of a subset of the elements in X_{ki} , and γ_{2k} is a vector of returns to the X_{2ki} variables. Finally, the worker's discount rate, r_i , is a function of family background variables T_i :

$$r_i = T_i \delta + \varepsilon_{3i} \tag{2.3.7}$$

If the individual takes the local job, the expected wage at time t (where t is normalized to zero at the time the worker quits his first job) is

$$y_{ci}(t) = y_{ci}^{s} e^{g_{ci}t}$$
(2.3.8)

Assuming it takes M units of time to move, the expected wage at time t in the new location is

$$y_{mi}(t) = 0 \qquad \text{for } t \le M$$

= $y_{mi}^{s} e^{g_{mi}(t-M)} \qquad \text{for } t > M$ (2.3.9)

This study makes the following assumptions. First, gross utility is the present value of wages. Second, workers face an infinite horizon. Third, the discount rate, r_i , is constant for each individual, where $r_i > \max(g_{si}, g_{mi})$. Finally, the costs of changing jobs and migrating enter the net utility function exponentially. Under these assumptions, the net utility of changing jobs locally can be written as¹⁴

$$U_{ci} = V_{ci} - C_{ci} = \left(\frac{y_{ci}^{s}}{(r_{i} - g_{ci})}\right) e^{-W_{i}\lambda_{c} - u_{ci}}$$
(2.3.10)

where λ_c is a vector that weights individual characteristics to reflect the costs of local job changing. The net utility of changing jobs across labor markets can be written as

¹⁴ I assume that individuals choose sensible parameters for y_k^s and g_k , k = c, m, and then act as if there is no uncertainty. This seems reasonable given the data and empirical model.

$$U_{mi} = V_{mi} - C_{mi} = \left(\frac{y_{mi}^{S}}{(r_{i} - g_{mi})}\right) e^{-r_{i}M} e^{-W_{i}\lambda_{m} - u_{mi}}$$
(2.3.11)

where λ_m is a vector of weight parameters reflecting migration cost.

I define $I_i = \ln(U_{mi}/U_{ci})$. Substituting from equations 2.3.10 and 2.3.11 and taking a Taylor series approximation around the population mean values of $(\overline{g}_s, \overline{g}_m, \overline{r})$ yields

$$I_{i} = \alpha_{0} + \ln y_{mi}^{s} - \ln y_{ci}^{s} + \alpha_{1}g_{mi} + \alpha_{2}g_{ci} + \alpha_{3}r_{i} + \alpha_{4}W_{i} + u_{ci} - u_{mi}$$
(2.3.12)

where $\alpha_1 = 1/(\overline{r} - \overline{g}_m) > 0$, $\alpha_2 = -1/(\overline{r} - \overline{g}_c) > 0$, $\alpha_3 = -M + (\overline{g}_m - \overline{g}_c)/[(\overline{r} - \overline{g}_c)(\overline{r} - \overline{g}_m)]$ and $\alpha_4 = \lambda_c - \lambda_m$.

Substituting equations (2.3.5), (2.3.6), and (2.3.7) into (2.3.12) yields the migration decision rule

$$I_i = \theta Z_{1i} + \varepsilon_i^* > 0 \tag{2.3.13}$$

where $\varepsilon_i^* = \varepsilon_{1mi} - \varepsilon_{1ci} + \alpha_1 \varepsilon_{2mi} - \alpha_2 \varepsilon_{2ci} + \alpha_3 \varepsilon_{3i} + u_{ci} - u_{mi}$, and Z_{1i} contains the unique elements of X_{ki} , W_i and T_i . I assume that all error components in ε_i^* are correlated. Thus the resulting selection rule (2.3.13) depends on variables affecting the new job's initial wage and growth rate in both locations, variables affecting the discount rate, and variables affecting the cost of moving.

Consider now the outcome equations. I define two types of outcome measurements. The first measurement is the starting wage on the second job, which is defined in equation (2.3.5) for both movers and stayers. The second measurement is a difference-in-difference (DID) variable, the starting wage on the second job minus the

ending wage on the first job; this second measurement eliminates the individual fixed effects. Assume that the ending wage on the first job is determined by

$$\ln y_i^E = X_{4i} \gamma_4 + \varepsilon_{4i} \tag{2.3.15}$$

For those who move, the DID outcome is

$$\ln y_{mi}^{s} - \ln y_{i}^{E} = X_{1m} \gamma_{1m} - X_{4} \gamma_{4} + \varepsilon_{1mi} - \varepsilon_{4i}$$
(2.3.16)

and for those who stay, the DID outcome is

$$\ln y_{ci}^{s} - \ln y_{i}^{E} = X_{1c} \gamma_{1c} - X_{4} \gamma_{4} + \varepsilon_{1ci} - \varepsilon_{4i}$$
(2.3.17)

The selection problem come from the correlation between the error term in the selection equation (2.3.13) and the two outcome equations (2.3.5) and (2.3.17). Like Willis and Rosen (1979), I take as a maintained assumption that the family background variables, father absent at age of 14 and residing in the birth county at age 14, are exogenous and do not affect wages.

As discussed in Section 1.4, matching estimators and selection models provide distinctive solutions to the sample selection problem. Exclusion restriction choice is discussed in Section 2.5 and the matching conditioning variable candidates is discussed in Section 2.6.

2.4 Data Description

The primary data source is the 1979-1996 waves of the NLSY79. The survey began in 1979 with a sample of 12,686 men and women born between 1957 and 1964. Interviews were conducted annually from 1979 to 1994, and biennially thereafter.

The NLSY79 provides a comprehensive data set ideally suited for studying migration and job mobility together. First, the longitudinal nature of the data makes it possible to track the same individuals over time as they move across jobs and labor

markets. Furthermore, the NLSY79 data files include detailed longitudinal records of the employment history of each respondent. Second, the confidential geocode includes the exact latitude and longitude of the respondent's residence at the time of each interview. This, in turn, allows calculation of a distance-based measure of migration, so results can be compared with more orthodox measures based on change of county or state. The distance-based measure of migration corresponds more closely to the theoretical notion of changing local labor markets than do the alternative measures. A change-of-county definition of migration misclassifies as migrants individuals who move short distances across county lines but do not change labor markets. A change-of-state definition of migration misclassifies as stayers individuals who move hundreds of miles and change labor markets but remain in the same state.¹⁵ Finally, the NLSY79 data focus on individuals at the outset of their work careers, a stage that exhibits the greatest moving and job changing.

In order to construct a sample suitable for empirical analysis, I introduce several selection criteria (see Table 2.1). The sample is limited to young men since the moving decisions of women are more complicated. Because the focus is postschooling labor market activity, I follow individuals from the time that they leave school. The longitudinal structure of the NLSY79 allows determination of when most workers make a permanent transition into the labor force. Conceptually, this study defines the working career as beginning the first time a respondent leaves formal schooling. To avoid counting summer breaks or other inter-term vacations as leaving school, a schooling exit is defined as the beginning of the first nonenrollment spell lasting at least 12 consecutive

¹⁵ Kennan and Walker (2003) use a change-of-state definition of migration in their structural model. It would not be feasible for them to consider this distance-based definition of migration.
months. Accordingly, respondents are excluded from the sample if the date of schooling exit cannot be clearly ascertained from the data. For example, respondents who are continuously enrolled throughout the observation period or who have incomplete or inconsistent schooling information are excluded from the sample.

Of the 6,403 male respondents in the initial sample, 184 individuals were deleted because they never reported a job. Another 78 were deleted because they never left school, or because an exact date of school leaving could not be determined. Another 215 were deleted because they never reported a civilian job. An additional 3,569 observations were eliminated because I could not identify a "clean" job-to-job transition spanning two consecutive interviews with valid address data. For a job to be considered, it had to last at least 6 months and require at least 25 hours per week. A "clean" job-tojob transition is defined as one with no more than 2 months of nonemployment between the jobs and no more than 3 months overlap while both jobs are being held. Finally, each job had to span the date of an interview. In order to identify migrants, the two jobs had to overlap with at least two consecutive interviews for which I were able to match the respondent's address to a latitude and longitude. After excluding another 279 respondents who did not have valid data for all the variables used in this analysis, ¹⁶ I reached a final sample of 2,078 men (see Table 2.1).

¹⁶ If an hourly wage on the first or second job was reported as greater than \$50 dollars or less than \$3 (in real terms) it was considered invalid.

Criteria	Sample Size	
Male respondents in NLSY79	6,403	
Number after deletion because respondent reported no job	6,219	
Number after deletion because information was missing on the week first left school for at least 12 months	6,141	
Number after deletion because no civilian job observations	5,926	
Number after deletion because no "clean" job-to-job transition with location information*	2,357	
Number after deletion because lacked information on variables used in this analysis was lackingor because the reported hourly wage was less than 3 dollars or more than 50 dollars	2,078	

Source: National Longitudinal Surveys of Youth 1979, 1979-1996 waves.

* The job must last at least 6 months and require at least 25 hours per week. A "clean" job-to-job transition is defined as one involving a quit on the first job and no more than 2 months of nonemployment between the jobs and no more than 3 months overlap while both jobs are being held.

Table 2.1: Sample Selection Criteria and Resulting Sample Size

This study explores migration conditional on voluntarily quitting the previous job. Movers are men who quit their first job and move to a new location, while stayers are those who quit their first job but do not move. In this study, migration was defined to have occurred if the respondent moved at least 50 miles or changed Metropolitan Statistical Areas (MSA) and moved at least 20 miles.¹⁷ I focus on wage growth between the first two jobs with a "clean" job-to-job transition as just defined, simply called job 1 and job 2 hereafter. More than 18 percent of all voluntary job changes involved migration.

Table 2.2 presents descriptive statistics. The third column provides means for the whole sample, and the fourth and fifth columns show means for movers and stayers respectively. The last column presents the difference in means between movers and stayers.

The individual characteristics shown in Table 2.2 are reported as of the end of job 1. The men in the sample are, on average, 26 years old at the time of the job change. The movers are slightly younger. African Americans are more likely than non-Hispanic whites to be stayers, while Hispanics are equally represented in both groups. On average, the movers have higher education, higher AFQT scores, and are more likely to be married. Movers are less likely to own a house prior to job change.

¹⁷ Adjacent county centroids are typically about 25 miles apart, so a move of 50 miles roughly corresponds to a move two counties away.

Variable Name	Variable Definition	tion Means			
Migration Dummy		Whole Sample	Movers	Stayers	Difference
Migrate	=1 if respondent moved at least 50 miles or changed MSA and moved at least 20 miles	0.18 (0.39)			
Individual Characteristics					
Age	Age in years	26.08 (4.37)	25.99 (3.59)	26.10 (4.53)	-0.106 (0.215)
Black	=1 if African American	0.23 (0.42)	0.15 (0.35)	0.25 (0.44)	-0.108 (0.021)
Hispanic	=1 if Hispanic	0.13 (0.34)	0.13 (0.33)	0.14 (0.34)	-0.009 (0.019)
Dropout	=1 if highest grade completed is less than 12	0.17 (0.38)	0.12 (0.33)	0.18 (0.39)	-0.06 (0.019)
High_school	=1 if highest grade completed is equal to 12	0.47 (0.50)	0.33 (0.47)	0.50 (0.50)	-0.166 (0.027)
Some_college	=1 if highest grade completed is greater than 12 and less than 16	0.18 (0.39)	0.18 (0.39)	0.18 (0.38)	0.002 (0.021)
College	=1 if highest grade completed is 16 or greater	0.18 (0.38)	0.36 (0.48)	0.14 (0.34)	0.227 (0.026)
AFQT	Age-adjusted performance on Armed Forces Qualifying Test	-4.20 (21.60)	2.95 (20.38)	-5.79 (21.55)	8.74 (1.17)
AFQT_missing	=1 if AFQT is missing	0.04 (0.21)	0.04 (0.20)	0.05 (0.21)	-0.01 (0.015)
Married	=1 if married, spouse present	0.40 (0.49)	0.44 (0.50)	0.39 (0.49)	0.057 (0.028)
Home_Owner	=1 if own home on job 1	0.20 (0.40)	0.14 (0.35)	0.21 (0.41)	-0.068 (0.020)

Continued

Table 2.2: Variable Definitions and Descriptive Statistics

TC 11	~ ~	1
Table	2.2	continued

Variable Name	Variable Definition	Means			
		Whole			
Migration Dummy		Sample	Movers	Stayers	Difference
Job 1 Characteristics					
log(startwage1)	Logarithm of starting wage on job 1 (\$1990)	2.02 (0.42)	2.13 (0.45)	1.99 (0.41)	0.141 (0.025)
log(endwage1)	Logarithm of ending wage on job 1 (\$1990)	2.09 (0.44)	2.21 (0.47)	2.07 (0.43)	0.142 (0.026)
Tenure	Tenure of job 1	2.60 (2.37)	2.66 (2.35)	2.59 (2.37)	0.070 (0.134)
MSA1	=1 if reside in MSA at time of job 1	0.86 (0.35)	0.84 (0.37)	0.87 (0.34)	-0.029 (0.020)
Private1	=1 if private sector employee on job 1	0.87 (0.33)	0.89 (0.32)	0.87 (0.34)	0.010 (0.01)
Professional1	=1 if professional/managerial occupation on job 1	0.22 (0.42)	0.39 (0.49)	0.18 (0.39)	0.210 (0.027)
Job 2 Characteristics					
Private2	=1 if private sector employee on job 2	0.80 (0.40)	0.87 (0.34)	0.78 (0.41)	0.090 (0.022)
Professional2	=1 if professional/managerial occupation on job 2	0.23 (0.42)	0.39 (0.49)	0.19 (0.39)	0.200 (0.026)
MSA2	=1 if reside in MSA at time of job 2	0.86 (0.34)	0.85 (0.35)	0.87 (0.34)	-0.020 (0.035)
County Earnings	Weekly county earnings in job 2 in \$100 1990 dollars	5.17 (1.11)	5.24 (1.17)	5.16 (1.09)	0.080 (0.062)
log(startwage2)	Logarithm of starting wage on job 2 (\$1990)	2.19 (0.47)	2.30 (0.51)	2.16 (0.45)	0.142 (0.029)
Family Background Cha	racteristics				
Same_county	=1 if respondent resides at age 14 in the same county as county of birth	0.57 (0.50)	0.49 (0.50)	0.59 (0.49)	-0.096 (0.028)
Father_absent	=1 if father absent from the household at age 14	0.20 (0.40)	0.15 (0.35)	0.21 (0.41)	-0.06 (0.019)
Father_college	=1 if father was college graduate	0.14 (0.35)	0.25 (0.43)	0.12 (0.32)	0.132 (0.024)

Note: Sample size equals 2,078, and the sample includes of 378 movers and 1,700 stayers. Standard errors are in parentheses.

The NLSY79 provides detailed information on each respondent's job history and on characteristics of each job, a feature that makes matching an appealing strategy to identify the migration effect. On job 1, movers and stayers are equally likely to live in an MSA, while movers are more likely to have a professional job. The chance that movers and stayers work in the private sector is roughly the same. Movers on average have higher starting and ending wages and have slightly longer tenure.

On job 2, movers and stayers are still equally likely to live in an MSA. Movers are more likely to work in the private sector and have a professional job. County earnings are roughly the same for movers and stayers. Between job 1 and job 2, movers and stayers experience, on average, roughly the same wage gain (around 9 percent).

Family background characteristics are hypothesized to affect the costs of changing jobs across labor markets relative to the costs of changing jobs locally. Characteristics such as whether the father has a college degree are likely to affect resources available to finance a move. Movers are more likely to report that their father had a college degree. As a proxy for ties to the local community, I use one variable that measures whether the respondent was residing at age 14 in the same county in which he was born. Not surprisingly, stayers are more likely than movers to have lived in their birth county at age 14. Another variable used is whether the father was absent at age 14, and this variable should influence both family resources and ties to the local community. Movers are more likely to have an absent father.

Table 2.2 presents systematic differences between movers and stayers. Thus, there is reason to suspect, a priori, that selection will be a serious problem that must be addressed to estimate a return to migration.

31

2.5 Sample Selection Models

2.5.1 Basic Model

This section considers the maximum likelihood sample selection estimator (Heckman 1974; Lee 1978). I will compare the equation describing the initial wage on the second job of those who move to the respective equation for those who do not move (see equation 2.3.5) while correcting for selection bias. As discussed in Section 2.3, the selection rule (2.3.13) depends on variables affecting the new job's initial wage and growth rate in both locations, variables affecting the discount rate, and variables affecting the cost of moving. Some of these variables will be exogenous while others will be endogenous with respect to the outcome equation. The exogenous variables are appropriate for use in the selection model.

Assume that the index function for moving is given by

$$y_i^* = Z_i \beta + \varepsilon_i \tag{2.5.1}$$

where Z contains only exogenous variables in Z_1 of equation (2.3.13). The individual migrates ($D_i = I$) if $y_i^* > 0$. In what follows I refer to the equation for the logarithm of the initial wage on a person's second job as the log wage equation. In an obvious change of notation, I write the log wage equation for a randomly chosen individual who migrates as

$$w_{mi} = \alpha_m + X_{mi} \gamma_m + v_{mi} \tag{2.5.2}$$

The log wage equation for a randomly chosen individual who stays is given by

$$w_{si} = \alpha_s + X_{si} \gamma_s + v_{si} \tag{2.5.3}$$

The next step consists of specifying which variables enter equation (2.5.1). We are particularly interested in variables that do not enter either X_{mi} or X_{si} , since these variables identify the model (unless one is willing to use functional form assumptions to identify the model). As discussed above, these variables should be independent of the error terms v_{mi} and v_{si} for us to obtain consistent parameter estimates of equations (2.5.2) and (2.5.3). The model in Section 2.3 suggests that four types of factors determine migration: a) factors affecting wages in the new job if the individual does not migrate; b) factors affecting wages in the job if the individual migrates; c) factors affecting the costs of moving; and d) factors affecting the individual discount rate (or the ability of an individual to finance a move). These factors are not mutually exclusive.

The most difficult variables to use are those determining wages in the individual's best alternative. The definition of migration in this study is that the individual moves at least 50 miles or changes Metropolitan Statistical Area (MSA) with a move of at least 20 miles. As a result, the number of alternative destinations is large, and specifying the best alternative is problematic. I do not attempt to model explicitly the role of factors affecting wages in the individual's best alternative. Thus the selection equation (2.5.1) will not have a structural interpretation, but this will not affect the consistency of estimates of (2.5.2) and (2.5.3).

As a result there are two potential sources of identification. The first source of identification comes from factors affecting the cost of moving. Here I use a measure of whether the individual was in his county of birth at age 14. This variable acts as a proxy for the level of attachment to the current community; the more attached he is, the less likely he is to leave. Following Willis and Rosen (1979), the second source of

identification is family background variables thought to determine the individual's discount rate. I use measures such as father's education and the family structure when the respondent was 14.¹⁸

Two other variables affect the probability of moving, but it is problematic to use them in the selection model since they are likely to be endogenous with respect to the error terms in equations (2.5.2) and (2.5.3). The first is home ownership. This is likely to affect the probability of moving since the transaction costs of selling a house are much higher than those of leaving a rental unit. However, the unobserved factors that lead to increased home ownership are also likely to lead to higher wages on the second job. The second variable is the wages on the first job. We would expect the lag wages to affect the probability of moving since they give valuable information on the respondent's earning power in both labor markets. However, we would also expect them to be correlated with unobserved components in the wage equation for the second job.

I assume that the errors $(\varepsilon_i, v_{mi}, v_{si})$ have a joint normal distribution with mean zero and variance terms given by

$$V(v_{ki}) = \sigma_k, Cov(\varepsilon_i, v_{ki}) = \sigma_{\varepsilon k}, V(\varepsilon_i) = 1 \quad k = m, s$$

The likelihood function is

$$L = \prod_{D_i=1} \Pr\left(Z_i \beta + \varepsilon_i > 0 \mid w_{mi}\right) \frac{1}{\sigma_m} \phi\left(\frac{w_{mi} - \alpha_m - X_{mi} \gamma_m}{\sigma_m}\right) \bullet$$

$$\prod_{D_i=0} \Pr\left(Z_i \beta + \varepsilon_i \le 0 \mid w_{si}\right) \frac{1}{\sigma_s} \phi\left(\frac{w_{si} - \alpha_s - X_{si} \gamma_s}{\sigma_s}\right)$$
(2.5.4)

¹⁸ I also considered mother's education, but this variable does not affect the migration decision in the sample.

where $\phi(\cdot)$ is the standard normal density function. I use maximum likelihood estimation instead of two-step estimation since the number of observations is limited for those who move, and this will help to obtain the most efficient estimates possible for calculating the migration effects. In fact, the data does not permit precise estimates of the wage equation (2.5.3) for those who move. As a result, I consider more restrictive forms of (2.5.2) and (2.5.3). In the first modification, I only allow the intercept to differ between the two equations (while continuing to allow the equations to have separate error terms). The log wage equation for a randomly chosen individual who migrates is given by

$$w_{mi} = \alpha_m + X_{mi} \gamma + v_{mi} \tag{2.5.5}$$

The log wage equation for a randomly chosen individual who stays is given by

$$w_{si} = \alpha_s + X_{si} \gamma + v_{si} \tag{2.5.6}$$

While the use of equations (2.5.5) and (2.5.6) conserves degrees of freedom, it may be too restrictive a specification. When I examined the unrestricted model given by (2.5.2) and (2.5.3), it seemed clear that the return to migration differed by educational group. I define three educational dummy variables $ED1_i$, $ED2_i$ and $ED3_i$ for high school dropouts, high school graduates and those with some college respectively. (College graduates are the control group.) I define \tilde{X}_{mi} and \tilde{X}_{si} as the remaining vector of explanatory variables for movers and stayers respectively. Then we have for the movers

$$w_{mi} = \tilde{\alpha}_m + \tilde{X}_{mi}\tilde{\gamma} + \delta_{m1}ED1_i + \delta_{m2}ED2_i + \delta_{m3}ED3_i + \tilde{v}_{mi}$$
(2.5.7)

The wage equation for a randomly chosen individual who stays is

$$w_{si} = \tilde{\alpha}_s + \tilde{X}_{si}\tilde{\gamma} + \delta_{s1}ED1_i + \delta_{s2}ED2_i + \delta_{s3}ED3_i + \tilde{v}_{si}$$
(2.5.8)

2.5.2 Different Measues of the Effect of Migration on Wages

First discussed are the average treatment effect (ATE) parameters, denoted as ATE^{s} for the overall sample. For the model determined by (2.5.5) and (2.5.6), a randomly chosen person will gain $ATE^{s} = \alpha_{m} - \alpha_{s}$ by moving. For the model determined by (2.5.7) and (2.5.8), where the effect of migration differs by education group, a randomly chosen individual in the respective group will experience the following gain (which may be negative) from migrating:

$ATE_{hd} = \tilde{\alpha}_m - \tilde{\alpha}_s + \delta_{m1} - \delta_{s1}$	(high school dropout)
$ATE_{hg} = \tilde{\alpha}_m - \tilde{\alpha}_s + \delta_{m2} - \delta_{s2}$	(high school graduate)
$ATE_{sc} = \tilde{\alpha}_m - \tilde{\alpha}_s + \delta_{m3} - \delta_{s3}$	(some college)
$ATE_{cg} = \tilde{\alpha}_m - \tilde{\alpha}_s$	(college graduate)

I then formulize the average treatment effect on the treated (ATET) parameters, denoted as $ATET^{s}$ for the overall sample. For the model given by (2.5.5) and (2.5.6), this takes the form

$$ATET^{S} = \alpha_{m} - \alpha_{s} + E(v_{mi} | D_{i} = 1) - E(v_{si} | D_{i} = 1)$$
$$= \alpha_{m} - \alpha_{s} + E(v_{mi} | Z_{i}\beta + \varepsilon_{i} > 0) - E(v_{si} | Z_{i}\beta + \varepsilon_{i} > 0)$$
$$= \alpha_{m} - \alpha_{s} + \sigma_{\varepsilon m}\lambda(Z_{i}\beta) - \sigma_{\varepsilon s}\lambda(Z_{i}\beta)$$

where $\lambda(Z_i\beta) = \phi(Z_i\beta)/\Phi(Z_i\beta)$ is the well-known inverse Mills ratio term and

 $\Phi(\cdot)$ is the standard normal cumulative distribution function. The mean of this expression is the overall effect. For the model where the effect of migration differs by education group, the effect of treatment on the treated for individuals in the different educational groups is given by

• high school dropout

$$ATET_{hd}^{i} = \tilde{\alpha}_{m} - \tilde{\alpha}_{s} + \delta_{m1} - \delta_{s1} + \sigma_{\varepsilon m} \lambda(Z_{i}\beta) - \sigma_{\varepsilon s} \lambda(Z_{i}\beta)$$

• high school graduate

$$ATET_{hg}^{i} = \tilde{\alpha}_{m} - \tilde{\alpha}_{s} + \delta_{m2} - \delta_{s2} + \sigma_{\varepsilon m} \lambda(Z_{i}\beta) - \sigma_{\varepsilon s} \lambda(Z_{i}\beta)$$

• some college

$$ATET_{sc}^{i} = \tilde{\alpha}_{m} - \tilde{\alpha}_{s} + \delta_{m3} - \delta_{s3} + \sigma_{\varepsilon m} \lambda(Z_{i}\beta) - \sigma_{\varepsilon s} \lambda(Z_{i}\beta)$$

• college graduate

$$ATET_{cg}^{i} = \tilde{\alpha}_{m} - \tilde{\alpha}_{s} + \sigma_{\varepsilon m}\lambda(Z_{i}\beta) - \sigma_{\varepsilon s}\lambda(Z_{i}\beta)$$

To obtain the treatment effects for each group, I take averages of the relevant expression over the group. Note that if $\sigma_{\varepsilon m} = \sigma_{\varepsilon s}$, the conditional and unconditional treatment effects are equal. This will be satisfied if the error terms in the mover and stayer wage equations do not differ.

2.5.3 The Effect of Endogenous Explanatory Variables in the Probit Equation

Some variables suggested by theory (for example, home ownership) are likely to help predict migration status but are also likely to be correlated with the error terms in the wage equation. To examine the problem that arises if one of the probit explanatory variables is correlated with the errors v_{mi} and v_{si} , consider the two-step version of the model (Heckman, 1979) that is often used for estimation. For movers we have

$$w_{mi} = \alpha_m + X_{mi} \gamma + \sigma_{\varepsilon m} \lambda (Z_i \beta) + v'_{mi}$$
(2.5.9)

For stayers we have

$$w_{si} = \alpha_s + X_{si} \gamma - \sigma_{\varepsilon s} \lambda \left(-Z_i \beta\right) + v'_{si}$$
(2.5.10)

If the Z_i variables are correlated with v_{mi} and v_{si} , then all parameters in (2.5.9) and (2.5.10) will be biased and inconsistent unless (X_{mi}, X_{si}) are independent of Z_i . In the empirical work I have the maintained assumption that family background variables are independent of v_{mi} and v_{si} and are not included in X_{mi} or X_{si} . As in Willis and Rosen, this assumption identifies the model. Now consider variables, such as home ownership or the previous wage, for which the independence assumption with respect to v_{mi} and v_{si} is more suspicious. I denote such variables by y_i . Then as a diagnostic I re-estimate the maximum likelihood model, modifying equations (2.5.5) and (2.5.6) as

$$w_{mi} = \alpha_m + X_{mi} \gamma + \delta_m y_i + v_{mi}$$
(2.5.11)

$$w_{si} = \alpha_s + X_{si} \gamma + \delta_s y_i + v_{si}$$
(2.5.12)

I then test whether δ_m and δ_m are significantly different from zero. If the coefficients are not significantly different from zero, y_i can be used as an explanatory variable in the probit; if the coefficients are significantly different from zero they should be excluded from the model.¹⁹

2.5.4 Empirical Results from the Selection Model

The base specification for the probit model includes age; age-squared; and dummy variables coded one for being African-American, being Hispanic, three educational categories (high school dropout, high school graduate, and some college), and being married with spouse present. The base specification also includes two dummy variables from the first job; these are coded one if the respondent worked in a professional occupation and if he lived in an MSA. Two dummy variables function as proxies for family resources: a dummy variable coded one if the father is absent from the household when the respondent was age 14 and a dummy variable coded one if the father

¹⁹ Alternatively, I could use a Hausman (1978) test of the validity of y_i as explanatory variables in the probit equation by comparing the treatment effects with and without these variables in the probit equation.

had a college degree. The first variable reflects the fact that female-headed families are likely to be poorer than other households. The second variable reflects the fact that families generally are better off if the husband has a college degree. Finally, to capture the non-pecuniary costs of moving, a dummy variable is coded one if the individual was living at age 14 in the county in which he was born. Presumably individuals still in the same county at age 14 will have stronger ties to the community and higher psychic costs of moving.

Identification of the selection model essentially comes from the last three variables, which are in the probit equation but not the outcome equations. As in Willis and Rosen (1979), the maintained identification strategy is that family background variables affect discount rates or costs but do not affect current wages. These variables need play a statistically significant role in the probit equation; otherwise there is a "weak instruments" problem in the selection model. The chi-square statistic for the null hypothesis that their coefficients are jointly zero in the probit equation equals 12.26, substantially larger than the critical value of 7.81 at the 5 percent level for 3 degrees of freedom. Moreover, the family background variables have the expected signs. Thus I believe the selection model for the base specification is well identified.

With regard to the outcome equations, I assume that wages on the second job depend on education, age, race, marital status, AFQT score, local labor market conditions, and whether the individual is working in an MSA, in the private sector and in a professional occupation on the second job. The selection estimates from the unrestricted wage equation (2.5.5 and 2.5.6.) model are reported in Table 2.3. It is obvious that these estimates are quite noisy. Therefore, I consider two restricted selection

models. The first model allows the constant to differ for movers and stayers; these results are presented in the first three columns of Table 2.4. The second model allows the effect of migration on wages to differ by educational category. (The null hypothesis that the treatment effects were constant across educational groups is rejected by the data.) These results are shown in the last three columns of Table 2.4.

The average treatment effect (ATE) and average treatment effect for the treated (ATET) are reported in Table 2.5. Panel A of Table 2.5 contains estimates from the first restricted model (allows only the constant to differ for movers and stayers) while Panel B of Table 2.5 contains estimates from the second restricted model (also allows the effect of migration on wages to differ by education group). Formulae phrased in Section 2.5.2 are used for calculating ATE and ATET, and the standard errors are calculated using the delta method.

	Probit	Lnwage2 Migrants	Lnwage2 Stayers
	-5.901	2.403	1.752
Constant	(1.272)	(1.283)	(0.341)
A = -	0.463	-0.036	0.004
Age	(0.096)	(0.09)	(0.027)
Ann Saurand	-0.009	0.001	0.0001
Age_ Squared	(0.002)	(0.002)	(0.0005)
Diast	-0.145	-0.079	-0.051
Віаск	(0.095)	(0.078)	(0.027)
TT:	0.082	0.028	-0.033
Hispanic	(0.103)	(0.071)	(0.029)
	-0.502	-0.627	-0.199
Diopout	(0.134)	(0.118)	(0.049)
Uigh gabool	-0.549	-0.321	-0.19
Hign_school	(0.108)	(0.093)	(0.043)
Come callere	-0.422	-0.295	-0.136
Some_college	(0.115)	(0.084)	(0.041)
AFOT		0.003	0.005
Arųi		(0.002)	(0.001)
AFOT missing		-0.043	0.006
ArQ1_missing		(0.108)	(0.046)

Continued

Table 2.3: Maximum Likelihood Estimates of the Mover-Stayer Model of Wages on the Second Job - Unrestricted Model

Table 2.3 continued

	Probit	Lnwage2 Migrants	Lnwage2 Stayers
Manufal	0.036	0.166	0.082
Married	(0.073)	(0.051)	(0.021)
Dec. C. a	0.367		
Profession11	(0.087)		
	-0.294		
MSAI	(0.095)		
Father about	-0.119		
Father_absent	(0.093)		
	-0.166		
Same_county	(0.068)		
	0.185		
Father_college	(0.097)		
		0.024	-0.057
Private2		(0.072)	(0.022)
		0.041	0.065
Professional2		(0.055)	(0.028)
		-0.016	0.018
MSA2		(0.067)	(0.032)
		0.064	0.072
County_earnings2		(0.020)	(0.009)
		0.422	0.398
Sigma		(0.021)	(0.011)
		0.12	-0.374
Rho		(0.630)	(0.350)
Log Likelihood	-1930.93		

Note: standard errors are in parentheses.

	Only V	Constant Dif Vage Equatio	fers in ns	Education Coefficients Differ in Wage Equations			
	Probit	Lnwage2 Migrants	Lnwage2 Stayers	Probit	Lnwage2 Migrants	Lnwage2 Stayers	
	-5.485	1.047	1.401	-5.479	1.159	1.414	
Constant	(1.283)	(0.359)	(0.324)	(1.281)	(0.370)	(0.329)	
	0.428	0.053	0.053	0.427	0.047	0.047	
Age	(0.098)	(0.025)	(0.025)	(0.097)	(0.026)	(0.026)	
	-0.008	-0.001	-0.001	-0.008	-0.001	-0.001	
Age Squared	(0.002)	(0.0005)	(0.0005)	(0.0020)	(0.0005)	(0.0005)	
	-0.24	-0.08	-0.08	-0.229	-0.082	-0.082	
Black	(0.091)	(0.025)	(0.025)	(0.091)	(0.026)	(0.026)	
	-0.011	-0.017	-0.017	-0.015	-0.017	-0.017	
Hispanic	(0.103)	(0.027)	(0.027)	(0.105)	(0.027)	(0.027)	
	-0.564	-0.378	-0.378	-0.595	-0.643	-0.296	
Dropout	(0.134)	(0.043)	(0.043)	(0.133)	(0.098)	(0.047)	
	-0.627	-0.336	-0.336	-0.618	-0.378	-0.285	
High_school	(0.108)	(0.035)	(0.035)	(0.107)	(0.073)	(0.041)	
	-0.476	-0.257	-0.257	-0.471	-0.356	-0.205	
Some_college	(0.114)	(0.034)	(0.034)	(0.114)	(0.074)	(0.039)	
		0.005	0.005		0.005	0.005	
AFQT		(0.001)	(0.001)		(0.001)	(0.001)	
		0.001	0.001		0.004	0.004	
AFQT_missing		(0.040)	(0.040)		(0.042)	(0.042)	
	0.199	0.098	0.098	0.182	0.101	0.101	
Married	(0.078)	(0.019)	(0.019)	(0.078)	(0.019)	(0.019)	

Continued

Table 2.4: Maximum Likelihood Estimates of the Mover-Stayer Model of Wages on the Second Job – Restricted Models

Table 2.4 continued

	Only Constant Differs in Wage Equations			Education Coefficients Differ in Wage Equations			
	0.298			0.332			
Professional1	(0.083)			(0.085)			
	-0.317			-0.314			
MSA1	(0.093)			(0.094)			
		-0.049	-0.049		-0.049	-0.049	
Private2		(0.020)	(0.020)		(0.021)	(0.021)	
		0.073	0.073		0.07	0.07	
Professional2		(0.024)	(0.024)		(0.025)	(0.025)	
		-0.007	-0.007		-0.004	-0.004	
MSA2		(0.028)	(0.028)		(0.029)	(0.029)	
		0.073	0.073		0.072	0.072	
County_earnings 2		(0.008)	(0.008)		(0.008)	(0.008)	
	-0.059			-0.092			
Father_absent	(0.089)			(0.090)			
	-0.106			-0.119			
Same_county	(0.066)			(0.067)			
	0.159			0.181			
Father_college	(0.095)			(0.095)			
	-0.607			-0.606			
Home_owner1	(0.105)			(0.108)			
		0.458	0.401		0.466	0.396	
Sigma		(0.026)	(0.012)		(0.034)	(0.010)	
		0.44	0.488		0.505	0.328	
Rho		(0.267)	(0.215)		(0.304)	(0.334)	
Log Likelihood	-1930.048			-1919.726			

Note: standard errors are in parentheses.

The ATE for the overall sample from the first restricted model is very small and insignificant. The ATE for the different educational groups from the second restricted model suggests a large negative effect for high school dropouts and a large positive effect for college graduates, but neither of these effects is statistically significant. ATET estimates are reported in the second column of Table 2.5. All conditional returns to migration are positive, although only the conditional mean for college graduates is close to statistical significance at the 10% level. Thus while the base specification model is well identified, the selection model is unable to precisely estimate unconditional or conditional effects.

I also consider two other alternative specifications which exploit potentially endogenous information on the costs and gains of moving. In the first model I add the previous wage to the probit equation. This would usually be considered endogenous. The second alternative specification adds home ownership while on the first job to the probit equation, since given the fixed costs of selling a house, home ownership would be expected to raise the costs of moving. Again one could argue that such a variable is endogenous. To investigate whether the additional variables are rejected by the data, a simple specification test that consists of including the variable to be tested in the two wage equations. If the variable has a statistically significant coefficient in the wage equations, then it is invalid to use the variable to identify the selection model.²⁰ Each of the alternative specifications is decisively rejected using this test and remaining is the

²⁰ The variable will be significant in the wage equation if it is endogenous (correlated with the wage equation error terms) or if it truly belongs in the wage equation.

base specification, which only uses the family background variables to identify the selection model.²¹

	ATE	ATET
A. Returns to Mig	gration Constant Acro	oss Education Groups
All Education	0.027	0.279
Groups	(0.127)	(0.308)
B. Returns to M	ligration Vary Across	Education Groups
Dropouts	-0.21	0.102
	(0.174)	(0.249)
High_school	0.029	0.421
	(0.161)	(0.317)
Some_college	-0.025	0.259
	(0.152)	(0.229)
College Grads	0.096	0.313
	(0.124)	(0.193)

Table 2.5: Migration Effects from Restricted Selection Models

 $[\]frac{1}{21}$ To save space, those test statistics are not reported here.

2.6 Propensity Score Matching

2.6.1 Conditioning Variables and Finer Balancing

As discussed in Chapter 1, the identification strategy for matching is very different from that of selection models. While selection models solve the sample selection problem by explicitly modeling the error term correlation between the selection equation and outcome equation, matching methods address this issue by choosing the correct set of conditioning variables to eliminate this correlation. The conditioning variables are chosen based on the theoretical framework set up in Section 2.3. First, only pre-migration variables from the selection rule of equation (2.3.13) are valid candidates for the propensity score model because the post-migration variables have been contaminated by the treatment. Second, additional variables other than the pre-migration variables in the selection rule are needed to achieve the CIA since there is no reason to assume ε_i^* is independent of $(\varepsilon_{1mi} - \varepsilon_{4i})$ and $(\varepsilon_{1ci} - \varepsilon_{4i})$, a necessary condition for CIA to hold (see equations (2.3.13) and (2.3.16)-(2.3.17)). We would expect that taking the difference between the starting wage on the second job and the ending wage on the first job would help to meet CIA but would not completely satisfy it.²² Instead it is assumed that there are other variables Z_{2i} such that

$$\varepsilon_i^* = f(Z_{2i}) + \tilde{\varepsilon}_i$$
, where $\tilde{\varepsilon}_i \perp (\varepsilon_{1ki} - \varepsilon_{4i}), k = m, c$ (2.6.1)

Of course, the question arises as to what to include in Z_{2i} . We can think of ε_i^* as containing unobservable traits such as ability, motivation, and inclination towards

 $^{^{22}}$ Matching models were estimated with outcome defined as both the starting wage of the second job and the DID variables as the LHS variables of equations (2.3.16) and (2.3.17). I found the latter produces more stable results with respect to model specification and common support constraints. To save space, I only report the matching estimates with DID outcome.

turnover. Variables in Z_{2i} , beginning wage, ending wage, and tenure of the first job, are proxies for the individual specific traits listed above.

Third, the migration decision variables (Z_{1i} in equation 2.3.13) are included in the conditioning variables only if they also affect the outcome, or they are correlated with the error terms in the outcome equations. Variables such as age, education, professional status, marital status, and race will directly affect wages. We would expect that home ownership would affect moving costs and would be correlated with unobservables in the wage equation. We would expect background variables indicating the wealth of the individual's parents to affect the discount rate. Whether these background variables enter the wage equation or are correlated with the unobservables in the wage equation is an open question.²³ I include them in X_i but also experiment with excluding them from the propensity score. Also included in X_i is a dummy variable indicating if an individual lived in an MSA during the first job period.²⁴

Fourth, a Z_{1i} variable will not be included in the conditioning variables if it affects only moving costs but not wages, since excluding such a variable will not introduce selection bias into the outcome evaluation. Furthermore, as discussed in Section 1.4, excluding it from the propensity score will help the matching identification. For example, we would expect individuals who, at age 14, lived in the same county where they were born to have higher physic costs of moving, but we would not expect this variable to affect wages.

²³ Willis and Rosen (1979) assume that family background does not enter the wage equation, nor is it correlated with the error in the wage equation. However, others may find this assumption too strong.
²⁴ Previous studies show workers in cities earn more than their nonurban counterparts after controlling for

earning capability. Glaeser and Mare (2001) suggest that the urban wage premium comes from living in the city, not from innate characteristics associated with urban residence.

In summary, the variables in the propensity score are: labor market history variables, human capital variables, family background variables, and home ownership. Although the CIA assumption is not testable, a credible case for its holding in this economic problem can be made.

The migration effect may depend on the level of schooling. This would occur, for example, if it is much easier for college graduates to search for a higher wage and find a job in a new location without moving there than it is for other educational groups. Let S denote schooling class and s denote a particular schooling level. Now it is possible to estimate

$$\Delta_s = E(Y_1 - Y_0 \| D = 1, S = s) = E(Y_1 \| D = 1, S = s) - E(Y_0 \| D = 1, S = s)$$
(2.6.2)

The first term in equation (2.6.2) is the mean increase in wages for those in schooling class *s* who move. To obtain the second term, we again use matching but only match individual *j* to individual *i* if individual *j* is in individual *i*'s schooling class.²⁵ As noted earlier, Rosenbaum and Rubin (1983) define such a procedure as finer balancing. Following Rosenbaum and Rubin (1985), we can first estimate the propensity score using the entire sample and then match the movers with the stayers in the same educational group based on the estimated propensity score. The empirical work below demonstrates that it is important to allow the treatment effect to vary by educational group.

²⁵ In other words the summation $\sum_{j=1}^{n} [\cdot]$ in (3.7) becomes $\sum_{j \in s} [\cdot]$.

2.6.2 Propensity Score Models

Table 2.6 reports the probit estimates of three models of the propensity score for the migration decision. All models contain variables representing demographics, characteristics of job 1, and home ownership. The models differ in their inclusion of the father's education and same-county variables. From the literature it is not clear whether father's education affects only the resources to finance a move, in which case it does not belong in the propensity score, or whether it is also a proxy for unobserved earning ability, in which case it does belong. The same-county variable, I believe, only represents psychic costs of moving, and its inclusion should not affect the final estimates.²⁶

Model I, the baseline model, contains the core variables and father's education. Model II contains only the core variables. Model III contains father's education, the same-county variable and an interaction between the same-county variable and the professional occupation variable. The interaction term is added to achieve balance between the movers and the matched sample. By comparing the matching estimates of returns to migration in Models I and II, we can investigate whether father's education affects only the resources to finance a move. By comparing the matching estimates in Models I and III, we can test the robustness of Model I to the inclusion of a variable that should not affect estimated returns to migration.

The demographic variables have the expected signs in all three models. Consistent with most migration studies, the results show that the probability of migration starts to

²⁶ We did not choose the other family background variable, father absent at age 14, because it does not significantly influence moving decisions as shown in Table 2.3.

decline at about age 25. Hispanics are more likely and African Americans less likely to move than non-Hispanic whites, although the coefficient on the Hispanic dummy is statistically insignificant. Individuals with less schooling than a college degree are less likely to move than are those with a college degree. Married men are more likely to move than are unmarried men. Men in professional occupations on job 1 are more likely to migrate. Homeownership has a negative and statistically significant effect on migration, as does living in an urban area. The three work history variables (starting wage, ending wage and tenure of job 1) are not significant individually, but a likelihood ratio test shows they are jointly significant. On average, movers have higher hourly wages prior to migration than do stayers.

Although respondents whose fathers have a college education are more likely to move (see Model II), a comparison between Models I and II shows that all of the other coefficients are not sensitive to inclusion of father's education. The Model III results indicate that living at age 14 in the county of birth reduces the probability of migration. The interaction between the professional occupation variable and the same-county variable, which is included only for purposes of balancing, alters only the coefficient on professional occupation. With this one exception, the coefficients are stable across the three models.

	Model I	Model II	Model III
Intercept	-5.86	-5.82	-5.86
	(1.31)	(1.30)	(1.31)
Age/10.0	4.12	4.13	4.15
	(0.99)	(0.98)	(0.99)
Age**2 /100.0	-0.82	-0.83	-0.83
	(0.19)	(0.19)	(0.19)
Hispanic	0.05	0.03	0.04
	(0.10)	(0.10)	(0.10)
Black	-0.24	-0.27	-0.21
	(0.09)	(0.09)	(0.09)
Dropout	-0.50	-0.58	-0.52
	(0.14)	(0.13)	(0.14)
High school	-0.54	-0.60	-0.55
	(0.11)	(0.11)	(0.11)
Some college	-0.38	-0.42	-0.40
	(0.11)	(0.11)	(0.11)
Married	0.17	0.16	0.17
	(0.08)	(0.08)	(0.08)
MSA1	-0.32	-0.30	-0.26
	(0.10)	(0.10)	(0.11)
Professional1	0.34	0.34	0.67
	(0.09)	(0.09)	(0.24)
Jome Owner1	-0.53	-0.53	-0.54
	(0.11)	(0.11)	(0.11)
og(startwage1)	0.18	0.18	0.18
·g(·········g···)	(0.11)	(0.11)	(0.11)
Fenure	0.02	0.02	0.02
	(0.02)	(0.02)	(0.02)
og(endwage1)	0.06	0.07	0.06
og(enuwager)	(0.11)	(0.11)	(0.11)
ather college	0.21		0.19
ather_conege	(0.10)		(0.10)
_	()		(0.10)
Same_county			-0.16
			(0.07)
Same_county*Professional1			-0.37
	8.08	0.72	(0.25)

Note: Values in the parentheses are standard errors.

* Chi-square statistics are from the likelihood ratio tests against the model without the three job 1 variables, starting wage, ending wage and tenure. The critical value at 5 percent significance level is 7.82.

Table 2.6: Propensity Score Coefficient Estimates

2.6.3 Balancing Tests and a Specification Test

Table 2.7 shows the results of the balancing tests of the three models. Panel A shows the paired *t*-statistics on the differences in the variable means between movers and the matched sample of stayers. Panel B presents the joint *F* statistics for the differences in the means of all variables at each quartile of the propensity score. All the tests are conducted with the mover sample and the matched sample from nearest neighbor matching. The *t*-tests in Panel A show that, under all three models, the conditioning variables are well balanced.²⁷ Matching does a good job with regard to pre-migration variables such as race, professional job dummy, and past wages that differ considerably between movers and stayers (see Table 2.2). The joint *F* tests in Panel B demonstrate that the conditioning variables are well balanced iointly at each quartile of the estimated propensity score.

Table 2.8 presents a specification test that examines the "treatment effect" of moving on annual wage growth on job 1 by educational category. Since this variable is pre-treatment, any significant "treatment effect" can only reflect selection bias that finer matching fails to correct. Panel A reports test statistics based on local linear regression matching and Panel B reports those based on nearest neighbor matching. None of these "effects" is significantly different from zero.

²⁷ For Model I the table shows the balancing statistic for the same-county variable even though it is not included in the model. For Model II the table shows the balancing statistics for father's education and the same-county variable, even though these variables are not included in the model. This basically confirms the probit results that both these variables affect the migration decision.

Panel A: t- tests								
	Model I		Model II		Model III			
	Difference	Paired <i>t</i> Statistics	Difference	Paired <i>t</i> Statistics	Difference	Paired <i>t</i> Statistics		
Age	-0.0880	-0.2361	0.0176	0.4798	-0.0147	-0.3986		
Hispanic	0.0117	0.4583	0.0147	0.5846	-0.0147	-0.5768		
Black	0.0235	0.8726	0.0029	0.1123	0.0117	0.4645		
Married	-0.0059	-0.1609	-0.0235	-0.6167	-0.0235	-0.6483		
Father_college	0.0147	0.5020	0.1026	3.7381**	0.0264	0.9761		
MSA1	-0.0147	-0.5620	-0.0088	-0.3414	-0.0059	-0.2261		
Professional1	-0.0411	-1.3488	-0.0235	-0.8525	-0.0059	-0.1922		
Home_Owner1	0.0117	0.5158	-0.0205	-0.7773	-0.0205	-0.8679		
log(startwage1)	-0.0005	-0.0172	-0.0222	-0.7169	-0.0014	-0.0476		
Tenure	-0.0381	-0.2142	-0.0064	-0.0379	-0.0618	-0.3505		
log(endwage1)	0.0023	0.0714	-0.0200	-0.6047	-0.0203	-0.6165		
Same County	-0.1085	-2.8939**	-0.1085	-2.8939**	0.0147	0.4236		

Panel B: F Tests

	Model I	Model II	Model III
1st quartile	0.40	0.84	0.86
2nd quartile	0.52	0.65	0.68
3rd quartile	0.64	0.81	1.29
4th quartile	1.60	0.90	0.99
Critical value at 5% level	F(11, 74) = 1.95	F(10, 75) = 1.99	F(12, 73) = 1.92

Note: All tests based on nearest neighbor matching with q = 5 trimming. ** Significant at the 5% level.

Table 2.7: Balancing Tests (Nearest Neighbor Matching)

	Model I	Model II	Model III
High school dropouts	2.30%	2.33%	3.13%
	(3.61%)	(3.44%)	(3.67%)
High school graduates	-3.16%	-3.18%	-3.07%
	(3.41%)	(3.38%)	(3.28%)
Some college	4.26%	4.46%	2.93%
	(3.52%)	(3.45%)	(3.49%)
College graduates	-0.30%	0.01%	0.28%
	(2.30%)	(2.47%)	(2.47%)

Panel A: Local Linear Regression Matching

Panel B: Nearest Neighbor Matching

	Model I	Model II	Model III
High school dropouts	4.32%	5.41%	1.28%
	(5.34%)	(5.34%)	(5.59%)
High school graduates	-6.74%	-2.41%	-4.50%
	(4.38%)	(4.33%)	(4.16%)
Some college	8.74%	1.25%	5.01%
	(5.34%)	(5.52%)	(5.26%)
College graduates	1.76%	2.54%	0.90%
	(3.45%)	(3.77%)	(3.42%)

Note: All tests based on matching with q = 5 trimming, and and in parentheses are standard errors. In the specification tests, the wage growth is standardized by job tenure. Since this variable is pre-migration, any significant "treatment effect" for this variable can only reflect selection bias that finer balancing matching fails to correct.

** Significant at the 5% level.

Table 2.8: Specification Tests: "Effect" of Migration on Wage Growth on Job 1 by Education Group

2.6.4 Estimates of the Migration Effects from the Baseline Model

Table 2.9 presents the matching estimates of the effect of migration on wage growth from the baseline propensity score model (Model I). Note that all estimates represent the treatment effect on the treated. For all three estimators in Table 2.9, Panel A (with a q = 5 trimming level), I conduct 200, 300, and 1,100 bootstrap repetitions to illustrate the importance of choosing a sufficiently large number of repetitions in calculating standard errors. Appendix B presents an algorithm for choosing the minimum required number of bootstrap repetitions based on the three-step method of Andrews and Buchinsky (2000, 2001). The minimum numbers are 218, 248 and 1,074 for the nearest neighbor, local linear, and local cubic estimators respectively.²⁸ Most of the literature uses at most 200 repetitions. For the nearest neighbor and local linear estimators, the standard errors from 200 repetitions are relatively close to those from 300 or 1,100 repetitions because 200 repetitions are not significantly less than the required minimums of 218 and 248 respectively. However, for the local cubic estimator, the standard errors from 200 or 300 repetitions are dramatically underestimated. For high school dropouts, the estimated standard error increases threefold when the number of repetitions is increased from 200 to 1,100. The large standard errors produced by local cubic matching indicate the problem of overparameterization (Fan and Gijbels, 1996). This problem is masked when standard errors are calculated using only 200 repetitions.

This discussion focuses on the local linear estimates with standard errors calculated from 300 bootstrap repetitions. A 25% bandwidth gives a wide enough

²⁸ For each estimator, I calculate the minimum repetitions required for the overall effect. I then calculate the minimum repetitions for each education group separately. Finally I take the maximum of the five numbers as the required number of repetitions.

window when data are disaggregated by educational class.²⁹ If education level is not distinguished, there is a very small, and statistically insignificant, effect of migration. When the data are disaggregate by education, the effect of migration for high school dropouts is estimated to be -12%. This estimate is significantly different from zero at the 10 percent significance level but not at the 5 percent level. College graduates who migrate experience 10% greater wage growth, and this estimate is statistically significant at the 5 percent level. There is no statistically significant difference in wage growth from migration for job changers who have a high school diploma or some college. Nearest neighbor matching provides noisier estimates of returns to migration because the procedure uses the data less efficiently than local linear matching. Local cubic regression matching produces very large standard errors, which as noted above, indicates the problem of overparameterization as noted above.

Three points are worth noting with respect to the negative estimated effect for high school dropouts. First, I estimate a contemporaneous effect of migration on wage growth. Insignificant or negative contemporaneous effects do not necessarily imply that migration is an irrational decision from the perspective of the human capital approach. As noted in Section 2.2, some previous studies have found that positive returns to migration often are not realized until five or six years after the original

 $^{^{29}}$ To implement finer balancing matching, I first choose a variable bandwidth to give a comparison group equal to 25% of the stayers. I then use only those in the group who are in the same educational category as the mover in question. Each mover gets far less than 25% of stayers in the local regression. I find that the results are not sensitive to a 1% to 2% bandwidth change.

Estimator	Overall	Dropouts	High school	Some college	College Grads			
Local linear 25% bandwidth	-0.56%	-12.46%	-4.73%	-0.75%	10.20%			
(200 repetitions)	(2.35%)	(6.83%)	(3.79%)	(5.66%)	(5.03%)			
[300 repetitions]	[2.32%]	[7.25%]	[3.65%]	[5.66%]	[5.06%]			
{1100 repetitions}	{2.41%}	{7.25%}	{3.99%}	{5.56%}	{5.18%}			
Nearest neighbor (one)	-0.84%	-12.23%	2.46%	-8.34%	5.28%			
(200 repetitions)	(3.53%)	(10.91%)	(5.98%)	(8.94%)	(7.37%)			
[300 repetitions]	[3.58%]	[11.54%]	[5.84%]	[8.87%]	[7.20%]			
{1100 repetitions}	{3.66%}	{11.10%}	{5.90%}	{8.76%}	{7.35%}			
Local cubic 25% bandwidth	-1.64%	-12.51%	-4.90%	-4.48%	9.28%			
(200 repetitions)	(3.48%)	(8.72%)	(5.28%)	(10.75%)	(5.81%)			
[300 repetitions]	[3.62%]	[12.52%]	[5.89%]	[9.93%]	[5.62%]			
{1100 repetitions}	{6.28%}	{29.26%}	{15.05%}	{8.74%}	{6.22%}			
	Pane	el B: Trimming lev	$\operatorname{vel} q = 3$					
Local linear 25% bandwidth	0.63%	-12.46%	-4.73%	-0.70%	12.56%			
(300 repetitions)	(2.27%)	(7.40%)	(3.83%)	(5.73%)	(4.88%)			
Nearest neighbor (one)	0.52%	-12.23%	2.46%	-8.76%	9.07%			
(300 repetitions)	(3.48%)	(11.56%)	(5.93%)	(8.81%)	(6.82%)			
Panel C: Trimming level q = 7								
Local linear 75% bandwidth	0 88%	12 46%	1 730/	0 75%	10 86%			
(300 ropotitions)	-0.0070	-12.4070	-4.75%	-0.7370	(5 20%)			
(Sub repetitions) Nearost neighbor (ana)	(2.4470) 1 200 /	(7.0070)	(3.3370) 2 469 /	(J.0470) 8 340/	(J.2770) 5 00%			
(300 repetitions)	-1.20%	(11.05%)	(5.99%)	(8.92%)	(7.48%)			
(500 repetitions)	(5.0070)	(11.0570)	(3.7770)	(0.7270)	(7.7070)			

Panel A: Trimming level q = 5

Table 2.9: Matching Estimates of the Effect of Migration on Wage Growth from Model I

migration, and the initial returns are negative. It is interesting to note that some of the previous studies found negative returns for the entire sample, while this study finds them (to the extent the difference is significant) only for high school dropouts. Migration may involve an assimilation process. A short-term loss in wage need not, and probably does not, imply a drop in life-time utility. In terms of the model in Section 2.3, the lifetime utility increases for migrants if the growth rate effect dominates a negative or zero initial wage gain. Of course, it may be the case that the model is not appropriate for dropouts. They could be insufficiently skilled to solve the optimization problem, even approximately. Alternatively, they may not be able to see wages in the other location without visiting it.³⁰

Second, unlike most migration studies, this study estimates a migration effect that has netted out the effect of job changing and thus the results do not imply that any group experiences a negative return to job changing. Third, it is possible that return and repeat migration drive the negative returns for high school dropouts, and the data do indication more repeat and return migration for high school dropouts than for other education categories.³¹ To explore this possibility, I excluded those with repeat or return migration from the mover sample. This modification, however, did not change the negative migration effect for high school dropouts or the positive effect for college graduates.

Panels B and C of Table 2.9 present estimates based on alternative trimming levels of q = 3 and q = 7 respectively. (the local cubic estimator is dropped given its poor performance in Panel A.) For both cases (and for the results presented later) I use 300

³⁰As one seminar participant put it, "College graduates can search and then move, while dropouts must move before they can search."

³¹ Return migration within two years is 36% for dropouts and 24% for the overall sample. Repeat migration within two years is 36% for dropouts and 22% for the whole sample.

bootstrap repetitions, more than the required minimum from the three-step Andrews-Buchinsky method. In each case, nearest neighbor matching still produces imprecise estimates. The estimates from local linear matching are not sensitive to this change in the trimming level, except for a two-percentage point difference in the return to college graduates between q = 3 and q = 5. This difference may reflect the widespread finding in the matching literature that the right tail of the distribution of returns is more sensitive to the trimming level than are other parts of the distribution.

2.6.5 Robustness of the Propensity Score Specification

Table 2.10 represents the migration effects estimated from the three alternative propensity score models. Panel A contains the estimates from the baseline model, repeated from Table 2.9, for 300 bootstrap repetitions and serves as a benchmark. Panel B reports estimates from Model II, in which the variable indicating whether the father has a college degree is excluded from the propensity score. The effects estimated with the local linear estimator are almost identical under Model I and Model II. Recall that under Model II, the father's college degree variable is not balanced between the movers and the matched sample of stayers (see Table 2.7). Since father's education is significant in Model I and since migration effects are similar regardless of whether this variable is included in the propensity score, it appears that father's education significantly affects the moving decision but does not provides extra information with regard to unobserved earning ability, after controlling for all the other individual characteristics and the lagged variables. The nearest neighbor matching estimates are very noisy under both Model I and II and, not surprisingly, the point estimates vary substantially across the two models.

Panel C of Table 2.10 reports estimates from Model III, in which the same-county variable is added. This variable is not balanced in Models I and II (see Table 2.7). Further, its coefficient is significant in Model III (see Table 2.6). Again the estimates from the local linear estimator are very close to those from Model I and Model II. Taken together these results suggest that living at age 14 in the birth county is a migration cost variable and does not affect wage growth. In summary, the results from the two alternative models suggest that the baseline model is well-specified and robust to alternative specifications.

2.6.6 Alternative Definitions of Migration

The distance-based measure of migration made possible by the confidential geocode of the NLSY79 data corresponds more closely to a change of labor markets than do two alternative definitions of migration commonly used in the literature: changing state of residence and changing county of residence. Table 2.11 presents summary statistics for all three definitions. The first column of Panel A shows the number of movers and stayers under each definition. The number of people who are considered movers differs substantially according to the definition used. There are only 258 movers (out of 2,078 job changers) who are defined as movers when a move is defined as crossing a state line. In contrast, there are 542 movers when a move is defined as crossing a county line. The distance-based measure produces an intermediate number of movers (378).
Panel A: Model I							
Estimator	Overall	Dropouts	High_school	Some_college	College Grads		
Local linear 25% bandwidth	-0.56%	-12.46%	-4.73%	-0.75%	10.20%		
(300 repetitions)	(2.32%)	(7.25%)	(3.65%)	(5.66%)	(5.06%)		
Nearest neighbor (one)	-0.84%	-12.23%	2.46%	-8.34%	5.28%		
(300 repetitions)	(3.58%)	(11.54%)	(5.84%)	(8.87%)	(7.20%)		
		Panel B: Mo	del II				
Local linear 25% bandwidth	-0.32%	-12.20%	-4.79%	0.03%	10.42%		
(300 repetitions)	(2.46%)	(7.10%)	(3.96%)	(5.61%)	(5.35%)		
Nearest neighbor (one)	-2.21%	-19.08%	-4.56%	4.25%	4.07%		
(300 repetitions)	(3.91%)	(11.22%)	(5.81%)	(8.81%)	(8.08%)		
		Panel C: Moo	lel III				
Local linear 25% bandwidth	-0.02%	-12.80%	-4.12%	-1.48%	12.03%		
(300 repetitions)	(2.51%)	(7.39%)	(4.14%)	(5.66%)	(5.16%)		
Nearest neighbor (one)	-2.09%	-9.01%	-1.73%	-9.78%	5.94%		
(300 repetitions)	(3.76%)	(10.87%)	(6.32%)	(9.75%)	(7.84%)		

Note: All three panels use q = 5 trimming level. See Table 3 for model specifications

Table 2.10: Matching Estimates of the Effect of Migration on Wage Growth Based on Three Alternative Models

Panel A. Distance Between Consecutive Locations (Miles)

Definition of Migration	Migration Status	Ν	Average	Minimum	Maximum
Distance-Based Measure	Mover	378	535	20	3772
	Stayer	1700	4	0	49
Change-of-State Measure	Mover	258	722	1	3772
	Stayer	1820	12	0	668
Change-of-County Measure	Mover	542	379	1	3772
	Stayer	1536	2	0	38

Panel B. Misclassification of Movers and Stayers When Move is Defined as Change-of-State or Change-of-County Relative to a Distance-Based Measure

Change-of-State Measure	Undercounts of Movers Overcounts of Movers	N 136 16	Average 120 16	Minimum 20 1	Maximum 668 44
Change-of-County Measure	Undercounts of Movers	0	-	-	-
	Overcounts of Movers	164	17	1	49

Table 2.11: Comparisons Between Movers and Stayers under Three Definitions ofMigration

Panel A of Table 2.11 also shows the average, minimum and maximum distances between consecutive locations for those classified as movers and stayers under each definition. The average distance for movers ranges from 379 miles under the change-ofcounty definition to 722 miles under the change-of-state definition. The average distance for movers under the distance-based measure is 535 miles. The average distances, however, mask the potential for misclassification inherent in the other two definitions.

Under the distance-based measure, the minimum distance between consecutive locations for movers is 20 miles, conditional on changing residence from one MSA to another. The maximum distance for stayers is 49 miles, conditional on not changing MSA. However, under the change-of-state definition of migration, the minimum distance for movers in this data set is 1 mile, and the maximum distance between consecutive locations for stayers is 668 miles. When a change-of-county definition is used, the minimum distance for movers is 1 mile and the maximum distance for stayers is 38 miles. Both the change-of-state definition and change-of-county definition incorrectly classify as movers those making short-distance changes in residence across a boundary. The change-of-state definition also incorrectly classifies as stayers individuals who make large-distance changes in residences.

Panel B of Table 2.11 shows the magnitude of the potential for misclassification of movers and stayers using definitions of migration based on crossing a state or county boundary. Row 1 describes individuals who are classified as movers under a distancebased measure but are classified as stayers under the change-of-state definition. These 136 individuals (36% of all movers under the distance-based measure) have an average distance of 120 miles between consecutive locations and a maximum distance of 668 miles. Row 2 describes individuals who are classified as stayers under the distance-based measure but are classified as movers under the change-of-state measure. These 16 individuals (less than .5% of all stayers under the distance-based measure) have an average distance between consecutive locations of 16 miles, a minimum distance of 1 mile and a maximum distance of 44 miles. The last row describes individuals who are classified as stayers under the distance-based measure but are classified as movers under the distance-based measure but are classified as movers under the change-of-county definition.³² These 164 individuals (10% of stayers under the distance-based measure) have an average distance between consecutive locations of 17 miles and a minimum distance of 1 mile.

Table 2.12 re-estimates all stages of the matching model using the two alternative definitions of migration. This permits investigation of the potential impact of misclassification, as described in Table 2.11, on the matching estimates. All results are based on Model I, the baseline model with q = 5 trimming. Once again I use 300 bootstrap repetitions, which is more than the minimum number calculated using the three-step method. Compared to the distance-based measure of migration, the alternative definitions yield smaller (in absolute value) and statistically insignificant estimates of the effect of migration on wage growth for dropouts and college graduates. None of the question of how previous estimated returns to migration in studies using different methodologies would change with a distance-based measure of migration.

³² Not surprisingly, the change-of-county definition does not classify as stayers any movers under the distance-based measure.

Panel A: Distance-Based Measure

Estimator	Overall	Dropouts	High_school	Some_college	College Grads
Local linear 25% bandwidth	-0.56%	-12.46%	-4.73%	-0.75%	10.20%
(300 repetitions)	(2.32%)	(7.25%)	(3.65%)	(5.66%)	(5.06%)
Panel B: Change-of-Count	ty Measure				
Local linear 25% bandwidth	-1.98%	-7.35%	-6.23%	-0.48%	7.27%
(300 repetitions)	(2.05%)	(5.70%)	(2.98%)	(5.20%)	(4.81%)
Panel C: Change-of-State	Measure				
Local linear 25% bandwidth	0.03%	-7.13%	-4.71%	-1.67%	8.59%
(300 repetitions)	(2.91%)	(8.49%)	(5.41%)	(6.86%)	(6.70%)

Note: The baseline model , q = 5 trimming level.

Table 2.12: Matching Estimates of the Effect of Migration on Wage Growth Based on Three Definitions of Migration

2.7 Conclusion

This paper estimates the effect of U.S. internal migration, for those who quit their first job, on wage growth between the ending wage on their first job and the starting wage on their second job. The analysis of migration differs from previous research in three important aspects. First, I exploit the confidential geocode in the NLSY79 to obtain a distance-based measure of migration rather than defining migration as a movement across county or state lines. Second, I investigate the effect of migration on wage growth between the first and second jobs separately by schooling level. Third, I use both selection models and propensity score matching to address selection issues and estimate the effect of migration on the wage growth of young men who move. An economic model helps determine which variables should be exclusion restriction variables and which ones should be included in the propensity score used in the matching procedure. Matching is a "data hungry" estimation strategy, and the NLSY79 data set provides a rich array of variables on which to match. This, in turn, makes the Conditional Independence Assumption, which underlies all matching, quite plausible in this case. Specifically, I use variables on previous labor market history, family background, demographics, and homeownership.

The selection model is well identified. However, the selection model cannot produce a precise unconditional or conditional estimate of the effect of moving for any specific educational group or for the sample as a whole. Using matching estimators, I find a significant positive effect of migration on the wage growth of college graduates and a marginally significant negative effect for high school dropouts. I do not find any significant effect for other educational groups or for the overall sample. The results are robust to changes in the model specification and matching method, and the models pass balancing tests and a specification test. These results suggest that the advantage of matching over the selection method is that the former can utilize valuable information contained in lag endogenous variables, of which this data set provides plenty.

I find that better data matters: using a measure of migration based on moving across wither county or state lines, the significant effects of migration on the wage growth of college graduates and dropouts disappear. Finally, this study provides useful information to applied researchers on the highest order of the polynomial needed when using local regression in the matching procedure, and on the number of bootstrap repetitions needed when calculating standard errors.

CHAPTER 3

ARE THERE ABNORMAL RETURNS AFTER SEASONED EQUITY OFFERINGS?

3.1 Introduction

The long-run stock performance of firms participating in Seasoned Equity Offerings (SEOs) has commanded extensive interest in the finance community over the past decade. Loughran and Ritter (1995) and Spiess and Affleck-Graves (1995) report that SEO firms show subpar stock performance 3 to 5 years after the issuing dates. These findings are confirmed by several later studies (for example, Loughran and Ritter 2000 and Jegadeesh 2000). Because long-run underperformance after SEOs is persistent, and the magnitude of abnormal performance is large, this has become one of the strongest challenges to the Efficient Market Hypothesis (EMH). However, other studies raise doubts about these results. Brav, Geczy and Gompers (2000) show that stock returns after SEOs follow a more pervasive return pattern for small stocks, and hence, abnormal returns following SEOs may not be an anomaly. Eckbo, Masulis and Norli (2000) argue that the SEO puzzle may be explained by the inability of the matching techniques to properly control for risk factors. Therefore, evidence on the existence of the long-run underperformance of SEO firms is still open to question.

Furthermore, supplementary studies on earnings manipulation and insider trading around SEOs also yield mixed results. Brous, Datar and Kini (2001) find that investors are not systematically disappointed by earnings announcements following SEOs, leading to questions about the validity of the optimistic expectations hypothesis that is commonly used to explain the underperformance phenomenon. On the contrary, Jegadeesh (2000) finds that a large portion of the SEO underperformance occurs on or around earnings announcement dates. Lee (1997) fails to find that insiders take advantage of the windows of opportunity and sell "overvalued" shares around SEOs. In contrast, Kahle (2000) establishes a link between SEO underperformance and measures of abnormal insider trading, and Clarke, Dunbar and Kahle (2001) find increases in insider selling prior to SEOs.

Hence, evidence on long-run SEO underperformance is far from conclusive. The major difficulty arises from the fact that market efficiency can only be jointly tested with a well-specified model for "normal" returns. Unfortunately, all models estimating "normal" returns are inadequate in capturing all risk factors, and thus Fama (1998) points out that all these studies may be plagued by the "bad model problem."

There are two approaches in current literature for estimating "normal" returns: factor models and matching methods. Several papers examine the reliability of factor models in studies of long-run stock performance. Barber and Lyon (1997), Kothari and Warner (1997), and Loughran and Ritter (2000) all find that factor model benchmarks are misspecified. These results should not be surprising since neither CAPM nor the Fama-French three-factor model have good explanatory power for small stocks, where most SEO firms belong (Banz 1981, Fama and French 1992, and Fama and French 1993). Therefore, prior evidence usually shows that the long-run abnormal performance mostly resides with small stocks; when the portfolio is value-weighted, the long-run underperformance tends to disappear (e.g., Fama 1998 and Mitechell and Stafford 2000).

Long-run stock performance after SEOs can be investigated using a typical evaluation study framework that has been discussed in great detail in Chapter 1. Much of the literature finding SEO underperformance uses traditional matching methods to solve the sample selection problem. These researchers compare *ex post* returns of event firms with those of non-event firms matched by *ex ante* size, book-to-market ratio and, sometimes, past returns. Unfortunately it is difficult to match firms in several dimensions concurrently, and this study shows that the traditional matching methods are not able to yield good matches *ex ante*. Furthermore, there is no guarantee that all risk factors have been incorporated in the model. Therefore, the concern that traditional matching methods are not able to sufficiently control for all risk factors is valid, and any conclusions based on these models may not be reliable.

To cope with the multi-dimensional matching problem, this paper implements propensity score matching. This method is able to match firms simultaneously in as many dimensions as needed, and thus it not only overcomes the "dimensionality curse" that troubles traditional matching methods, but it also provides a convenient way to investigate impact from additional risk factors. Therefore I am able to match risk factors in a way that is not feasible in earlier studies.

In terms of addressing the sample selection problem, another approach, as used in Chapter 2, is the selection models. As discussed before, selection models require at least one variable significantly affecting the SEO participation decision but uncorrelated with

71

post-event stock market performance. On the other hand, matching methods require a rich data set that contains *ex ante* variables related to both the SEO decision and *ex post* stock market performance. Long-run stock performance after major corporate events is a type of evaluation problem for which it is difficult to find an exclusion restriction variable but for which a rich set of conditioning variables is available, and matching therefore makes a plausible identification strategy to long-run stock performance study.

This study investigates various market and accounting variables. I confirm that size, book-to-market ratio and past returns, the three factors commonly believed to affect stock returns, are the main factors that affect both a firm's decision to issue new equities and estimated abnormal returns after SEOs. I find that additional variables have only very marginal effects. When the propensity score is estimated conditional on size, book-to-market ratio and past returns, I am able to match these *ex ante* characteristics well. Using data from the period 1986-1998, I find that abnormal returns after SEOs disappear.

This paper does not attempt to address the broader question of market efficiency. It merely shows that the previous evidence on the long-run abnormal returns after SEOs is not reliable. SEO firms do not behave differently from similar firms that are small in size andhave low book-to-market ratios and high recent returns.

The rest of the chapter is organized as follows. Section 3.2 describes the data in this study. Section 3.3 shows the results on SEO long-run underperformance using traditional matching methods and explains the problem of "dimensionality curse" that faces these methods. The technical issues of propensity score matching related to this study is discussed in Section 3.4. Section 3.5 illustrates the procedure used to implement the propensity score model using the year 1993 sample. Section 3.6 conducts a full-scale

investigation of stock returns following SEOs during the period from 1986 to 1998. The final section draws conclusions.

3.2 Sample Selection

SEO observations are obtained from the Securities Data Corporation (SDC) New Issues database. This study includes only those that involve primary offerings; pure secondary offerings, units and warrant offerings are excluded. There is a total of 5,399 SEO observations (3,664 issuing firms) to start with during the period from 1986 to 1998.

The following restrictions are imposed to construct the SEO sample set:

For each SEO observation in year *t*,

- All issuer stocks are found in the Center for Research in Securities Prices (CRSP) monthly stock return database at the time of the SEO date. The security must have a CRSP share code of 10 or 11 (US common stocks).
- 2. A firm's equity market value (price multiplied by number of shares outstanding) must be available in the CRSP database on December 31, *t*-1.
- 3. Accounting data relevant to book-value must be available on Compustat. I collect annual data reported as of data year *t*-2 in Compustat, which usually becomes available in mid *t*-1. Observations with a negative book value for equities are excluded.
- 4. Only the first issue is included in the study if a firm makes multiple issues during a 36-month period.
- 5. Utility firms are excluded.

After imposing these screens, 2,212 SEO observations remain (these are referred to as sample 1 in this paper) during the study period. This is the base sample size for two-

dimensional matching. The sample size for three-dimensional matching is smaller (1,890 firms, referred to as sample 2 in this paper) due to missing past returns. More observations are lost from propensity score matching (samples 3 and 4) due to outliers and common support constraints. Table 3.1 presents the respective sample sizes in this study.

	2 Dimensional	3 Dimensional		
Method	Matching	Matching	Propensity So	core Matching
			w/o common support	w/ common support
year	(1)	(2)	(3)	(4)
1986	167	152	152	140
1987	127	101	101	91
1988	48	40	40	36
1989	88	78	78	72
1990	78	75	75	68
1991	226	208	208	186
1992	204	181	181	165
1993	237	199	197	182
1994	150	121	121	112
1995	225	187	187	170
1996	286	241	239	205
1997	233	187	186	166
1998	143	120	120	112
Total	2212	1890	1885	1705

Table 3.1: SEO Sample in This Study (1986 – 1998)

Size, book-to-market, and past return variables are defined as follows:

BE (book value of equities): defined following Fama and French (1997) as stockholders' equity (data 216), plus balance sheet deferred taxes and investment tax credit (if available, data 35) and postretirement benefit liabilities (if available, data 330), minus the

book value of preferred stocks (estimated in the order of the redemption (data 56), liquidation (data 10), or par value (data 130), depending on availability).

Mcap (market capitalization or size): stock price times the number of shares outstanding on December 31, *t*-1.

BM (Book-to-market ratio): BE divided by Mcap.

QR1: buy-and-hold returns during the fourth quarter of year *t*-1.

QR2: buy-and-hold returns during the third quarter of year *t*-1.

QR3: buy-and-hold returns during the second quarter of year t-1.³³

Pre9: buy-and-hold returns from April to December during year t-1.³⁴

Note that I pay special attention to past returns and strive for better matching using the past three quarters' returns. Fama and French (1996) show that short-term returns are a dimension not completely captured by size and book-to-market ratio. Since recent stock returns are an important factor in SEO decisions, I use past returns in addition to size and book-to-market ratio to capture the unidentified risk factors indirectly. Furthermore, I break past returns into three quarters because quarterly returns provide a dynamic pattern instead of a snapshot. For example, an SEO firm's returns may follow a down-up-up pattern in the last three quarters before SEO, while a selected matching firm's returns may follow an up-up-down pattern during the same period. The aggregate return may fail to discern this difference in the past return pattern between the SEO firm and the match because of the cancel-out effects. However, price run-up or

³³ Past returns from earlier quarters are not included because I find that they do not seem to play an important role either in a firm's SEO decision or in long-run returns. Section 3.4 explains why these variables, which are not important in either treatment assignment or treatment outcome, do not need to be included in the propensity score model.

³⁴ This variable is the aggregate of QR1, QR2 and QR3. This variable is used in order to show that quarterly returns are better able to capture past return dynamics than aggregate returns. This will be discussed in detail in Section 3.5.

market timing has been proposed as the main reason for equity issuance, and it is therefore very important to capture this price run-up pattern. Using returns over shorter intervals, such as quarterly returns, will overcome this problem.³⁵ Use of quarterly returns poses a major hurdle to traditional matching methods because it increases the number of dimensions to be matched. Fortunately, the propensity score matching method can accommodate extra dimensions and can easily capture the dynamic pattern in past returns.

I also investigate other variables, which are listed in Appendix C. These variables are found to play insignificant roles in firms' equity issuance decisions and thus are not included in the propensity score models, as will be explained in Section 3.5.

3.3 Traditional Matching Methods

This study implements both 2-dimensional and 3-dimensional matching on the 1986-1998 data to illustrate the "dimensionality curse" problem. In 2-dimensional matching, I follow the method in Eckbo, Masulis and Norli (2000). For each SEO observation in year *t*, I first generate a list of all non-SEO firms that have not made equity offerings during the previous 36 months and are within 30 percent of the size of the SEO firm at the end of year *t*-1. The reason the size benchmark is tailored to each SEO firm instead of using the entire size group is to provide a closer size match. This list serves as the original list of matching candidates for the SEO firm. From this list, I further select the firm that has the book-to-market ratio closest to the SEO firm. If the matching firm is dropped at that point. The firm with the next closest book-to-market ratio from the original list of matching candidates is used as the substitute for the period afterwards. If

³⁵ Section 3.5 elaborates on this point.

an SEO firm issues or is delisted within the post-event 36-month window, I use its cumulative return until the issuance date or delisting date, whichever comes first.

In 3-dimensional matching, this study follows the method in Brav, Geczy and Gompers (2000). For year t, I first assign each firm (both SEO and non-SEO firms) that has not issued stocks during the previous 36 months into one of four size groups. Within each size group, I rank firms according to book-to-market ratio and sort firms into four book-to-market sub-groups. Furthermore, within each size and book-to-market group, I sort firms in terms of their past 9-month returns, and classify firms into 4 additional subgroups. ³⁶ Accordingly, every firm is assigned to one of the 4 x 4 x 4 groups. As in earlier studies, NYSE firms are used to determine break points for size, book-to-market ratio and past 9-month returns. Each SEO firm is matched with a portfolio consisting of all non-SEO firms residing in the same group. As in the case of 2-dimensional matching, non-SEO firms in the matching portfolios are dropped whenever they issue stocks again or are delisted. If an SEO firm issues or is delisted before the 36th month, its cumulative return is used until the issuance date or delisting date, whichever comes Because 3-dimensional matching requires past 9-month returns, some SEO first. observations are dropped due to missing past returns, and the total sample size decreases to 1,890 for this method (referred to as sample 2 in the paper). Notice that under both matching methods, IPO firms during the three-year post-event window are excluded from the matching group.

Table 3.2 shows the difference in *ex post* returns during the 36-month window between SEO firms and the matching firms using traditional matching methods. From

³⁶ I use past nine-month returns to be consistent with the three past quarterly returns in the propensity score models.

1986 to 1998, the 36-month abnormal returns range from -37.31 percent to 17.11 percent under 2-dimensinal matching, and from -39.04 percent to 1.57 percent under 3dimensional matching. The average over the whole period is -14.58 percent under 2dimensional matching and -16.40 percent under 3-dimensional matching; both are statistically significant.³⁷ Therefore, consistent with earlier studies, this sample provides evidence supporting long-run stock underperformance after SEOs using traditional matching methods with equal-weighting.

	2 Dimensional Matching			3 Dimensional Matching				
		Sample (1	l)		Sample (2	2)		
	SEO	Matching		SEO	Matching			
year	firms	Firms	Difference	firms	Firms	Difference		
1986	10.17%	13.30%	-3.13%	12.17%	19.34%	-7.17%		
1987	-10.67%	-7.89%	-2.78%	-9.84%	-2.15%	-7.69%		
1988	21.29%	6.35%	14.94%	18.36%	16.79%	1.57%		
1989	6.77%	40.71%	-33.93%	6.33%	38.38%	-32.05%		
1990	25.67%	58.44%	-32.77%	26.82%	65.85%	-39.04%		
1991	33.46%	38.30%	-4.84%	35.24%	56.77%	-21.52%		
1992	38.36%	50.52%	-12.16%	41.64%	49.97%	-8.33%		
1993	26.22%	63.53%	-37.31%	27.63%	54.36%	-26.74%		
1994	42.39%	53.19%	-10.80%	51.78%	62.41%	-10.63%		
1995	34.77%	52.60%	-17.84%	41.76%	47.67%	-5.91%		
1996	4.98%	33.48%	-28.49%	5.90%	32.69%	-26.79%		
1997	27.88%	39.93%	-12.05%	34.38%	40.24%	-5.86%		
1998	13.88%	-3.23%	17.11%	9.09%	30.06%	-20.96%		
0	00 450/	27.020/	14.500/	25.100/	41.500/	16 400/		
Overall	22.45%	37.03%	-14.58%	25.12%	41.52%	-16.40%		
S.E.	[2.53%]	[5.48%]	[4.26%]	[2.81%]	[0.86%]	[2.83%]		

Table 3.2: *Ex post* 36-Month Returns for SEO and Matching Firms

³⁷ Standard errors are obtained using bootstrap.

However, I argue that the traditional matching methods are not able to control properly for risk factors. Table 3.3 shows the mean differences in size, book-to-market ratio and past returns between SEO and matching firms during the period 1986-1998. Standard errors, shown in brackets, are obtained from bootstrap. Although Panel A shows that 2-dimensional matching is able to match neither size nor book-to-market ratio in the statistical sense, the differences in these dimensions are not economically significant, since SEO firms are less than 6 percent smaller in size than their matching firms and only about 2 percent higher in book-to-market ratio. Therefore, this method is able to achieve relatively good matching quality in these two dimensions. The problem for 2-dimensional matching resides in matching of past returns, the dimension that it does not attempt to match. Note that the past 9-month returns for SEO firms are more than 60 percent higher than those of the matching firms, and that the difference is statistically significant. This difference demands attention for two reasons. First, the capital structure literature shows that stock returns are a major determinant of firms' equity issuance decisions, so it is very important to control for this dimension.³⁸ This rationale is explained in detail in the next section. Second, there is no reason to believe that size and book-to-market ratio are able to control for all risk factors (Fama and French 1996). Since any risk factor not captured by size and book-to-market ratio should be reflected in returns, failure to balance past returns implies that some risk factors are not accounted for, and the SEO firms may be fundamentally different from their matching firms *ex ante*.

Panel B shows *ex ante* balance from 3-dimensional matching. Note that this method is not able to balance size between SEO and matching firms, with SEO firms about 25 percent smaller than matching firms. Because it has been well documented that

³⁸ For example, Hovakimian, Opler and Titman (2001) and Baker and Wurgler (2002).

size is a major risk factor, failure to balance size indicates that SEO and matching firms may be very different. Additionally, SEO firms show significantly lower past returns than

Panel A: 2-dimensional Matching (size and book-to-market) - Sample (1)							
	SEO firms	Matching firms	Difference				
Difference in <i>ex ante</i> variables							
Market capitalization	495,784	525,193	-29,409				
			[8,823]**				
Book-to-market ratio	0.4608	0.4528	0.0081				
			[0.0037]**				
Past 9-month returns	37.59%	14.12%	23.47%				
			[2.12%]**				
Panel B: 3-dimensional Matchir	ng (size, book-to-ma	rket and past 12-month	returns) -Sample (2)				
	SEO firms	Matching firms	Difference				
Difference in <i>ex ante</i> variables		0					
Market capitalization	545,247	682,382	-137,135				
			[45,364]**				
Book-to-market ratio	0.5151	0.5270	-0.0118				
			[0.09]				
Past 9-month returns	39.80%	46.15%	-6.35%				
			[1.54%]**				

Table 3.3: Difference in ex ante Variables from Traditional Matching

non-SEO firms. It is infeasible to increase the number of size groups to enhance the overall matching quality. The reason is that increasing the number of size groups will sharply reduce the number of firms in each size subgroup and thus will adversely affect

matching quality in the dimensions that are subsequently matched, namely book-tomarket ratio and past returns.

The point illustrated here is that traditional methods are not able to control properly for risk factors. The reason is that it is very difficult to match <u>all</u> dimensions simultaneously. Although 2-dimensional matching is able to achieve good matches in the 2 dimensions it strives to match, it ignores matching quality in another key dimension. As matching moves from 2 dimensions to 3 dimensions, it becomes very hard to balance all dimensions; matching quality in one dimension is usually achieved at the expense of one or more other dimensions. This problem is referred to as the "dimensionality curse." Since SEO and matching firms appear to be considerably different when traditional methods are used, the concern that risk factors may be insufficiently controlled for is valid, and any conclusion based on these matching methods may not be reliable. Furthermore, because traditional methods are not able to match these three dimensions simultaneously, they would be hard-pressed to match additional characteristics if the need arises. Hence, a better matching method is needed.

3.4 Propensity Score Matching

In this study, the SEO decision is defined as the treatment. Accordingly, the *treatment group* consists of SEO firms, while the *comparison group* consists of all non-SEO firms. The outcome is the *ex post* stock market returns, namely the 36-month buy-and-hold stock returns after the SEO event.³⁹

It is worth noting that propensity score matching and the traditional matching used in Section 3.3 are based on exactly the same identification conditions as discussed in

81

Chapter 1. The only difference between them is the matching mechanism: the former conditions on a vector X, whereas the latter conditions on an index P(X). Propensity score matching is essentially a more efficient matching mechanism due to the reduction in number of dimensions. As illustrated in Section 3.3, traditional matching does suffer from the "dimensionality curse" in this case. The next section will use 1993 data to illustrate how the "dimensionality curse" can be overcome by propensity score matching.

Since this study deals with a complicated situation involving multiple issuances, delisting, and switching from the comparison group into the treatment group over time, it would be computationally demanding to implement local regression matching. I only implement nearest neighbor matching with replacement.

The most crucial part of propensity score matching is model specification, which boils down to choice of X. In this study, I draw on finance theory and empirical evidence to select variables that may affect both the SEO decision and future stock performance. Section 3.5 illustrates the approach used to determine whether a variable should be included in the propensity score model. I also add a common support constraint and conduct balancing tests as described in Section 1.7.

3.5 Illustration of the Propensity Score Matching Method Using 1993 Data

For each year from 1986 to 1998, I estimate a separate propensity score model. This section describes the procedure used to choose an appropriate propensity score model illustrated by 1993 data. I delete ratio outliers more than three standard deviations away from their means because these outliers may contaminate the propensity score models. For example, if the propensity score model includes size, book-to-market ratio

³⁹ As in most long-run performance studies, there is a relatively small treatment group and a relatively large comparison group. The relative sample sizes make matching from the treated (SEO) firms to the untreated

and past returns, I exclude those observations whose book-to-market ratios are more than three standard deviations away from the mean.

As discussed in Section 1.4, the model selection criteria require incorporation of variables that affect <u>both</u> SEO decisions and estimated *ex post* abnormal returns. I report the coefficient estimates from the empirical propensity score models in Table 3.4 and present balancing tests of *ex ante* variables in Table 3.5.

I start model specification with firm characteristics that have been previously identified to affect returns: size, book-to-market ratio, and past returns. I present models with aggregate past returns and quarterly returns, respectively, in order to highlight the advantage enjoyed by quarterly returns. In both tables, Models 1 and 2 use size, book-to-market ratio and aggregate past 9-month returns as RHS variables. Model 1 uses linear terms only, and Model 2 is a refined version of Model 1 with the addition of higher order and interaction terms. Models 3 and 4 use size, book-to-market ratio and past quarterly returns as RHS variables; Model 3 uses linear terms only and Model 4 adds higher order and interaction terms. Higher order and interaction terms are added in Models 2 and 4 in order to balance *ex ante* variables, as required by the two balancing tests.

The two linear models, Models 1 and 3, both indicate that SEO firms are usually smaller in size, have lower book-to-market ratios, and show stronger past returns (Table 3.4, first and third column) than non-SEO firms. These results are consistent with empirical evidence from earlier studies (e.g. Jung, Kim and Stulz 1996).

⁽non-SEO) firms feasible.

	Model 1	Model 2	Model 3	Model 4
Intercept	-2.26	-2.28	-2.43	-2.64
	(0.10)**	(0.11)**	(0.12)**	(0.13)**
BM (Book-to-market)	-0.61	-0.65	-0.51	-0.43
	(0.14)**	(0.15)**	(0.14)**	(0.14)**
Mcap (size)	-0.83	-4.75	-0.76	-0.83
(in the scale of E-4)	(0.41)**	(1.30)**	(0.40)*	(0.43)*
QR1			0.71	2.58
			(0.16)**	(0.42)**
QR1sq				-1.71
-				(0.43)**
QR2			1.37	1.83
			(0.21)**	(0.32)**
QR2sq				-0.4613
				(0.26)*
QR3			0.42	0.42
			(0.24)*	(0.24)*
Pre9	0.53	1.21		
(Pre 9-month returns)	(0.10)**	(0.18)**		
BM*Mcap		0.05		
-		(0.01)**		
Pre9*Pre9		-0.24		
		(0.07)**		
Pre9*Mcap		4.09		
_		(1.75)**		

Table 3.4: Propensity Score Model Coefficient Estimates

	Model 1	Model 2	Model 3	Model 4
Panel A: Mean tests (T-te	sts)			
BM (Book-to-market)	0.08	0.02	0.02	0.02
	(3.33)**	(0.50)	(0.46)	(0.50)
Mcap (size)	53,756	-60,030	-17,346	-42,238
	(0.46)	(-0.53)	(-0.15)	(-0.32)
QR1	3.46%	5.14%	5.14%	-0.21%
	(1.59)	(1.67)*	(2.12)**	(-0.09)
QR2	3.56%	0.40%	-2.85%	-0.53%
	(1.91)*	(0.20)	(-1.94)*	(-0.35)
QR3	0.99%	-2.97%	2.13%	0.79%
	(0.54)	(-1.44)	(1.05)	(0.37)
QR4	-4.89%	-1.71%	-9.01%	-5.93%
	(-1.21)	(-0.46)	(-1.17)	(-1.57)
Pre 9-month returns	9.47%	1.25%	6.40%	0.27%
	(4.12)**	(0.28)	(2.34)**	(0.10)
Panel B: Joint tests (F-tes	ts)			
First quartile	2.94	1.20	1.30	0.16
Second quartile	3.31	0.76	2.98	0.29
Third quartile	1.38	1.60	0.22	0.99
Fourth quartile	1.29	1.44	0.85	1.17
Critical Value	2.79	2.79	2.79	2.79

Table 3.5: Balancing Tests of Propensity Score Matching: Difference in *ex ante* Variables (1993 Sample)

I then explore whether these models have balanced all *ex ante* variables in Table 3.5. Panel A shows mean differences between SEO and matching firms for each ex ante variable, and Panel B reports joint F tests.⁴⁰ Note that these linear models, Models 1 and 3, are not able to balance ex ante variables between SEO and matching firms. Specifically, SEO firms have significantly higher book-to-market ratios than their matches and higher past 9-month returns under Model 1. In addition, the F statistics are significant for the first and second quartiles. Model 3 is able to balance book-to-market ratio but the difference in past returns between SEO and matching firms is still significant, and the second quartile exhibits a significant F statistic. In summary, both t and F statistics show that the linear models are inadequate in balancing ex ante variables. As discussed before, higher order or interaction terms are needed in the propensity score models. With the addition of non-linear and interaction terms, Models 2 and 4 (columns 2 and 4 of Tables 3.4 and 3.5) are able to balance the ex ante variables across SEO firms and their matches. Furthermore, F statistics (Panel B) show that SEO firms and their counterparts are simultaneously matched in every quartile.

Furthermore, comparison of Model 2 and Model 4 shows the advantage that quarterly returns hold over aggregate returns. Although Model 2 is able to balance past-9-month returns between the two groups, the difference in returns during the latest quarter (QR1) is significantly different between the two groups, with SEO firms significantly outperforming matching firms. The magnitude of difference in QR3, although not statistically significant, is relatively high, with SEO firms underperforming matching firms. Accordingly, Model 2 is not able to capture the dynamic pattern of price

⁴⁰ A 6 percent trimming rule is used in the table. Results using 5 and 7 percent trimming rules are very similar and are available upon request.

run-up right before equity offering that is documented in the capital structure literature (Hovakimian 2003 and Baker and Wurgler 2002). However, because the mismatches in QR1 and QR3 are in different directions, they cancel each other out in the aggregated returns. In other words, even though the aggregate returns are balanced under Model 2, the quarterly returns of the SEO and matching firms are not, with SEO firms typically exhibiting significantly higher returns leading up to the SEO. Since price run-up is an important determinant of equity offering decisions, failure to capture the past dynamic return patterns implies that Model 2 is inadequate in solving the selection problem. I believe Model 4 is a better model, and I use past quarterly returns instead of aggregate returns in the propensity score models during the whole sample period.

I also investigated past returns in earlier quarters. In particular, the obvious candidate is QR4 (returns in the first quarter of year t-1), since several earlier studies use previous 12-month returns. This variable is excluded for three reasons. First, I find that the coefficient of this variable is not statistically significant in the propensity score model, so it seems that QR4 does not significantly affect a firm's SEO decisions. Second, even without being added to the propensity score model, this variable is balanced between the SEO and matching groups (Table 3.5), which further shows that the three quarterly return variables used have already captured the risk factors carried in QR4. Third, use of QR4 significantly reduces the sample size. The same story holds for returns during earlier periods. Therefore, the model includes only the latest three quarters' returns.

I further investigated whether other accounting variables may affect both a firm's decision to issue seasoned equities and *ex post* stock returns. Accounting variables were

considered that have been mentioned in the capital structure literature, such as tax rate (TR), target capital structure (AD), growth rate (BCHG1), profitability (PROAA), and internal fund adequacy (IFCAPX, FFAR, and Free Cash Flow Index). I conducted a step-wise test to examine the marginal impact of these variables and found that coefficients of these variables are not stable, varying from year to year, with most of them statistically insignificant and sometimes showing the wrong sign. Therefore, it seems that the roles played by these accounting variables in firms' SEO decisions are ambiguous. Furthermore, inclusion of these variables significantly reduces the sample size, and thus I do not include any of the accounting variables.

3.6 Results Over the Period 1986-1998

To conduct a broad investigation of long-run stock performance after SEOs, I apply the propensity score matching method for each year from 1986 to 1998. I report *ex ante* balance in Table 3.6 and long-run *ex post* returns in Table 3.7. To facilitate direct comparison, I recalculate 2-dimensional and 3-dimensional matching using the same sample.⁴¹ I investigate only the event-time, equally weighted buy-and-hold returns for three years after issuance, since this method usually shows the most persistent underperformance (e.g., in Brav, Geczy and Gompers 2000). All models use the same five variables – Mcap, BM, QR1, QR2 and QR3, and the propensity score functional forms are chosen based on *ex ante* variable balance. As discussed above, if linear terms do not lead to balance on *ex ante* variables, higher order or interaction terms constructed from the above five variables are added until good *ex ante* balance is achieved between the SEO and matching firms. Therefore, although the same set of *ex ante* variables is

⁴¹ Sample 3 differs from sample 2 in that it drops firms with outliers. Sample 4 is constructed using 6 percent trimming.

used for all years, the functional forms of the propensity score models differ from year to year. This is because variations in the macro economy and the business environment cause the *ex ante* variables to follow different distributions over time, thus requiring different functional forms to fit the data.

Table 3.6 presents balancing tests over the period from 1986 to 1998. Panel A shows the aggregate mean differences between SEO and matching firms for each *ex ante* variable.⁴² Standard errors, shown in brackets, are obtained from bootstrap. Note that all of the *ex ante* variables are very well balanced between SEO and matching firms. Panel B reports joint *F* tests over each quartile for each year, and none of the *F*-statistics is above the critical level. Therefore, the propensity score method is able to balance *ex ante* variables properly, implying that the major risk factors have been sufficiently controlled for. This is an important improvement upon traditional matching methods.

The purpose of adding the common support constraint is to avoid comparing firms that are not truly comparable, which most likely occurs among those small SEO firms with unusually high probabilities of offering (at the right tail of propensity score distribution). Notice that *ex ante* balance without trimming (sample 3) is as good as that under the 6 percent trimming rule (sample 4).⁴³ This shows that the matching procedure has yielded high quality matches even for the small SEO firms at the right tail.

 $^{^{42}}$ I conduct *t* and *F* tests for each year. Only aggregate mean difference results are reported in Panel A because of space limit. The *ex ante* variables are well balanced in terms of means for each year during the sample period. Results on a yearly basis are available upon request.

⁴³ Other trimming rules do not lead to any discernable changes in matching quality *ex ante*.

Panel A: Mean tests (T-tests)

1998

	BM	Мсар	Qr1	Qr2	Qr3	Previous Returns
2-Dimensional Matching	0.0033	_33 250				23 54%
Sample (3): without common support	[0 0012]**	-33,239 [0,727]**				[2] 160/]**
Sample (5). wanout common support	[0.0013]**	[9,737]**				[2.10/0]**
3-Dimensional Matching	-0.0251	-137,519				-6.33%
Sample (3): without common support	[0.0081]**	[45,094]**				[1.42%]**
Propensity Score Matching						
Sample (3): without common support	0.0111	-42,133	0 73%	-0.82%	-0.28%	0.13%
	[0 0096]	[72 909]	[0 74%]	[0 77%]	[1.04%]	[1 45%]
	[0.0090]	[12,909]	[0.7470]	[0.7770]	[1.0470]	[1.4570]
Sample (4): with common support	0.0160	-38 127	0.46%	-0.41%	0.51%	0.81%
Sample (1): Wat common support	[0.0105]	[84 110]	[0.68%]	[0 74%]	[0.84%]	[1 15%]
Panel B: Joint tests (F-tests)	First	Second	Third	Fourth	Critical	-
i ear	Quartile	Quartile	Quartile	Quartile	value	
1986	0.90	1.95	0.45	1.18	2.84	
1987	0.38	1.43	0.09	0.36	3.07	
1988	0.94	1.42	0.13	1.29	4.35	
1989	0.41	0.33	1.29	1.16	3.24	
1990	1.35	0.36	1.87	1.18	3.29	
1991	0.29	1.88	0.10	0.85	2.79	
1992	0.36	0.33	1.43	1.08	2.79	
1993	0.16	0.29	0.99	1.17	2.79	
1994	0.58	0.93	1.37	0.32	2.98	
1995	1.03	1.57	0.94	0.41	2.79	
1996	0.67	0.65	0.62	0.35	2.76	
1997	1.84	1.93	1.68	0.49	2.79	

Table 3.6: Balancing Tests of Propensity Score Matching - Difference in *ex ante* Variables (1986 - 1998)

1.15

2.14

0.99

2.98

1.13

Table 3.7 presents the event-time equal-weighting difference in 3-year buy-andhold returns between SEO and matching firms during the period from 1986 to 1998 without trimming (sample 3) and with 6 percent trimming (sample 4) under the propensity score matching method. In order to facilitate direct comparison so that the results are not driven by deleted outliers, I also recalculate the equal-weighting 3-year

	Raw 36-month returns		Difference in 36-month returns between SEO firms and matching firms						
Year			2 Dimensional Matching		3 Dimension	al Matching	Propensity Score Matching		
	w/o common support	w/ common support	w/o common support	w/ common support	w/o common support	w/ common support	w/o common support	w/ common support	
	Sample (3)	Sample (4)	Sample (3)	Sample (4)	Sample (3)	Sample (4)	Sample (3)	Sample (4)	
1986	12.17%	12.17%	-0.89%	-0.34%	-7.17%	-5.15%	0.32%	-0.46%	
1987	-9.94%	-9.83%	-12.53%	-15.55%	-7.69%	-8.47%	-10.59%	-12.42%	
1988	18.36%	16.20%	10.94%	10.96%	1.57%	-0.14%	22.60%	22.71%	
1989	6.33%	5.89%	-38.85%	-37.78%	-32.05%	-32.21%	-22.58%	-24.43%	
1990	26.82%	21.60%	-29.59%	-28.41%	-39.04%	-40.86%	-18.28%	-17.12%	
1991	35.24%	40.45%	-5.60%	-3.65%	-21.52%	-18.67%	-6.84%	-3.19%	
1992	41.64%	45.01%	-14.84%	-13.48%	-8.33%	-9.19%	2.63%	1.87%	
1993	28.54%	30.10%	-37.32%	-37.85%	-26.02%	-23.52%	-7.38%	-7.86%	
1994	52.21%	54.11%	-6.37%	-11.61%	-10.63%	-11.47%	-0.40%	-2.38%	
1995	41.76%	39.79%	-16.46%	-13.04%	-5.91%	-9.01%	8.78%	7.47%	
1996	6.03%	8.92%	-30.64%	-33.13%	-26.56%	-26.75%	-5.77%	-5.60%	
1997	35.26%	27.82%	-0.76%	-15.03%	-5.31%	-12.45%	-16.73%	-15.80%	
1998	9.09%	8.42%	21.21%	19.06%	-20.96%	-21.78%	14.21%	12.43%	
Overall	25.28%	25.91%	-13.89%	-15.36%	-16.22%	-16.82%	-3.63%	-3.75%	
S.E.	[2.85%]	[2.64%]	[4.54%]	[4.56%]	[2.84%]	[2.76%]	[3.69%]	[3.61%]	

Table 3.7: Difference in 3-Year Returns (1986 - 1998)

buy-and-hold returns using traditional matching methods. Standard errors are obtained from bootstrap as well. Notice that 2-dimensional and 3-dimensional matching still yields significant SEO underperformance, with the magnitude ranging from -13.89 to -16.82 percent and standard errors from 2.76 to 4.56 percent. However, with exactly the same sample, propensity score matching shows minimal underperformance. Over the whole sample period, SEO firms underperform matching firms by less than 4 percent, with standard errors around 4 percent. Since good matches were found for those small firms at the tails, as shown in Table 3.6, it is not surprising that the estimated abnormal returns are quite similar between sample 3 and sample 4. This shows that the results are quite robust and are not driven by the few observations at the tails.

In summary, Table 3.7 shows that the differences in the estimated magnitude of long-run abnormal returns after SEOs are not caused by sample differences, but purely by different matching methods. Since the propensity score matching method is able to achieve a much better *ex ante* match, I argue that the results are more reliable. These results indicate that there is no significant long-term abnormal performance following seasoned equity offerings.⁴⁴

3.7 Conclusion

Prior research has widely documented abnormal stock returns following SEOs, which presents a major challenge to the efficient market hypothesis. However, these results may be driven by the matching methods commonly used in this type of research. Traditional matching methods typically match firms in terms of size, book-to-market ratio and, sometimes, past returns – the factors that are generally found to affect stock returns. However, these methods are not able to achieve good matches on all three dimensions simultaneously, which implies that they may not be able to control properly for risk factors. Therefore, this evidence against the efficient market hypothesis may be questionable.

⁴⁴ I also tried aggregate past 9-month returns instead of quarterly returns. As expected, *ex ante* balance is not as good as that from quarterly returns, but it is still significantly better than that from traditional methods. Use of aggregate returns leads to *ex post* 36-month abnormal returns at around 6 percent, which is slightly higher than that from the use of quarterly returns but still statistically insignificant.

To overcome this imperfect matching problem, this study implements the propensity score matching method. This method not only enables matching on multiple dimensions concurrently, but it also provides a venue to investigate and incorporate other variables that may affect both a firm's SEO decision and its subsequent long-run stock performance. Hence, I am able to address the sample selection problem and the "dimensionality curse."

Using SEO observations during the period from 1986 to 1998, I find that size, book-to-market ratio and past returns are important SEO decision factors, while other variables have only marginal impact. It is also more appropriate to use past quarterly returns than aggregate returns because returns over several shorter intervals are able to better capture dynamic return patterns. I estimate propensity scores along five dimensions: size, book-to-market ratio, and three past quarterly returns. This study only investigates the event-time, equally-weighted buy-and-hold returns three years after issuance, since these methods usually show the most persistent underperformance. I find that the abnormal returns after SEOs evaporate with propensity score matching; the estimated underperformance is less than 4 percent, with standard errors around 4 percent. In contrast, significant underperformance is shown using traditional methods for the same sample. Hence, low returns may be a prevalent problem facing all small stocks, and the widely documented abnormal returns after SEOs may result from inadequate control for risk factors. Once risk factors are properly controlled for, the abnormal returns cease to exist. These results are consistent with the arguments put forward by Brav, Geczy and Gompers (2000) and Eckbo, Masulis and Norli (2000).

The important contribution of this paper is an implementation of the propensity score matching method. This method not only overcomes the "dimensionality curse" that troubles traditional matching methods, but it also provides a convenient way to investigate impact from additional risk factors. Therefore, it is a method that may be very useful and can be generally applied to long-run stock performance studies in the future.

CHAPTER 4

ASSESSING THE ABILITY OF MATCHING TO ADDRESS ATTRITION BIAS USING DATA FROM THE RAND HEALTH EXPERIMENT

4.1 Introduction

Attrition has long attracted the attention of econometricians and labor economists. As is well known, attrition can create a sample selection problem in any data set. Further, since attrition can occur in a social experiment, there is generally no randomization strategy that will eliminate this bias in such an experiment. Matching, meanwhile, is becoming an increasingly popular tool in economics for dealing with selection issues.

Recently economists have been evaluating matching methods by trying to mimic the experimental evidence in social experiments where individuals were randomly assigned to job training. Heckman, Ichimura and Todd (1997, 1998) and Heckman, Ichimura, Smith and Todd (1998) present evidence of mixed success with the Job Training Partnership Act data. Dehejia and Wahba (1999) argue that matching can adequately address selection issues in the National Supported Work Demonstration data, but Smith and Todd (2003) take a more pessimistic view. It is important to have evidence on how well matching works for dealing with selection issues outside of job training, in part because selection issues in job training are very difficult to deal with (LaLonde 1986).

This study uses data from the well-known RAND Health Experiment. I artificially introduce non-random attrition in the data, and then use matching to correct for the resulting selectivity bias. Since the data are from a social experiment, I can assess the efficacy of matching to deal with the biases due to attrition. Specifically, I consider the effect of being in a health insurance plan with a 95 percent co-payment (hereafter the 95% plan) compared with being in a plan with no co-payment (hereafter the free plan). Individuals are followed for 3 to 5 years, and attrition can occur in any year. The artificial attrition takes the form of excluding in one group 50% of the individuals with expenditures above the median (for the group) in a given year. Matching is used to address the bias introduced by this attrition. I consider two treatment effects: (i) the treatment effect for the remaining sample, i.e the sample with attrition constitutes the treatment group and the sample without attrition constitutes the comparison group; (ii) the treatment effect for the whole sample before attrition, i.e the sample without attrition constitutes the treatment group and the sample with attrition constitutes the comparison group. Since matching may perform differently for different components of health expenditures, I consider the problem of estimating the experimental effect of the plans on (i) expenditures on chronic conditions and (ii) total health care expenditures.

Finally, this study investigates a modification of standard matching procedures that seems particularly applicable to health data but should also have applications outside of health. Essentially, matching proceeds as follows. First I estimate a probit model for attrition for the sample in which I introduce attrition. Since attrition is positively correlated with expenditures in year 2, we need to include variables that are correlated with year 2 expenditures. We also need variables having the same meaning for individuals in the plan that has attrition and the plan that does not, so that propensity scores can be compared across the two groups. A natural variable to predict year 2 expenditure is year 1 expenditure. However, year 1 expenditure is post-baseline, and it is not comparable across individuals in different plans, because I know from the experimental evidence that individuals in the 95% plan have lower expenditures than those in the free plan. (This, of course, simply reflects the idea that generally I do not want to use post-treatment variables if they are affected by treatment.) One solution is to use pre-baseline expenditures, since these are comparable across individuals in the two plans. However, in the data pre-experiment health expenditures are self-reported, and the unreliability of this measure has been widely documented. As an alternative to using prebaseline data in the propensity score, I proceed as follows. I assume that the relative magnitude of health expenditures within a plan is not affected by the plan's generosity, so individuals who have relatively high expenditures in year 1 in the free plan are comparable to those who have relatively high expenditures in year 1 in the 95% plan because the plan assignment was a randomized process. Thus I use the individual's health expenditure ranking within his group in year 1, normalized by sample size, as a conditioning variable.

I find that matching performs moderately well. The empirical results can be summarized as follows. First, matching does a better job in general dealing with attrition in chronic care expenditures than in total expenditures. Second, matching does a better job of estimating the first policy effect described above, the treatment effect for the remaining sample, than for the second policy effect, the treatment effect for the whole sample. This latter treatment effect is more difficult to estimate since the sample size of
the comparison group is relatively small and there are fewer high-expenditure individuals in the comparison sample. Third, the modification of the matching procedure using health expenditure ranking in the previous year (as opposed to pre-baseline expenditure) seems to work relatively well.

The rest of the chapter proceeds as follows. Section 4.2 provides the background for the RAND data used in this study. Section 4.3 describes the method of introducing simulation bias into the RAND data. Section 4.4 contains the empirical results and Section 4.5 concludes the paper.

4.2 Rand Health Insurance Experiment

4.2.1 Overview

The Rand Corporation conducted a health insurance experiment, the Rand Health Insurance Experiment (HIE), from 1974 to 1982 at six sites across the United States: Dayton, Ohio; Seattle, Washington; Fitchburg and Franklin counties, Massachusetts; and Charleston and Georgetown counties, South Carolina. In each site, the experiment lasted for 3 or 5 years. The experiment was designed to assess how different rules for sharing health care costs between individuals and health insurance companies would affect utilization of services, the quality of health care provided, patient satisfaction and health status. A side issue addressed by this experiment was to study two alternative modes of delivery – fee for service (FFS) plans versus health maintenance organizations (HMO).

The experimental samples (with the family as the unit of sampling) were randomly chosen from each site subject to eligibility criteria (see *User's Guide to HIE Data*, Rand Corporation). The participating families then were randomly assigned to one of up to fifteen experimental FFS health insurance plans (distinguished by coinsurance

rates and maximum out-of-pocket expenditures) in each location, and one HMO in Seattle.

4.2.2 Sample Description and the Design of the Experiment

To avoid the complications of dealing with site effects, I focus on Seattle, which has the largest number of enrollees among all sites. Table 4.1, panel A presents the sample selection criteria. There were 3,351 insured enrollees in Seattle.⁴⁵ To simplify the analysis, I constrain the sample to consist of adults⁴⁶ only, reducing the sample to 2,435. I next discard the nonrandom individuals who were insured by a staff model HMO, the Group Health Cooperative of Puget Sound (GHC), before the experiment. The remaining 1,833 enrollees were randomly assigned to 11 FFS health insurance plans and GHC (the HMO) by the experiment. I next delete the 888 enrollees assigned to GHC because categorized health expenditures are not available, and the analysis looks at chronic care expenditures. Among the remaining of 945 enrollees, 393 were assigned almost evenly to 9 FFS plans. These 393 individuals are deleted because the sample sizes for these plans are too small to estimate the parameters of interest. The final sample consists of 552 individuals assigned to The Plan 13 (the 95% plan).

⁴⁵ HIE over sampled low-income individuals in all sites except Seattle.

⁴⁶ HIE defines adults to be enrollees who were 14 years of age or older at the time of exit and who completed at least one of the enrollment and exit medical history questionnaires.

Panel A: Sample Selection	
Seattle insured enrollees (not including adjunct enrollees)	3351
After constraining to be adults (were under 14 years of age or older at the time of exit and completed at least one of the enrollment and exit medical history questionnaires)	2435
After excluding the nonrandom sample from GHC, the HMO participating in the HIE in Seattle.	1833
After excluding the enrollees assigned to GHC during the experiment	945
After deleting the enrollees scattered in 9 FFS insurance plans	552
Final sample composition	
Plan 13 (95% Plan)	218
Plan 11 (Free Plan)	334
Total	552

Panel B: Two Samples Used in this Study

	Plan 11 (Free Plan)	Plan 13 (95% Plan)
By random assignment	334	218
After constraining socioeconomic variables used in the propensity score to be available	272	192
After constraining year 1 health expenditures to be available	263	177
After constraining year 1 and year 2 health expenditures to be available: <i>Sample I</i>	233	153
After constraining pre-experiment self-reported health expenditures and year 1 and year 2 health expenditures to be available: <i>Sample II</i>	195	118

Table 4.1: Sample Selection Criteria and Resulting Sample Size

In Plan 11, participants paid nothing out-of-pocket for covered services. For Plan 13, participants paid nothing out-of-pocket for covered inpatient services, but paid 95% of covered outpatient services until the deductible limit is met; then the plan paid 100%. The deductible limit is \$150 per person or \$450 per family. To ensure all families drawn by the experiment were equally likely to participate in their plans, the experiment developed a method of side payments called Participation Incentive payments. The Participation Incentive payments were calculated as the maximum loss risked by changing to the experimental plan from existing coverage.

Table 4.1, panel B, describes the construction of the two samples, *Sample I* and *Sample II*, used in this study. Sample I contains all the observations for which the socioeconomic variables used in the propensity score, as well as year 1 and 2 health expenditures, are available. Sample II contains all Sample I observations for which the pre-experiment self-reported health expenditures are also available.

4.3 Methodology

4.3.1 A Simulation Procedure

This section describes the procedure of creating artificial attrition. I introduce attrition into the 95% plan sample on the basis of chronic care expenditures, while the free plan sample remains untouched. I rank individuals in the 95% group by their year 2 chronic condition expenditures. I then randomly exclude half of the individuals above median (i.e., a quarter of the whole 95% sample is deleted). This seemed to us to be a realistic case, representing individuals with high expenditures dropping out of a less generous plan. The goal is to investigate how well matching can adjust for this attrition and produce unbiased results for two treatment effects. The first effect I consider is the

"experimental effect" for the whole sample. That is, I ask the question "What would happen to health care expenditures of those in the free plan if I switched them to the 95% plan?" For this question the experiment gives us the true parameter. The second effect is the "experimental effect" for those remaining in the 95% plan after attrition. That is, "What would have happened to their expenditures if I shifted the remaining members of the 95% plan to the free plan?" In this case the experiment does not give us the true effect, but I can use the experimental data to simulate the true effect so that I have a benchmark.

I start with the second effect, and I call this simulation/estimation procedure *Scenario I*. First I define this problem in a formal evaluation study setting. Staying in the 95% plan is the treatment; the treatment group consists of the remaining 95% plan sample after attrition and the comparison group consists of the whole free plan sample. The outcome is year 2 chronic condition expenditures. The goal is to estimate ATET as defined in Chapter 1. I adopt the following procedure:

- (1) Create the same attrition in the free plan group, deleting half of the individuals above the median in terms of year 2 chronic condition expenditure. Since both the full 95% and the free groups were created by random assignment, the sample difference between the 95% and free plan groups (both with simulated attrition) is considered to be the "true effect" for the 95% sample after attrition.
- (2) Estimate the probability of attrition within the 95% plan using a probit model. Individuals who are deleted by attrition is coded as 1 and otherwise coded as 0. I will discuss the conditioning variables in the next section.

- (3) Impute the probability of attrition for each individual in the free plan by using the free plan data and the parameters estimated in step (2) from the 95% plan.
- (4) Implement propensity score matching to estimate the ATET. I will defer the discussion of matching methods to the next section.
- (5) Simulate the "true experimental effect". The samples follow the usual distribution of health expenses, where around one-third of the enrollees use no medical service and a small percent of users account for half or more than half of the total expenses. This "true experimental effect" varies considerably depending on the random sample drawn. To solve this problem, repeat step (1) through step (4) 50 times. The final estimate of the 'true effect' is the average of true effects from 50 simulations. The final estimated ATET is obtained by averaging the treatment effects from the matching procedure. We want to see how close the average of the matching estimates is to the average of the true effects.
- (6) Obtain the standard errors for the estimates. I draw 100 bootstrap samples. For each bootstrap sample, I repeat step (1) through step (5) and calculate the "true effect" and ATET for each bootstrap sample. I use the standard deviation of the 100 bootstrap estimates as their respective standard error. Note that this is a computationally intensive procedure, since it requires 5000 (50*100) simulations and 5000 matching estimates.

For the equivalent estimates for total care expenditures I follow exactly the same procedure, substituting total care expenditures for chronic care expenditure, including ranking the individuals on the basis of their total care expenditures. I next consider the first effect as stated above, and I call this simulation/estimation procedure *Scenario II*. I again create attrition in the 95% plan. For this case the treatment effect is staying in the free plan, so the treatment group consists of the free plan and the remaining 95% plan participants are in the comparison group. The outcome is still year 2 chronic condition expenditures. Due to the problem setting, here I have one of those rare cases where ATET is equivalent to ATE. I replace step (1) above with step (1a):

(1a) The "true experimental effect" comes from the difference between the average expenditures for the full 95% plan sample (before attrition) and the full free plan sample, i.e. the experimental parameter.

I conduct matching correspondingly considering the switching of the treatment and comparison group.

4.3.2 Model Specification and Matching Procedure

This section discusses the conditioning variables in the probit model and the matching procedures used. The baseline questionnaire provides a rich set of socioeconomic variables. Table 4.2 reports the summary statistics of variables used in this study. Due to the random assignment of sample members, all socioeconomic variables are similar between the two plans.⁴⁷ On average the 95% plan participants have lower pre-baseline expenditures than the free plan participants, but the difference is not statistically significant.

⁴⁷ Due to the social experiment setting, there is no need for this study to add a common support constraint because the inequality identification condition of matching has been satisfied by the random assignment.

I choose to use sex, age, marital status, highest grade completed, household income⁴⁸ and a dummy variable for having health insurance at the workplace. In addition, the most important conditioning variable is lagged medical utilization. Conventionally, the conditioning variables should be pre-treatment variables as discussed in Chapter 1. In one specification I use pre-experimental health care spending. However this variable is self-reported and likely to be unreliable, which is a common issue for self-reported health expenditure data. As an alternative, I assume that the plan that the individual is in affects only the level of his health care spending, but not the rank of his expenditures within his plan. I then order the individuals in terms of their health spending in year 1 within their plan and use their rank in their group in year 1 in the propensity score. The basic idea is that I want to match high-expenditure individuals in the 95% plan with high-expenditure individuals in the free plan. If the assumption is reasonable, this year 1 rank variable will allow us to fully utilize the social experimental setting, randomized assignment at baseline and accurate report of in-experiment health expenditure.

⁴⁸ I use the logarithm of family income normalized by family size.

Variable Name	Variable Definition	Sample I		Sample II	
		95% Plan	Free Plan	95% Plan	Free Plan
Sociaeconomic Varial	bles				
Sex		0.50	0.48	0.47	0.47
Age	Age in years	35.20	35.21	35.14	35.46
HGC	Highest grade completed	12.90	12.66	12.90	12.79
Income	logarithm family income normalized by family size	9.52	9.44	9.57	9.47
Work-Ins	=1 if having insurance at the workplace	0.80	0.82	0.84	0.84
MSP	=1 if married, spouse present	0.70	0.67	0.76	0.67
Pre-experiment Healt	h Expenditure				
TotExp0	Total health expenditure in the year before the experiment (self-reported)	-	-	85.59	120.14
Health Expenditures	During the Experiment				
TotExp1	Total health expenditure in year 1	122.34	208.58	99.36	212.48
TotExp2	Total health expenditure in year 2	141.53	175.93	138.96	168.99
Chronic1	Chronic condition health expenditure in year 1	94.77	103.27	103.73	102.23
Chronic2	year 2	102.42	117.74	123.08	116.39

Table 4.2: Variable Definitions and Descriptive Statistics

As discussed in Section 1.5, asymptotically local linear or local cubic regression would be optimal in this case. However, I should also consider the finite properties of the matching estimators. Frölich (2001) suggests that kernel regression is more robust to misspecification in the bandwidth than local linear regression in finite samples, especially when the comparison group is relatively small. The treatment and comparison groups in the study are similar in terms of sample size, raising concern about the local linear regression estimator. Since the migration study in Chapter 2 already indicates overparameterization problem of local cubic regression, I do not implement it for the smaller sample in this study. The matching procedures used in this study are nearest neighbor with replacement, kernel regression and local linear regression.

4.4 Empirical Results

Before examining the empirical results, it is worth discussing when I might expect matching to do well. First, I would expect matching to perform reasonably well when the comparison group is relatively large compared to the treatment group and the comparison group has not lost many high-expenditure people due to attrition. Thus, I would expect matching to do better with scenario I than with scenario II. In scenario I the treatment group (the 95% plan group after attrition) has 115 observations in Sample I and 89 observations in Sample II, while the comparison group (the whole free plan group) has 233 observations in Sample I and 195 observations in Sample II. Scenario II has the treatment and comparison groups in Scenario I switch roles. Second, I would expect matching to perform better when the outcome is defined as expenditures related to chronic conditions than total expenditures. The total expenditures include expenditures related to acute conditions and expenditures related to chronic conditions. Acute condition expenditures are harder to predict by past utilizations due to their randomness. Third, I would expect matching to perform better when the conditioning variables include year 1 health expenditure ranking rather than the pre-experiment expenditure. Year 1 expenditure is the most recent past utilization, so it should have the most up-to-date information about individuals' expenditure patterns. Moreover, pre-experiment health expenditures are self-reported and assumed to be less reliable.

Table 4.3 presents estimates for Scenario I. First consider Panel A, which contains estimates based on Sample I. The propensity score model includes the socioeconomic variables stated in the last section and the rank of year 1 expenditure. In column (1), the "true experimental effect" from the simulations suggests that moving from the 95% plan after the attrition to the free plan would increase chronic care expenditures by \$9.78 in year 2 and would increase total health care expenditures by \$25.46.⁴⁹ From column (2) we see that if we simply compare the 95% plan sample with attrition to the free plan to be \$50.84 and \$80.96 for chronic care expenditures and total expenditures respectively. Thus the attrition bias is \$41.06=\$50.84-\$9.78 for chronic care expenditures and \$55.50=\$80.96-\$25.46 for total expenditures. This is the expected direction of the bias because the 95% plan sample with attrition contains relatively fewer high-expenditure individuals than the 95% plan under random assignment.

⁴⁹ Dollar amounts in year 0,1, and 2 are in 1979,1980 and 1981 dollars respectively.

Panel A: Sample I, Conditioning Variable = Order of Year 1 Expenditures				
	(1)	(2)	(3)	(4)
	True Parameter	Sample Diff	Kernel 15%	Linear 15%
Chronic Condition Expenditure STD (100 replications)	-9.78 (21.53)	-50.84 (25.29)	-18.88 (17.94)	-21.48 (17.94)
Total Expenditure STD (100 replications)	-25.46 (74.86)	-80.96 (81.24)	-70.23 (77.05)	-70.67 (75.28)

	True Parameter	Sample Diff	Kernel 15%	Linear 15%
Chronic Condition Expenditure STD (100 replications)	6.42 (21.97)	-33.17 (26.42)	-1.37 (22.96)	-3.65 (28.11)
Total Expenditure STD (100 replications)	-21.89 (87.47)	-73.80 (96.18)	-58.67 (81.04)	-58.51 (1346.04)

Panel C: Sample II, Conditioning Variable = Pre-experiment Expenditures

	True Parameter	Sample Diff	Kernel 15%	Linear 15%
Chronic Condition Expenditure	6.42	-33.17	-15.12	-16.68
STD (100 replications)	(21.97)	(26.42)	(24.20)	(24.19)
Total Expenditure	-21.89	-73.80	-68.67	-68.05
STD (100 replications)	(87.47)	(96.18)	(93.30)	(91.50)

Table 4.3: Simulating Attrition in the 95% Coinsurance Plan Group (Scenario I) 95% Plan Expenditure - Counterfactual from Free Plan Expenditure (Treatment Group = Remaining 95% Plan Group after the Simulated Attrition) In Table 4.3 columns (3) and (4) of row 1, I estimate the true effect of switching these individuals to the free plan to be \$18.88 and \$21.48 on chronic care expenditures under kernel and local linear regression matching respectively. Both of these estimates are reasonably close to the true effect of \$9.78. Row 2 indicates that kernel and local linear regression matching predicts the true effect to be \$70.23 and \$70.67 on total health care expenditures, both of which only slightly reduce the selection bias.

Table 4.3, Panel B, uses the same conditioning variables as in Panel A when I switch to sample II. Combining the results in columns (1) and (2), I see that the bias from ignoring attrition is \$39.59=\$33.17- (-\$6.42) and \$51.91=\$73.80- \$21.89 for the chronic care and total expenditures respectively. Again we see that matching is more successful in terms of correcting the selection bias when the outcome is defined as chronic condition expenditures rather than total expenditures. Note that the standard error on the local linear estimate has "blown up." This confirms that local linear regression may perform poorly when the comparison group is not large relative to the treatment group. Moreover, I will see this phenomenon again in Table 4.4 below. For the sample sizes, it seems better to use kernel matching.

Finally, conside Panel C of Table 4.3, where I replace year 1 health expenditure ranking with pre-experiment expenditures in the propensity score. The estimation is based on Sample II, which includes Sample I respondents for whom pre-experiment expenditures are available. The true effects are, of course, the same as in Panel B since the sample is the same. Overall matching estimators do worse in Panel C than in Panel B, which is not a surprise for the reasons I discussed above.

Table 4.4 contains results from simulation/estimation for Scenario II, which asks what would happen to health care expenditures of those in the free plan if we switched them to the 95% plan. As stated above, Scenario II poses a more difficult setting for matching than Scenario I. Table 4.4 is structured similarly to Table 4.3 except that the treatment and comparison groups are switched. In Scenario II, not only the treatment group (full free plan sample) is larger then the comparison group (95% plan sample after attrition), but also the comparison group has relatively smaller sample variation due to attrition. As discussed before, the samples in this study follow the usual distribution of health expenses, where more than one-third of the enrollees use no medical service. Thus much of the sample variation comes from individuals with expenditures above median. When attrition was created by randomly exclude half of the individuals above median expenditure, a significant amount of sample variation has lost. Overall matching estimators do worse for each corresponding case than in Table 4.3. This result indicates that it is important to have a relatively large comparison group for matching to succeed. Further, the sample variation in the comparison group certainly plays a crucial role even when I do not have a common support problem in this study.

Panel A: Sample I, Conditioning Variable = Order of Year 1 Expenditures					
	(1)	(2)	(3)	(4)	
	True Parameter	Sample Diff	Kernel 15%	Linear 15%	
Chronic Condition Expenditure	15.33	50.84	31.64	7.11	
STD (100 replications)	-	(27.33)	(29.23)	(30.18)	
Total Expenditure	34.39	80.96	78.09	73.06	
STD (100 replications)	-	(76.61)	(84.47)	(112.49)	
Panel B: Sampl	e II, Conditioning V	ariable = Order	of Year 1 Expenditu	ires	
1	, 8		ľ		
	True Parameter	Sample Diff	Kernel 15%	Linear 15%	
Chronic Condition Expenditure	-6.69	33.17	8.62	-21.13	
STD (100 replications)	-	(29.80)	(31.13)	(31.86)	
Total Expenditure	30.02	73.80	73.27	71.61	
STD (100 replications)	-	(89.69)	(112.58)	(131.01)	
Panel C: Sample II	, Conditioning Var	iable = Pre-experi	ment Health Expen	ditures	
`					
	True Parameter	Sample Diff	Kernel 15%	Linear 15%	
Chronic Condition Expenditure	-6.69	33.17	12.52	-30.80	
STD (100 replications)	-	(29.80)	(29.81)	(29.64)	
Total Expenditure	30.02	73 80	71 49	61 99	
STD (100 replications)	-	(89.69)	(102.97)	(103.53)	

Table 4.4: Simulating Attrition in the 95% Coinsurance Plan Group (Scenario II) Free Plan Expenditure - Counterfactual from Remaining 95% Plan Expenditure (Treatment Group = Whole Free Plan Group)

4.5 Conclusion

Attrition has long attracted the attention of econometricians and statisticians, since it can create a sample selection problem in any data set. In particular, it can contaminate a social experiment. Matching is becoming an increasingly popular tool in economics for dealing with such selection issues. This study uses data from the RAND Health Insurance Experiment to assess the ability of matching to deal with attrition. I consider the average difference in health expenditures between two insurance plans. I artificially introduce non-random attrition in one of the plans, and then use matching to correct for the resulting selection bias. Since the data are from a social experiment, I know the true difference in expenditures between the plans, and thus we can assess the efficacy of matching for dealing with attrition.

I modify standard matching procedures as follows: consider an experiment that lasts 2 years after the baseline. I introduce attrition in year 2 that is positively correlated with expenditures in year 2. For the matching to work, we need variables in the propensity score that are correlated with year 2 expenditures and have the same meaning for individuals in the two plans. A natural variable for predicting year 2 expenditures is year 1 expenditures. However, year 1 expenditures are post-baseline and are not comparable across individuals in the two plans because of treatment effects. Further, preexperiment expenditures are self-reported and the unreliability of such measure has been widely documented. To solve this problem, I assume that the relative magnitude of health expenditures within a plan is not affected by the plan's generosity, and thus individuals who have relatively high expenditures in year 1 in one plan are comparable to those who have relatively high expenditures in year 1 in the other plan. This allows us to match in year 2 on the basis of the individual's health expenditure ranking *within* his group in year 1.

I find that matching performs moderately well. The empirical results can be summarized as follows. First, matching does a better job in general dealing with attrition when I analyze chronic care expenditures rather than total expenditures. Second, matching does a better job when the treatment group is relatively small compared to the comparison group. Third, the modification of the matching procedure using health expenditure ranking in year 1 seems to work relatively well.

APPENDIX A

COMMON SUPPORT

As discussed in Section 1.7, the purpose of adding a common support constraint is to satisfy the identification condition P(D = 1|X) < 1. We know that, given properly specified P(X), P(D = 1|P(X)) < 1 ensures P(D = 1|X) < 1, so the common support constraint is added upon P(X). We follow the procedure proposed by Heckman, Ichimura and Todd (1997) as specified below.

Let \hat{S}_1 (\hat{S}_0) denote the estimated smoothed support of propensity scores for the treatment group (comparison group).

Step 1. Find the region over which both the participants' and non-participants' propensity score densities are positive:

$$\hat{S}_{10} = \{ p \in \hat{S}_1 \cap \hat{S}_0 : \hat{f}(p | D = 1) > 0 \text{ and } \hat{f}(p | D = 0) > 0 \}$$
(A1)

This is the overlap of the two supports. This step deletes the treatment group members with no counterparts in terms of P(X) in the comparison group.

Step 2. Find the region \hat{S}_q over which both treatment group and comparison group propensity score densities are *strictly* positive, above a "trimming level" c_q .

$$S_{q} = \{ p \in \hat{S}_{10} : \hat{f}(p | D = 1) > c_{q} \text{ and } \hat{f}(p | D = 0) > c_{q}$$
(A2)

where c_q satisfies

$$\sup_{c_q} \frac{1}{2N} \sum_{\{i \in \overline{I}_1\}} \{ l(\hat{f}(p_i | D = 1) < c_q) + l(\hat{f}(p_i | D = 0) < c_q) \} \le q$$
(A3)

where \bar{I}_1 is the set of observed values of estimated propensity scores from the treatment group that lie in \hat{S}_{10} and N is the number of elements in the set \bar{I}_1 . Here the arguments, p_i , in both indicator functions are treatment group members' estimated propensity scores which fall inside \hat{S}_{10} . This step deletes the treatment group members with few counterparts in terms of P(X) in the comparison group.

The above procedure is designed to trim q percent to 2q percent of participants. To test the sensitivity of the matching estimator to common support constraints, we specify different trimming levels in the migration and the SEO studies.

APPENDIX B

THREE-STEP METHOD FOR CHOOSING THE NUMBER OF BOOTSTRAP REPETITIONS

Andrews and Buchinsky (2000, 2001) propose a three-step method for choosing the number of bootstrap repetitions. We follow their procedure to set the proper number of bootstrap repetitions to calculate the standard errors for each parameter we estimate. The following is a special case in Andrews and Buchinsky (2001).

We first define the notation following Andrews and Buchinsky (2001). *B* is the number of repetitions, and *pdb* denotes the measure of accuracy, which is the percentage deviation of the bootstrap quantity of interest based on bootstrap repetitions from the ideal bootstrap quantity for which $B = \infty$. The magnitude of *B* depends on both the accuracy required and the data. If we require the actual percentage deviation to be less than *pdb* with a specified probability $1-\tau$, then the three-step method takes *pdb* and τ as given and provides a minimum number of repetitions B^* to obtain the desired level of accuracy. We use conventional accuracy level (*pdb*, τ) = (10,0.05).

Step 1. Calculate initial number of repetitions B_1

Set a starting value $\omega_1 = 0.5$ in equation (B1) below. This is the specification for calculating standard errors based on asymptotics, but this method is not sensitive to the starting value.

$$B_{1} = \operatorname{int} \begin{pmatrix} 10,000 * z_{1-\tau/2}^{2} * \omega_{1} \\ pdb^{2} \end{pmatrix}$$
(B1)

where $z_{1-\tau/2}$ is $1-\tau/2$ quantile of standard normal distribution. In our case $B_1 = 193$.

Step 2. Use the bootstrap results $\{\hat{\theta}: \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_{B_1}\}$ to update ω_1 to ω_B :

$$\mu_{B} = \frac{1}{B_{1}} \sum_{r=1}^{B_{1}} \hat{\theta}_{r}$$
(B2)

$$\gamma_{B} = \frac{1}{1 - B_{1}} \sum_{r=1}^{B_{1}} \left(\hat{\theta}_{r} - \mu_{B} \right)^{4} / se_{B}^{4} - 3$$
(B3)

where se_B is the standard deviation of $\{\hat{\theta}: \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_{B_1}\}$.

Then
$$\omega_B = \frac{\left(2 + \gamma_B\right)}{4}$$
 (B4)

Step 3. Calculate B_2 from

$$B_{2} = \inf\left(\frac{10,000 * z_{1-\tau/2}^{2} * \omega_{B}}{pdb^{2}}\right)$$
(B5)

and the minimum number of repetitions $B^* = \max(B_1, B_2)$.

APPENDIX C

OTHER VARIABLES INVESTIGATED IN CHAPTER 3

The following variables were investigated but not included in the final model.

QR4: Buy-and-hold returns during the first quarter of year *t*-1.

BCHG1: Change in net income during the previous year (a predefined variable in Research Insight).

IFCAPX: Internal funding of capital expenditures, defined as capital expenditure divided by the sum of retained earnings plus depreciation and amortization, multiplied by 100 (a predefined variable in Research Insight).

LTDCAP: Long-term debt as a percentage of total capital, defined as long-term debt divided by total invested capital, multiplied by 100 (a predefined variable in Research Insight).¹

AD: Industry-adjusted debt ratio, defined as (LTDCAP – MDLTDCAP)/MDLTDCAP, where MDLTDCAP is the mean industry LTDCAP, which is re-estimated after observations more than three standard deviations from the industry mean are deleted.²

PROAA: Pretax return on average assets (a predefined variable in Research Insight).

¹ Total invested capital is defined as the sum of total equity, long-term debt, minority interest and preferred stock.

² Industries are defined following Fama and French (1997).

FFAR: Funds flow adequacy ratio, defined as net cash flow divided by the sum of capital expenditures plus inventory change plus cash dividends, multiplied by 100 (a predefined variable in Research Insight).

Free Cash Flow Index: Operating cash flow (data 308) divided by capital expenditure (data 128).

TR: Tax rate, defined as income taxes divided by pretax income (a predefined variable in Research Insight).

BIBLIOGRAPHY

- Andrews, D. W. K. and Buchinsky, M. (2000), "A Three-Step Method for Choosing the Number of Bootstrap Repetitions", *Econometrica*, 67, 23-51.
- Andrews, D. W. K. and Buchinsky, M. (2001), "Evaluation of a Three-Step Method for Choosing the Number of Bootstrap Repetitions", *Journal of Econometrics*, 103, 345-386.
- Baker, M. and Wurgler, J. (2002), "Market Timing and Capital Structure", *Journal of Finance*, **57**,1-32.
- Banz, R. (1981), "The Relationship Between Return and Market Value of Common Stocks", *Journal of Financial Economics*, **9**, 3-18.
- Barber, B.M. and Lyon, J.D. (1997), "Detecting Long-run Abnormal Stock Returns: The Empirical Power and Specification of Test Statistics", *Journal of Financial Economics*, 43, 341-372.
- Bartel, A. P. (1979), "The Migration Decision: What Role Does Job Mobility Play?" *American Economic Review*, **69**, 775-786.
- Baumann, R. W. and Reagan, P. B. (2002), "Employment and Earnings of Economically Disadvantaged Whites: Understanding Why Appalachians are Particularly Disadvantaged" (Mimeo, Ohio State University).
- Borjas, G. J., Bronars, S. G. and Trejo, S. J. (1992a), "Assimilation and the Earnings of Young Internal Migrants", *The Review of Economics and Statistics*, **74**, 170-175.
- Borjas, G. J., Bronars, S. G. and Trejo, S. J. (1992b), "Self-selection and Internal Migration in the United States", *Journal of Urban Economics*, **32**, 159-185.
- Brav, A., Gezy, C., and Gompers, P.A. (2000), "Is the Abnormal Return Following Equity Issuance Anomalous?" *Journal of Financial Economics*, **56**, 209-249.
- Brous, P.A., Datar, V., and Kini, O. (2001), "Is the Market Optimistic about the Future Earnings of Seasoned Equity Offering Firms?" *Journal of Financial and Quantitative Analysis*, **36**, 141-168.

- Clarke, J., Dunbar, C., and Kahle, K. (2001), "Long-run Performance and Insider Trading in Completed and Canceled Seasoned Equity Offerings", *Journal of Financial and Quantitative Analysis*, **36**, 415-430.
- Dahl, G. B. (2002), "Mobility and the Returns to Education: Testing a Roy Model with Multiple Markets", *Econometrica*, **70**, 2367-2420.
- Dehejia, R. H. (Forthcoming), "Program Evaluation as a Decision Problem", *Journal of Econometrics*.
- Dehejia, R. H., and Wahba, S. (1999), "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, **94**, 1053-1062.
- Eckbo, B.E., Masulis, R.W., and Norli, O. (2000), "Seasoned Public Offerings: Resolution of the 'New Issues Puzzle'," *Journal of Financial Economics*, 56, 251-291.
- Falaris, E. M. (1987), "A Nested Logit Migration Model with Selectivity", *International Economic Review*, **28**, 429-444.
- Fama, E. (1998), "Market Efficiency, Long-term Returns, and Behavioral Finance", *Journal of Financial Economics*, **49**, 283-306.
- Fama, E., and French, K. (1992), "The Cross-section of Expected Stock Returns", *Journal of Finance*, **47**, 427-465.
- ———. (1993), "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics*, **33**, 3-56.
- ———. (1996), "Multifactor Explanations of Asset Pricing Anomalies", *Journal of Finance*, **51**, 55-84.
- ———. (1997), "Industry Costs of Equity," *Journal of Financial Economics*, **43**, 153-193.
- Fan, J. (1992), "Design Adaptive Nonparametric Regression", Journal of the American Statistical Association, 87, 998-1004.
- Fan, J. and Gijbels, I. (1992), "Variable Bandwidth and Local Regression Smoothers," *Annals of Statistics*, **20**, 2008-2036.

———. (1996), "Local Polynomial Modeling and Its Applications", Monographs on Statistics and Applied Probability 66 (London: Chapman & Hall)

- Frölich, M. (2001), "Nonparametric Covariate Adjustment: Paired-Matching versus Local Polynomial Matching" (Mimeo, University of St. Gallen).
- Gabriel, P. E. and Schmitz, S. (1995), "Favorable Self-selection and the Internal Migration of Young White Males in the United States", *Journal of Human Resources*, Summer 1995, 460-471.
- Glaeser, L. E. and Mare, D. C. (2001), "Cities and Skills", *Journal of Labor Economics*, **19**, 316-342.
- Goss, E. P., and Schoening, N. C. (1984), "Search Time, Unemployment and the Migration Decision", *Journal of Human Resources*, **19**, 570-579.
- Greenwood, M. J. (1997) "Internal Migration in Developed Countries", *Handbook of Population and Family Economics*, vol. 1B, 648-720 (Amsterdam, New York: Elsevier).
- Ham, J. C., Li, X. and Reagan, P. B. (2001) "Matching and Selection Estimates of the Effect of Migration on Wages for Young Men", (Mimeo, Ohio State University).
- Hanushek, E. A. (1973), "Regional Differences in the Structure of Earnings", *Review of Economics and Statistics*, **55**, 204-213.
- Härdle, W. (1990), "Applied Nonparmetric Regression", *Econometric Society Monographs* No. 19 (Cambridge: Cambridge University Press).
- Heckman, J. (1974), "Shadow Prices, Market Wages, and Labor Supply." *Econometrica*, Vol. **42**, 679-694.

. (1979), "Sample Selection as a Specification Error." *Econometrica*, 47, 153-161.

- Heckman, J., Ichimura, H. and Todd, P. (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program", *Review of Economic Studies*, **64**, 605-654.
- Heckman, J., Ichimura, H. and Todd, P. (1998), "Matching as an Econometric Evaluation Estimator", *Review of Economic Studies*, **65**, 261-294.
- Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998), "Characterizing Selection Bias Using Experimental Data", *Econometrica*, **66**, 1017-1098.

- Heckman, J. and Navarro-Lozano, S. (2003), "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models", (wp 9497, NBER).
- Heckman, J. and Robb, R. (1985), "Longitudinal Analysis of Labor Market Data", Econometric Society Monographs series, no. 10 (Cambridge, New York, and Sydney: Cambridge University Press).
- Hovakimian, A., Opler, T., and Titman, S. (2001), "The Debt-equity Choice", *Journal of Financial and Quantitative Analysis*, **36**, 1-24.
- Hovakimian, A. (2003), "The Role of Target Leverage in Security Issues and Repurchases," working paper, Baruch College.
- Hunt, J. C. and Kau, J. B. 1985. "Migration and Wage Growth: A Human Capital Approach", *Southern Economic Journal*, **51**, 697-710.
- Jegadeesh, N. (2000), "Long-term Performance of Seasoned Equity Offerings: Benchmark Errors and Biases in Expectations", *Financial Management*, **29**, 5-30.
- Jung, K., Kim, Y.C., and Stulz, R. (1996), "Timing, Investment Opportunities, Managerial Discretion, and the Security Issue Decision", *Journal of Financial Economics*, 42, 159-185.
- Kahle, K. (2000), "Insider Trading and the Long-run Performance of New Security Issues", *Journal of Corporate Finance*, **6**, 25-54.
- Kennan, J. and Walker, J. (2003), "The Effect of Expected Income on Individual Migration Decisions", (Mimeo, University of Wisconsin).
- Kothari, S.P., and Warner, J. (1997), "Measuring Long Horizon Security Price Performance", *Journal of Financial Economics*, **43**, 301-339.
- Lansing, J. B., and Mueller, E. (1967), "The Geographic Mobility of Labor", Survey Research Center, Ann Arbor, MI.
- Lechner, M. (1999), "Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification", *Journal of Business and Economic Statistics*, January 1999, 74-90.
- Lechner, M. (2000), "An Evaluation of Public-Sector-Sponsored Continuous Vocational Training Programs in East Germany", *Journal of Human Resources*, Spring 2000, 347-375.
- Lee, I., (1997), "Do Firms Knowingly Sell Overvalued Equity?" *Journal of Finance*, **55**, 1439-1466.

- Lee, L. (1978), "Unionism and Relative Wage Rates: A Simultaneous Equations Model with Qualitative Limited Dependent Variables." *International Economic Review*, 19, 415-433.
- Lalonde, R. J. (1986), "Evaluating the Econometric Evaluations of Training Programs Using Experimental Data", *American Economic Review*, **76**, 602-620.
- Linneman, P. and Graves, P. E. (1983), "Migration and Job Change: A Multinomial Logit Approach", *Journal of Urban Economics*, 14, 263-279.
- Loughran, T., and Ritter, J.R. (1995), "The New Issue Puzzle," *Journal of Finance*, **50**, 23-51.
- ———. (2000), "Uniformly Least Powerful Tests of Market Efficiency", *Journal of Financial Economics*, **55**, 361-389.
- McCall, B. P. and McCall, J. J. (1987), "A Sequential Study of Migration and Job Search", *Journal of Labor Economics*, **5**, 452-476.
- Nakosteen, R. A. and Zimmer, M. A. (1980), "Migration and Income: The Question of Self-selection", *Southern Economic Journal*, 46, 840-851.
- Nakosteen, R. A. and Zimmer, M. A. (1982), "The Effects on Earnings of Interregional and Interindustry Migration", *Journal of Regional Science*, **22**, 325-341.
- Newhouse, J. P. and the Insurance Experiment Group (1993), *Free for All? Lessons from the RAND Health Insurance Experiment*, (Harvard University Press).
- Plane, D. A. (1993), "Demographic Influences on Migration", *Regional Studies*, **27**, 375-383.
- Polachek, S. W. and Horvath, F. W. (1977), "A Life Cycle Approach to Migration: Analysis of the Perspicacious Peregrinator", in Ehrenberg, R. G. (ed.) *Research in Labor Economics*, 103-149.
- Raphael, S. and Riker, D. A. (1999), "Geographic Mobility, Race, and Wage Differentials", *Journal of Urban Economics*, **45**, 17-46.
- Robinson, C. and Tomes, N. (1982) "Self-Selection and Interprovincial Migration in Canada", *Canadian Journal of Economics*, **15**, 474-502.
- Rosenbaum, P. R. and Rubin, D. B. (1983), "The Central Role of the Propensity Score in Observational Studies for Casual Effects", *Biometrika*, **70**, 41-55.

- Rosenbaum, P. R. and Rubin, D. B. (1985), "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score", *The American Statistician*, **39**, 33-38.
- Rubin, D. B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Education Psychology*, **66**, 688-701.
- Ruppert, D., Sheather, S. J. and Wand, M. P. (1995), "An Effective Bandwidth Selector for Local Least Squares Regression", *Journal of the American Statistical Association*, **90**, 1257-1270.
- Spiess, D.K, and Affleck-Graves, J. (1995), "Underperformance in Long-run Stock Returns Following Seasoned Equity Offerings", *Journal of Financial Economics*, 38, 243-267.
- Shaw, K. (1991), "The Influence of Human Capital Investment on Migration and Industry Change", *Journal of Regional Science*, **31**, 397-416.
- Silverman, B. W. (1986) "Density Estimation for Statistics and Data Analysis." Monographs on Statistics and Applied Probability 26 (London: Chapman & Hall).
- Sjaastad, L. A. (1962), "The Costs and Returns of Human Migration", *Journal of Political Economy*, **70**, 80-93.
- Smith, J. and Todd, P. (Forthcoming), "Does Matching Address Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*.
- Topel, R. H., and Ward, M. P. (1992) "Job Mobility and the Careers of Young Men", *Quarterly Journal of Economics*, **107**, 439-479.
- Tunali, I. (2000) "Rationality of Migration", International Economic Review, 41, 893-920.
- Willis, R. and Rosen, S. (1979), "Education and Self-selection", *Journal of Political Economy*, **87**, s7-s36.
- Yankow, J. J. (1999), "The Wage Dynamics of Internal Migration", *Eastern Economic Journal*, **25**, 265-278.
- Yankow, J. J. (2003) "Migration, Job Change, and Wage Growth: a New Perspective on the Pecuniary Return to Geographic Mobility", *Journal of Regional Science*, 43, 483-516.