Essays in Applied Microeconomics: Signaling, Marriage Hazard Rates, and Ability Sorting

by

Patrick David Buckley

B.S., United States Military Academy (1993)

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2003

Signature of Author ....
Department of Economics
January 10, 2003

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

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

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

Essays in Applied Microeconomics: Signaling, Marriage Hazard Rates, and Ability Sorting

by

Patrick David Buckley

Submitted to the Department of Economics
on January 10, 2003, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This dissertation consists of three independent empirical examinations of the effects of information and earnings on individual behavior. The first essay examines the effects of information about the quality of previous performance on contemporaneous performance evaluation. I adapt a model of statistical discrimination to incorporate a signal that indicates whether previous performance exceeded some criterion, and I test the model's predictions using a regression discontinuity design and a policy change at the United States Military Academy involving insignias of academic awards on cadet uniforms. I estimate the signaling effects of an award indicating placement on the Dean's List on senior year GPA to be 0.05 grade points. The second essay examines the relationship between male economic conditions and female marriage rates. I overcome problems of endogeneity and spurious correlation by using the oil boom and bust in Texas from 1970 to 1990 to instrument for average male earnings in Texas counties. I estimate that a 10% increase in male earnings results in a 15% increase in the hazard rate of never married young women into marriage, which does not indicate a significantly large contribution of changes in male economic status to the decline in marriage rates over the past several decades. The third essay examines how pay spreads in tournament compensation schemes may induce higher ability individuals to sort themselves into tournaments with higher pay spreads. I exploit variation in the promotion rates of different military occupations resulting from the downsizing of the Army in the 1990s, and I find modest evidence that higher ability soldiers are more likely to reenlist when the effective pay spreads are higher.

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

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

Acknowledgements

I owe a tremendous amount of thanks to many individuals without whose help I could not have completed this dissertation. I appreciate the support of the Department of Social Sciences at the United States Military Academy especially Lieutenant Colonel Michael Meese who made this whole learnig experience possible and who believed in me from the beginning. I am tremendously grateful to my advisors, Josh Angrist and Daron Acemoglu, who did more than advise – they counseled, corrected, encouraged, and mentored me for the past few years. I am a better economist because of their patience and guidance. I also appreciate the helpful recommendations from my friends, colleagues, and instructors at MIT especially David Autor, Hoyt Bleakley, Peter Diamond, Jon Gruber, Melissa Kearney, David Lyle, Robin McKnight, Cindy Perry, David Sims, and Michael Steinberger. Their comments have significantly improved the quality of this research.

I could not have collected the data for this dissertation without the assistance of many individuals. I appreciate the help of Major Donovan Phillips and Jane Lumia of the Office of Policy, Plans, and Analysis at the U.S. Military Academy who assisted me with acquiring the cadet data for Chapter Two; Colonel Casy Wardynski, Major Daniel Olexio, Major Paul Kucik, and especially Staff Sergeant Mark Pettit who provided the data on the enlisted soldiers in the U.S. Army; and Major Robert Campbell from the Department of the Army Headquarters for assisting me with the personnel strength data for the U.S. Army.

Lastly, I am eternally greatful to my wife, Rita, and our sons, Blain and Cole, who have sacrificed a great amount to support me in this endeavor. I appreciate their love and commitment.

Contents

Chapter 1

Introduction

In this dissertation I present three separate empirical examinations of the effects of information and earnings on individual behavior. In Chapter 2, I examine the effects of a binary signal of an individual's quality on the evaluation of that individual's contemporaneous performance using a natural experiment from the United States Military Academy. Chapter 3 examines the relationship between male economic conditions and female marriage rates using the exogenous shock to male earnings resulting from the Texas oil boom and bust in the 1970s and 1980s. Lastly, Chapter 4 examines the effects of pay spreads in tournament compensation schemes on ability sorting – or the retention of high quality workers – using evidence from enlisted personnel in the United States Army.

1.1 Signaling and Performance Evaluation

This chapter examines the effects of information about the quality of previous performance on the evaluation of imperfectly measured contemporaneous performance. I adapt a model of statistical discrimination in the evaluation of current performance to incorporate a signal, such as an award, a credential, or reputation, that indicates whether previous performance on a similar task exceeded some criterion. I test the model's predictions of information updating by exploiting a natural experiment resulting from a policy change at the United States Military Academy at West Point involving insignias of academic awards on cadet uniforms. In 1993 the Academy created an award – a "silver star" - to recognize placement on the Dean's List, or earning a grade point average (GPA) of at least 3.0. Using a regression discontinuity design, I estimate the signaling effects of the silver star on senior year GPA to be approximately 0.05, or one half of a letter grade in one class out of a ten-class academic year. However, as a result of the beneficial signaling effect, individuals may exert effort to earn the award. I examine the effects of differential exertion of effort on estimates of the signaling effect.

1.2 Male Earnings and Female Marriage Rates

This chapter examine the relationship between male economic conditions and female marriage rates. Previous studies have been confounded by problems of joint endogeneity and spurious correlation, and studies that instrument for male earnings have been conducted under the assumption that OLS estimates are biased upwards. However, normal measurement error attenuation and the positive correlation between male and female earnings, which may actually deter marriage, may result in and downward bias of OLS estimates. I use the oil boom and bust in Texas from 1970 to 1990, which represent shocks to earnings in the predominantly male petroleum industry, to instrument for average male earnings in Texas counties. I estimate that a 10% increase in male earnings results in a 15% increase in the hazard rate of never married young women into marriage. While these estimates do indicate a downward bias, they do not indicate a significantly large contribution of changes in male economic status to the decline in marriage rates over the past several decades.

1.3 Tournament Pay Spreads and Ability Sorting

This chapter examines the effects of pay spreads in tournament compensation schemes on ability sorting. The tournament model literature indicates that pay spreads in tournaments may induce ability sorting of individuals whereby higher ability individuals sort themselves into tournaments with higher pay spreads. This phenomenon is of extreme importance to hierarchical organizations, like the military, which have tournament-like promotion systems. Differences in promotion rates for different levels of relative performance create spreads in career pay, and therefore, may affect the types of individuals who self-select into tournaments for further promotion. I exploit variation in the promotion rates of different military occupations resulting from the downsizing and reorganizing of the Army in the 1990s following the end of the Cold War to test whether higher ability soldiers are more likely to reenlist following their second enlistment period - effectively a decision to choose the military as a career - when the effective pay spreads are higher. Additionally, I instrument for the pay spreads using changes in the Congressionally authorized strength levels to account for potential spurious correlation and attenuation bias. Using different measures of ability, I detect a modest affect of pay spreads on ability sorting. Higher ability soldiers are attracted to tournaments with higher pay spreads. As such, efforts to manipulate promotion rates and pay schedules simply to meet personnel strength requirements may affect the ability mix of the retained force.

Chapter 2

Rising Stars: Evidence of Signaling and Information Updating from the U.S. Military Academy

2.1 Introduction

Economists have long recognized the possibility that individual characteristics may provide a signal of imperfectly measured individual productivity through statistical discrimination. Most of the research on statistical discrimination focuses on race and gender as signals of quality.[1] Evidence of such discrimination is important because it suggests that imperfect information can generate wage and income differentials – signaling effects – that persist even in perfectly competitive labor markets.

In addition to generating statistical discrimination on the basis of time-invariant personal characteristics, imperfect information can also lead to the use of past performance to modify estimates of current productivity. For example, if a workers' performances cannot be observed perfectly, an employee who has typically performed well may be promoted in spite of poor contemporaneous performance.[2] In this paper,

[1] See, for example, the seminal papers by Arrow (1972), Phelps (1972), and Aigner and Cain (1977).

[2] This example and the subject of this study are examples of Thorndike's (1920) original definition

I am concerned with estimating signaling effects in an educational setting when there are observable signals of past performance such as awards, credentials, or reputation. In particular, I ask whether students who perform well during previous terms are graded too generously in the current term. This type of statistical discrimination is similar to that based on race and gender in that it leads to systematic differences between productivity and rewards. However, in the academic context of this study, group designation is determined by the individual's previous performance rather than an immutable trait, and the group characteristics can be inferred by the correlation between current and previous performance. In addition to general theoretical interest in the consequences of imperfect information, as study of grading practices is of practical interest to human resource professionals and teachers interested in the consequences of alternative schemes of performance assessment and performance awards.

This study of the halo effect in grading exploits an unusual natural experiment resulting from a policy change at the United States Military Academy at West Point involving insignias of academic awards on cadet uniforms. At the end of the 1992-1993 academic year, the Academy created an award – a "silver star" – to recognize placement on the Dean's List and to be prominently, and compulsorily, worn on the cadets' daily uniforms. Thus, beginning in the Fall 1993, cadets who earned a GPA of 3.0 or higher during the previous year or cumulatively through all previous years wore a silver star. To the instructor, the silver star is as apparent as a cadet's race and gender, and the policy change allows me to test the signaling effect of the silver star on cadets' grades.

A number of previous studies have attempted to estimate the effects of statistical discrimination in the labor market. For example, in a study about the signaling effects of the General Education Development (GED) High School Equivalency Diploma, Tyler, et al. (2000) identify a counterfactual for their treatment groups by exploiting interstate variation in GED passing standards, and they estimate that the signaling

---

of the halo effect – a "marked tendency to think of the person in general as rather good or rather inferior and to color judgements of the [specific performance] by this general feeling." Interestingly, though, in their survey, Balzer and Sulsky (1987) were able to identify 108 operational definitions of halo in dfifty different studies.

effect of the GED may increase the earnings of white high-school dropouts by 10 to 19 percent. Goldin and Rouse (2000) study the effects of using a physical screen behind which musicians audition for positions in major symphony orchestras on the probability that a female would advance through the audition sequence and eventually be hired. Their findings modestly, though not conclusively, suggest that the use of the screen fosters greater impartiality in the hiring process. However, the authors do not specifically distinguish whether this effect is a result either taste discrimination or statistical discrimination. Others have also looked at college grading; for example Seaver and Quarton (1976) examine the effects of placement on the Dean's List on academic performance in subsequent quarters and find a positive effect of the award.[3]

This study improves on these earlier efforts in that the introduction of the silver star at West Point provides a controlled experiment for the study of halo effects. Control is facilitated both by the discontinuous nature of the rule for awarding the star and by the presence of unaffected cohorts. Additionally, the silver star's prominence and the small class sizes at the Academy ensure that the signal is observed by the instructor. This is unlike the case when the instructor is dependent upon a published Dean's List, as in Seaver and Quarton (1976). Using a regression discontinuity design, I estimate the signaling, or halo, effects of the silver star on senior year GPA to be approximately 0.05 grade points. This is equivalent to one half to a full letter grade in one class out of a ten-class academic year. Additionally, the evidence suggests that the signaling effect may become larger as the observed output becomes a nosier indicator of an individual's true productivity. Also, as a result of the beneficial signaling effect, individuals may exert effort to earn the award. I examine the effects of differential exertion of effort on estimates of the signaling effect, and I estimate that for juniors, differential exertion of effort by contaminate estimates of the signaling effect by as much as 0.08 to 0.10 grade points. While these observations occur in an academic context, these findings are commutable to any context in which an evaluator observes

[3]However, Cook and Campbell (1979) claim that the award effects found by Seaver and Quarton are a function of the linear specification of the mapping of previous GPA to current GPA in their regression discontinuity design. By adding a quadratic term, Cook and Campbell demonstrate that these effects disappear nullifying the results.

and noisy measure of one's quality. The most common example of such situations include performance evaluations by a supervisor, which may be influenced by signals of the quality of related performance.[4]

In the sections that follow I develop a theoretical framework of statistical discrimination that incorporates information from a signal of previous performance into the evaluation of current performance (Section 2), and I explain my empirical strategy to test the theoretical predictions of the model (Section 3.) I discuss the empirical results and provide some concluding remarks in Sections 4 and 5, respectively.

2.2 Theoretical Framework

2.2.1 Statistical Discrimination

An individual studies two academic subjects, one in each of two periods 0 and 1. Cognitive requirements for learning the subjects' material are similar so that the individual's levels of understanding in each course, q_i^0 and q_i^1, have a joint normal distribution $N\left(\alpha^0, \alpha^1, \sigma_0^2, \sigma_1^2, \sigma_{01}\right)$ where $\sigma_{01} > 0$, or there is a positive correlation between the levels of understanding. As such, q_i^0 and q_i^1 have the following linear relationship:

$$q_i^0 = \alpha^0 + \frac{\sigma_{01}}{\sigma_1^2}\left(q_i^1 - \alpha^1\right) + \eta_i.$$

To measure comprehension of the material, a teacher administers a test and observes the individual's performance for each time period, s. Because obtaining a perfect measure of true understanding is excessively costly, performance on the test, y_i^s, is a noisy measure of understanding, so that,

$$y_i^s = q_i^s + u_i^s,$$

where u_i^s is independently and normally distributed error term with a mean of zero

[4]Altonji and Pierret (2001) find that employers may discriminate among young workers on the basis of observable characteristics.

and a variance of $\sigma^2_{u^s}$. Thus, $E[y^s_i] = E[q^s_i] = \alpha^s$, the average level of understanding of the subject. The normality of q^s_i results in the following linear relationship:

$$q^s_i = \alpha^s (1 - \gamma^s) + \gamma^s y^s_i + \varepsilon^s_i,$$

where ε^s_i is normally distributed. In other words, true understanding is a weighted average of mean understanding and individual test performance plus some error, and γ^s is the coefficient of determination (r^2) between q^s and y^s and is a measure of the reliability of the test.

$$\gamma^s = \frac{Var(q^s_i)}{Var(q^s_i) + Var(u^s_i)}.$$

As a result of the imprecision of the test, the teacher's objective is to assign a grade, w^s_i, that minimizes the mean squared error, or loss, between the grade and true understanding conditional on the available information for all n students, or.

$$\min \frac{1}{n} \sum_{i=1}^{n} (w^s_i - q^s_i)^2 \mid \Omega^s.$$

In this case the information available about an individual simply equals the individual's performance on the test, or $\Omega^s_i = [y^s_i]$. Thus, the teacher's objective is to assign a grade commensurate with the student's conditional expected understanding,[5] and the grade assignment function is

$$\begin{aligned} w^s_i &= E[q^s_i \mid \Omega^s_i] \\ &= \alpha^s (1 - \gamma^s) + \gamma^s y^s_i. \end{aligned} \tag{2.1}$$

Suppose now that a separate teacher assesses the individual's performance in the second period. The teacher is unable to directly observe y^0_i or w^0_i, but can observe a binary signal indicating whether w^0_i exceeded some criterion, w^*. Individuals for

[5]There is much pedagogical debate as to the purpose and functions of grades and graded exercises. However, I hold that my assumption, as it applies to the *process* of grading, is reasonable – teachers use tests and other graded assignments to assess students' learning and understanding of the course material.

whom $w_i^0 \geq w^*$ are assigned the signal and are designated as being in group A. Similarly, those for whom $w_i^0 < w^*$ are in group B and do not have the signal. In practice, this signal may be an award, an honor, or even the individual's reputation. Thus, with this additional information $\Omega_i^1 = [y_i^1, I(w_i^0 \geq w^*)]$, and the resulting grade assignment function for the second period becomes

$$\begin{aligned} w_i^1 &= E\left[q_i^1 \mid \Omega_i^1\right] \\ &= \begin{cases} \alpha^A\left(1-\gamma^A\right)+\gamma^A y_i^1 & \text{if } w_i^0 \geq w^*\text{, or} \\ \alpha^B\left(1-\gamma^B\right)+\gamma^B y_i^1 & \text{if } w_i^0 < w^*, \end{cases} \end{aligned} \tag{2.2}$$

where

$$\begin{aligned} \alpha^A &= E\left[q_i^1 \mid w_i^0 \geq w^*\right], \\ \alpha^B &= E\left[q_i^1 \mid w_i^0 < w^*\right], \\ \gamma^A &= \frac{Var\left[q_i^1 \mid w_i^0 \geq w^*\right]}{Var\left[q_i^1 \mid w_i^0 \geq w^*\right] + Var\left[u_i^1\right]}, \text{ and} \\ \gamma^B &= \frac{Var\left[q_i^1 \mid w_i^0 < w^*\right]}{Var\left[q_i^1 \mid w_i^0 < w^*\right] + Var\left[u_i^1\right]}. \end{aligned}$$

These grade assignment functions are similar to the wage assignments in the Aigner and Cain (1977) model of statistical discrimination, whereby employers observe a noisy measure of a worker's true productivity and assign wages equal to workers' conditional expected productivities. To minimize the informational loss, employers consider an individual's race or gender and assign wages based upon group characteristics. However, in the academic setting described above, group designation is determined by the individual's previous performance rather than an immutable trait, and the group characteristics can be inferred through the correlation between the levels of understanding for the two subjects.

Because w_i^0 and q_i^1 are positively correlated, the mean understanding, in terms of the second course, of those in group A is greater than the population mean, α^1, while the mean ability of those in group B is less than α^1. In other words, on average, an individual who exceeds some standard of performance in a previous course will have a higher level of understanding in another course than another individual whose previous performance did not exceed the criterion. This is a result of the conditional distribution of q_i^1 for each group. Similarly, the conditional variances of q_i^1 for both groups is smaller than the population variance. Due to the joint normality of q_i^0 and q_i^1, these conditional moments for group A can be expressed as follows:

$$\begin{aligned}
\alpha^A &= E\left[q_i^1 \mid w_i^0 \geq w^*\right] \\
&= \alpha^1 + \frac{\sigma_{01}}{\sigma_\nu}\lambda^A(\theta), \text{ and}
\end{aligned}$$

$$\begin{aligned}
Var\left[q_i^1 \mid w_i^0 \geq w^*\right] &= Var\left[q_i^1 \mid \frac{\nu_i}{\sigma_\nu} \geq \frac{w^* - \alpha^0}{\gamma^0 \sigma_\nu}\right] \\
&= \sigma_1^2\left[1 - \left(\rho_{vq^1}\right)^2 \delta^A(\theta)\right] \\
&= \sigma_1^2\left[1 - \left(\frac{\sigma_{01}}{\sigma_1 \sigma_\nu}\right)^2 \delta^A(\theta)\right],
\end{aligned}$$

where

$$\begin{aligned}
\nu_i &= u_i^0 + \eta_i + \frac{\sigma_{01}}{\sigma_1^2}\left(q_i^1 - \alpha^1\right), \\
\theta &= \frac{w^* - \alpha^0}{\gamma^0 \sigma_\nu}, \\
\lambda^A(\theta) &= \frac{\phi(\theta)}{1 - \Phi(\theta)}, \\
\delta^A(\theta) &= \lambda^A(\theta)\left[\lambda^A(\theta) - \theta\right],
\end{aligned}$$

$\phi(\theta)$ and $\Phi(\theta)$ are the standard normal probability and cumulative density functions, respectively, and ρ represents the correlation coefficient.[6] In other words, the mean level of understanding for those whose previous grade exceeded the award criterion equals the mean level of understanding in the course plus a correction factor based upon the correlation between q_i^0 and q_i^1 and the manner in which the award criterion differentiates the population. The conditional variance for this group equals the

[6]See Johnson and Kotz (1970), p.81, for similar derivations.

population variance discounted by a factor based upon the correlation. Similarly the conditional mean and variance of q_i^1 for group B are

$$\alpha^B = \alpha^1 - \frac{\sigma_{01}}{\sigma_\nu}\lambda^B(\theta) \text{ and}$$

$$Var\left[q_i^1 \mid w_i^0 < w^*\right] = \sigma_1^2\left[1 - \left(\frac{\sigma_{01}}{\sigma_1\sigma_\nu}\right)^2 \delta^B(\theta)\right],$$

where

$$\begin{aligned}
\lambda^B(\theta) &= \frac{\phi(\theta)}{\Phi(\theta)}, \text{ and} \\
\delta^B(\theta) &= \lambda^B(\theta)\left[\lambda^B(\theta) - \theta\right].
\end{aligned}$$

The updating of *ex post* beliefs about an individual's understanding based upon group characteristics as described in equation (2) can result in grade differentials between individuals with equal test performance but who have different levels of previous performance. The grade difference between two individuals i and j whose test performances are equal but who differed in whether their previous grades exceeded the award criterion equals the expected difference in their true levels of understanding. If the award criterion provides an even division of the population, the conditional variances of understanding for the groups will be equal[7] or, equivalently, $\gamma^A = \gamma^B = \gamma$, and the resulting grade differential will be

$$\begin{aligned}
w_i^1 - w_j^1 &= E\left[q_i^1 - q_j^1 \mid y_i^1 = y_j^1, w_i^0 \geq w^*, w_j^0 < w^*\right] \\
&= \alpha^A\left(1-\gamma^A\right) + \gamma^A y_i^1 - \alpha^B\left(1-\gamma^B\right) - \gamma^B y_j^1 \\
&= \left(\alpha^A - \alpha^B\right)(1-\gamma) \qquad (2.3)
\end{aligned}$$

The individual who earned the award in the previous period is awarded a higher grade than the individual who did nor. This positive differential is an example of the

[7]This occurs when $w^* = \alpha^0$, so that the criterion evenly divides individuals into the groups A and B. This simplification is solely for clarification purposes, as it allows for the elimination of terms that otherwise would be nearly equal to zero. Results of this are that $\gamma^A = \gamma^B = \gamma$ and $\theta = 0$.

most commonly accepted definition of economic discrimination – equal output not being rewarded with equal remuneration. However, it is important to note that this effect is not a function of taste discrimination but of imperfect information. It is the signaling effect that results from the incorporation of the signaled information about the quality of previous academic performance.

The following example illustrates this effect in practice. Two students take a test in economics. One student is known by the teacher to have received a high grade in a previous class, while the other student is known to have received a low grade. Both students make an incorrect statement in answering an essay question on the test. The teacher must determine whether the incorrect statement indicates a lack of understanding of the material or is simply a misstatement. The student who received the higher previous grade is more likely to have made a mistake than the student with the lower grade, and the teacher who incorporates this information may award a higher grade to the first student.

This model of information updating resembles many psychological theories pertaining to the cognitive processes by which individuals appraise and evaluate performances by others and to the potential biases in the assessment, or halo-effects.[8] Social cognition and decision analysis theories describe an assimilation effect, by which the evaluation of present performance is biased toward the evaluation of the previous work.[9] Assimilation may be explained by the evaluator's use of heuristic devices of adjustment and anchoring, by which individuals reference evaluations to a benchmark, such as an *a priori* expectation, categorization, or prototyping based upon information about the quality of previous performance.[10] Other similar models, such as those by Slovic and Lichtenstein (1971) and Anderson (1970), posit that evaluators execute a Bayesian-like opinion revision when evaluating an individual's performance. All of the cognitive assimilation theories closely resemble the statistical mechanisms

[8]Ilger, Barnes-Farrell, and McKellin (1993) and Balzer and Sulsky (1992) provide summaries of the performance appraisal literature, while Feldman (1981) provides a survey of the cognitive processes pertaining to performance appraisal.

[9]See Murphy, et al. (1985).

[10]Tversky and Khaneman (1974), p.1128.

proposed above, albeit less quantitatively.

Figure 1a depicts the grade assignment function in which the information from the signal is not incorporated (equation (1)) and for the two groups when the information is incorporated (equation (2)). There are two differences between each group's grade assignment function and the original assignment function. The first is that the mean assigned grade is higher (lower) for those in group A (group B). The second is that the coefficient of determination, or slope of the regression line, is smaller than the original slope for both groups because of the decrease in the variance of q_i^1 within the groups. The grade differential between individuals of different groups for any level of test performance represents the signaling effect of the award as expressed in equation (3).

This framework allows for the detection of the signaling effect through the mapping of the grade assigned in the second period onto the grade for the first period. Due to the correlation between q_i^0 and q_i^1, an individual's expected grade in the second period given his or her grade in the first period is

$$E\left[w_i^1 \mid w_i^0\right] = \begin{cases} \alpha^A - \alpha^0\beta\gamma^A + \beta\gamma^A w_i^0 & \text{if } w_i^0 \geq w^*, \text{ or} \\ \alpha^B - \alpha^0\beta\gamma^B + \beta\gamma^B w_i^0 & \text{if } w_i^0 < w^*, \end{cases} \tag{2.4}$$

where $\beta = \frac{\sigma_{01}}{\sigma_0^2}$. This results in a discontinuity in the mapping of w_i^1 onto w_i^0 at the award criterion, or a difference between the expected grade of an individual whose previous grade equaled the criterion and of an individual whose grade was just below the criterion. This discontinuity is depicted in Figure 1b, and assuming again that $\gamma^A = \gamma^B = \gamma$, the mapping described by equation (4) results in a discontinuity equal to the signaling effect in equation (3), or $\left(\alpha^A - \alpha^B\right)(1-\gamma)$.

A question of interest is how the size of the signaling effect changes when the correlation between the abilities for the different tasks becomes weaker or stronger. In the simple case where $w^* = \alpha^0$, the grade differential between individuals in groups A and B with the same test performance is

$$\left(\alpha^A - \alpha^B\right)(1-\gamma) = \frac{\sigma_{01}}{\sigma_\nu}\left[\lambda^A(0) + \lambda^B(0)\right](1-\gamma)$$

An increase in the correlation, σ_{01}, has two effects. First, it increases the parameter $\frac{\sigma_{01}}{\sigma_\nu}$. (See Appendix A for a proof.) Second, it decreases the value of the coefficient of determination between q_i^1 and y_i^1 for group A, γ^A, due to a decrease in the conditional variance of q_i^1 for each group. Thus, an increase in the correlation between q_i^0 and q_i^1 increases the size of the signaling effect when the signal criterion is close to the mean level of understanding with respect to the subject in the first period.

Also of interest is how the size of the signaling effect changes with the reliability of test performance as an indicator of true comprehension of the subject. In the case where $w^* = \alpha^0$, the only effect of an increase in the variance of u_i^1 is through the coefficient of determination. As $\sigma^2_{u^1}$ increases, γ decreases. Thus, a decrease in the reliability of the test for the current period results in an increase in the size of the signaling effect – evaluators place more weight on the signal. Alternatively, an increase in the variance of u_i^0 results in an increase in σ_ν and an increase in γ due to an increase in the conditional variance of q_i^1 for each group. These two effects result in a decrease in the signaling effect. Thus, a decrease in the reliability of the observed output for the task upon which the signal is based results in a decrease in the size of the signaling effect – evaluators place less weight on the signal.

2.2.2 Effort

Up to this point the model only focuses on the signal's effect on the actions of the teacher or evaluator, while disregarding the potential incentive effects that an award may induce. Assume, however, that grades afford some utility to the student.[11] There may be great benefits from extremely high grades in the form of prestigious awards or scholarships, and, certainly, there may be great disutility in receiving failing grades, such as from expulsion. However, within the range of normal academic performance, assume that individual utility, U_i, is linear in grades and that grades are continuous. Assume also that individual effort, e_i, to improve comprehension decreases utility by

[11]In fact, academic grades are important to cadets at West Point because a cadet's class rank affects his or her assignment of occupational specialty and first duty station upon graduation from the Academy.

an amount ae_i^2. Test performance is a noisy measure of the sum of latent ability to understand and the additional comprehension resulting from exerted effort, so that

$$y_i = q_i + e_i + u_i.$$

Because utility is linear in grades, individuals exert the same amount of effort, $\overline{e}$, regardless of the latent level of comprehension. The teacher's objective is to reward overall comprehension; therefore, the grade assignment function is

$$\begin{aligned} w_i &= E\left[q_i + e_i \mid y_i\right] \\ &= \left(\alpha + \overline{e}\right)(1-\gamma) + \gamma y_i. \end{aligned}$$

The utility maximizing effort level is

$$\overline{e} = \frac{\gamma}{2a}.$$

While an award, such as the silver star, may provide some utility in the form of a sense of accomplishment, a more tangible benefit is its signaling effect during subsequent academic periods. For a uniform benefit from the signaling effect of b, an individual exerts the amount of effort that maximizes his or her expected utility, or the expected grade less the cost of effort plus the probability of receiving the award multiplied by the award's value.

$$\begin{aligned} e_i^* \text{ maximizes } E\left(U_i\right) &= E\left(w_i\right) + b \cdot \Pr\left[w_i \geq w^*\right] - ae_i^2 \\ &= E\left(w_i\right) + b \cdot \Phi\left[\frac{w^* - \left(\alpha + \overline{e}\right)(1-\gamma)}{\gamma\sigma_u} + \frac{\left(q_i + e_i\right)}{\sigma_u}\right] - ae_i^2. \end{aligned}$$

The first order condition for the solution to this maximization problem is

$$\gamma + \frac{b}{\sigma_u}\phi\left(\frac{w^* - \left(\alpha + \overline{e}\right)(1-\gamma)}{\gamma\sigma_u} + \frac{\left(q_i + e_i\right)}{\sigma_u}\right) - 2ae_i = 0. \tag{2.5}$$

Figure 2 graphs the optimal effort levels that solve equation (6) for the following example parameters: $w^* = 3$, $a = 2$, $b = 0.08$, $\gamma = 0.9$, $\sigma_u = 0.1$. The figure demonstrates that while most individuals exert effort $\overline{e}$, individuals whose latent levels of understanding are just below the award criterion will exert more effort to earn the award than will those whose ability is either greater than or far less than the award criterion. In other words, those who have a chance of earning the award will exert additional effort, such as studying more, to do so, while those who are fairly certain to earn or not to earn the award will exert no additional effort. This is similar to the traditional Spence (1973) and Stiglitz (1975) model of signaling whereby "low" type individuals may exert effort to appear to be a "high" type individual to benefit from a wage, or grade, differential between types.

2.3 Empirical Strategy and Data

2.3.1 Identification

The United States Military Academy at West Point, New York, trains and educates cadets for commissioning as officers in the United States Army. The Academy has an accredited four-year academic program that confers a baccalaureate of science degree upon graduates,[12] and despite the regimented lifestyle and military training, a cadet's academic experience is not unlike that of other undergraduate students. The typical class size is between 14 and 18 cadets,[13] there are no auditorium-sized academic classes, and as such, instructors are able and encouraged to know cadets by name. While grades generally are not normalized, GPAs do exhibit a fairly normal distribution with a mean that equals about 2.8 during the freshman year and that increases to about 3.2 for the seniors.

[12]The program at the Academy is designed to last exactly four academic years, thereby making it easy to identify whether an individual is a senior, junior, sophomore, or freshman. A cumulative GPA of 2.0 is required for graduation, and most cadets who fail to maintain academic proficiency are dismissed from the Academy. On extremely rare occasions, involving no more than 20 to 25 individuals per year, cadets may be allowed to graduate in December after having attended the Academy for four and a half years. Even more rare is the case when a cadet leaves the Academy for one year due to poor academic performance, attends another college during that year, and successfully applies for re-admittance.

[13]U.S. Military Academy Admissions Office, www.usma.edu/admissions/prosp_acad.asp.

There are three academic awards that a cadet can earn and which are required to be worn as insignias on the cadet uniform – the Distinguished Cadet Award, whose insignia is a gold star; the Superintendent's Award, which takes the form of a gold wreath; and a silver star for those who achieve placement on the Dean's List. The Academy introduced the silver star at the end of the 1992-1993 academic year, and awarded cadets wore it for the first time during the 1993-1994 academic year. Prior to the introduction of the silver star, the gold star was awarded to cadets who were in the top five percent of their class either based upon their previous academic year's GPAs or their cumulative GPAs. After 1993, the standard for earning the gold star became earning at least a 3.666 yearly or cumulative GPA. The standard for earning the silver star is earning at least a 3.0 yearly or cumulative GPA, while the Superintendent's award, the gold wreath, is awarded for earning at least a 3.0 on *each* of the academic, military, and physical performance scores for the previous year. Earning a gold star or a wreath supersedes being awarded the silver star, but if a cadet earns both a gold star and a wreath, he or she then wears them both. These awards are bestowed, or revoked, each year; therefore, a cadet must re-earn the award(s) each year. Because these awards convey information about a cadet's performance during the previous year and because previous academic performance is likely to be positively correlated with current academic performance, there may be some *ex post* signaling value to earning an award.

The discontinuous nature of the silver star award and the existence of regimes with and without the silver star afford several empirical techniques for identifying the effect of the silver star on cadets' grades. The simplest strategy, a technique that reflects the regression discontinuity design of Thistlewaite and Campbell (1960), involves the mapping of one academic year's GPA onto the previous year's GPA, as described in equation (4), for cadets in the silver star regime. The stochastic specification for this strategy is

$$\begin{aligned} GPA_{it} &= \alpha + \beta \cdot X_i + \\ &\quad \lambda_0 \left(GPA_{it-1} - 3 \right) + \lambda_1 \left(GPA_{it-1} - 3 \right) \cdot QUAL_i + \delta \cdot QUAL_i + \varepsilon_i \end{aligned} \tag{2.6}$$

where GPA_t and GPA_{t-1} represent the current and previous years' grade point averages, X is a vector of individual characteristics, and $QUAL$ is dummy variable for being qualified for placement on the Dean's List. In estimating equation (6) the regression coefficient δ represents the discontinuity at the award criterion and is an estimate of the signaling effect of the silver star described in equations (3) and (5). Including the square and cube of previous GPA ensures that the discontinuity is not a function of non-linearities in the mapping of previous GPA onto current GPA that might otherwise appear as a discontinuity in a linear mapping.

The existence of regimes with and without the silver star provides comparable treatment and control groups, so, to build upon the simple specification, the discontinuity detected for those in the silver star regime can be compared to the same discontinuity in a mapping of GPA for a counterfactual group. If the regression coefficient δ results from an insufficiently robust specification of previous GPA and not from a signaling effect of the silver star, then using the counterfactual group will identify this problem. Therefore, the treatment, or signaling, effect of having the silver star is identified because the policy change differentiates individuals as to whether or not they qualify for placement on the Dean's List and whether that award is observable.

The average effect of the treatment can be estimated by the following differences-in-differences specification:

$$\begin{aligned} GPA_{it} &= \alpha + \beta \cdot X_i + \lambda \cdot GPA_{it-1} + \\ &\quad \delta_1 \cdot QUAL_i + \delta_2 \cdot POST_i + \delta_3 \cdot QUAL_i \cdot POST_i + \varepsilon_i \end{aligned} \tag{2.7}$$

where $POST$ is a dummy variable for being in the silver star regime. In estimating equation (7) the regression coefficient δ_3 captures the average signaling effect for those who wore the silver star, or those who qualified for placement on the Dean's List and were in the silver star regime. As with specification (6), adding quadratic and cubic specifications of previous year's GPA ensures that the discontinuity is not a function of a non-linear relationship between GPA_t and GPA_{t-1}.

This specification, however, restricts the slopes of the GPA mapping to be the same for individuals with previous evaluations both above and below the criterion; in terms of the model, it restricts γ^A to be equal to γ^B and γ^0. Including interactions between GPA and the main effects relaxes these restrictions and allows me to test for discontinuities around the criterion point with the following stochastic equation:

$$\begin{aligned}GPA_{it} &= \alpha + \beta \cdot X_i + \\ &\lambda_0 \left(GPA_{it-1} - 3\right) + \lambda_1 \left(GPA_{it-1} - 3\right) \cdot QUAL_i + \\ &\lambda_2 \left(GPA_{it-1} - 3\right) \cdot POST_i + \lambda_3 \left(GPA_{it-1} - 3\right) \cdot QUAL_i \cdot POST_i + \\ &\delta_1 \cdot QUAL_i + \delta_2 \cdot POST_i + \delta_3 \cdot QUAL_i \cdot POST_i + \varepsilon_i. \end{aligned} \tag{2.8}$$

For this specification, the estimate for the parameter δ_3 measures the signaling effect at the award criterion. Again, this study benefits from the availability of an appropriate counterfactual to explain any differences between the qualifying and non-qualifying groups that are not explained by variation in the previous year's GPA or other covariates.

It is important to note that individuals whose true abilities are less than the award criterion might exert more effort than those whose abilities are greater than the criterion, as explained before, in order to earn the award. In such cases, the differences-in-differences estimates of the signaling effect may be contaminated. The estimates will equal the grade differential that results from the signaling effect minus the grade differential resulting from differential effort exertion. Additionally, when using the discontinuity designs in specifications (6) and (8), if the specification of previous GPA is not sufficiently robust, the relatively discontinuous nature of the optimal level of exerted effort at levels of understanding less that the award criterion – despite its continuity at the award criterion – may distort the estimates of the signaling effect.

However, using cadets' senior year as the current evaluation period and the junior year as the previous period avoids these problems. This is because earning the silver star during their senior year affords cadets no signaling effect during subsequent academic years, and as such, there should be no differential exertion of effort between those who did and did not earn the silver star during their junior year. These estimates should reflect the true signaling effect.

In addition to assuming that there is no incentive for seniors to earn the award, this framework ignores other motivational factors such as attempting to "live up to the signal" or any performance enhancement such as a sense of encouragement from having earned the award. Such motivational effects would cause the differences-in-differences and discontinuity estimates for seniors to be greater than the true signaling effect. I assume that these incentives or encouragement effects are negligible. The negative findings of Cook and Campbell (1979) regarding the Seaver and Quarton (1976) study of the effects of placement on the Dean's List on subsequent academic performance support this assumption.

However, the assumption of no differential exertion of effort cannot be made for juniors, because during their senior year they may benefit from having earned the silver star in their junior year. As such, some individuals who did not qualify for the award during their sophomore year may exert additional effort during their junior year – increasing their grades. One method of measuring the distorting effects of this additional effort is to measure the discontinuity at the award criterion using the first class in the silver star regime as the control group. Relaxing the implicit assumption of perfect information, one could argue reasonably that during the first academic year following the implementation of the silver star, cadets would not yet have recognized the potential benefits that earning the silver star may provide during subsequent years. As such, there would be no differential exertion of effort by the cadets in the Class of 1995 during their junior year, even though they were in the silver star regime.

Using the Classes of 1995 through 1999 and a differences-in-differences specification similar to that in equation (7) with the main effects of *Not Qualified for Silver Star* and *Not First* (the Class of 1995 being in the *First* period), the regression coefficient for *Not Qualified for Silver Star * Not First* would estimate the distorting effect that the exertion of additional effort to earn the award has on the estimates of the signaling effect. This is because any discontinuity at the award criterion for the first class in the silver star regime mostly would be a result of the signaling effect, while any discontinuity for all subsequent classes would result from the signaling effect minus the effect of the differential exertion of effort. Thus, differencing these would produce a lower bound for this distorting effect. This distortion would not be a measure of the effect that this additional effort has on grades but an indicator of the impact that the effort has on the estimation of the signaling effect

2.3.2 Data

The sample for this study consists of U.S. Military Academy cadets from the graduating classes of 1991 through 1999, for whom the GPA for each academic year is reported. Using these cadets' grades I examine the signaling effect of the silver star having been earned by performance during the previous year on the current year's GPA. For estimating the signaling effects of the silver star on senior year GPA, the experimental groups include cadets in the Classes of 1991 through 1993 as those in the control years and cadets in the Classes of 1995 through 1999 as those in the treated years. I exclude the Class of 1994 because this is the only class that was not completely in one of the two regimes for both the junior and senior years. Similarly, when estimating the signaling effects on the junior year GPA, I use the Classes of 1991 through 1994 as the control years and the Classes of 1996 through 1999 as the treated years, omitting the Class of 1995.

I exclude those cadets who earned a gold star or a wreath during their junior years to isolate the sample to those who either wore a silver star during their senior year or wore no insignia, and for the differences-in-differences estimates I limit the sample to those cadets whose junior year GPA was between 2.8 and 3.2 to focus on those cadets just above or below the 3.0 award criterion. Table 1a reports the means and variances of the senior and junior year GPA, cumulative GPA through the junior year, and Scholastic Aptitude Test (SAT) scores for cadets in their senior year. The table divides the sample into four groups, those who were and who were not on the Dean's List (qualified for the silver star) in both the non-silver star regime and the silver star regime. The non-qualifying and the qualifying groups are similar across regimes in terms of junior year GPA, cumulative GPA through the junior year, and

SAT scores. The groups are also similar across regimes in terms of senior year GPA, although some modest grade inflation over time is apparent. A small number of cadets in each regime with junior GPAs less than 3.0 still qualified for the silver star due to their cumulative GPAs. 7.8% of those in the non-silver star regime qualified by their cumulative GPAs alone, while a comparable 5.3% did so in the silver-star regime. This implies that the 3.0 criterion for junior year almost completely determines whether an individual in on the Dean's List – a "sharp" discontinuity design, to use Campbell's (1969) terminology, as opposed to a "fuzzy" discontinuity.

Table 1b reports similar descriptive statistics for sample groups used for estimating the signaling effects on the junior year GPA. These groups are also comparable across regimes. Two notable differences between the juniors and seniors are in the proportion of each that qualified for the Dean's List and in the means of qualification. Nearly one quarter of the seniors in the sample qualified for the Dean's List compared to a third of the juniors, and over twice as many of these seniors qualified by their previous year's grade alone. This is a function of higher average grades being awarded in junior year classes than those awarded during the sophomore year, or $\alpha^{Junior} > \alpha^{Sophomore}$

2.4 Results

2.4.1 The Signaling Effect

Table 2a reports estimates of the effect of the silver star on senior year GPA using the simple discontinuity design for the classes in the silver star regime (1995-1999) as described by equation (6). As discussed above, because there is no incentive for the seniors to re-earn the silver star, these estimates should not be contaminated by any differential effort induced by the award. The regression coefficient for *Qualified for Silver Star* measures the discontinuity at the award criterion, and in the linear specification of junior year GPA, the estimate of the signaling effect is 0.04 grade points and is statistically significant, as reported in column (1). This estimate increases to 0.07 with the cubic specification in column (3) implying that mapping of senior year GPA onto junior year GPA is non-linear and that the linear specification of GPA may result in inaccurate estimates of the signaling effect. These estimates are equivalent to one thirds to two thirds of a letter grade in one course during a ten course academic year.

The differences-in-differences specification in equation (7) takes advantage of the control afforded by the cadets not in the silver star regime and estimates the average effect of the silver star on senior year GPA. To focus on those cadets just above or below the 3.0 award criterion I limit the sample to those cadets whose junior year GPA was between 2.8 and 3.2. The regressors *Qualified for Silver Star* and *Post* are the main effects of being qualified for the silver star (on the Dean's List) and of being in the silver star regime, respectively, and the estimates of the signaling effect are represented by the coefficients for the regressor *Qualified for Silver Star * Post.* Column (1) of Table 2b reports a simple differences-in-differences estimate of the signaling effect of 0.052 grade points, approximately one half of a letter grade. The coefficient for the main effect *Post* is small (-0.020) and insignificant, implying that the treatment and control groups are adequately comparable.

The specification in column (2) includes junior year GPA as a covariate, and the fact that its regression coefficient is statistically significant (as it is in the previous design) demonstrates an understandably strong positive correlation between previous GPA and current GPA. The estimate for the signaling effect in this specification is 0.048, nearly the same the estimate in the previous specification, and is statistically significant. However, the estimates for the main effects *Qualified for Silver Star* and *Post* are both small and insignificant. This indicates that the variation in the covariates, specifically junior year GPA, explains the majority of the differences between the groups. Because there is no significant discontinuity in the GPA mapping at the award criterion for the control group, the estimates of the signaling effect from the simple discontinuity design should and do resemble the estimates in these specifications. Additionally, the estimate remains stable as the square and cube of junior year GPA are added as reported in columns (3) and (4). Figure 3a graphs the predicted senior year GPA on junior year GPA, according to the cubic specification in column

(4) without additional covariates. The parallel separation between the predicted senior year GPAs for those in the silver star regime who are not on the Dean's List and those who are depicts the average signaling effect of the silver star. These estimates of the signaling effect are also robust to the inclusion of cumulative GPA and its square and cube as additional covariates as reported in columns (5) through (7).

Table 3c reports the regression coefficients from estimates of the fully interacted specification in equation (7).[14] Unlike the differences-in-differences estimates, the fully interacted specifications allow the slopes of the relationship between previous and current GPA to differ for each group and estimate the discontinuity at the award criterion point of 3.0. The specification for columns (1) through (3) do not allow for different slopes above and below the criterion for the control group, while the specifications in columns (4) through (6) do. This restriction allows a comparison of the slopes for the silver star regime to the non-silver star regime. In terms of the model, it allow for a test of whether $\gamma^1 > \gamma^A$ and $\gamma^1 > \gamma^B$ as predicted.

The discontinuity at the award criterion of 3.0 between the predicted senior year GPAs for those in the silver star regime who are not on the Dean's List and those who are depicts the estimated signaling effect at the award criterion. The estimate of this discontinuity in the linear specification is small (0.008) because of the detected discontinuity in the control group ($Qual$ = 0.030). That this coefficient is not equal to zero indicates that the linear specification is not sufficiently precise in mapping current GPA onto previous GPA. However, in the quadratic and cubic specifications in columns (2) and (3) respectively, this coefficient is close to zero, and the estimates for the signaling effects increase to 0.04 and 0.07. Additionally, it appears as though the slopes of the mapping in the silver star regime are less steep ($\gamma^1 > \gamma^A$ and $\gamma^1 > \gamma^B$) due to the negative coefficients interactions of junior year GPA with the main effects. However, this simply may be a function of a decrease in the range of awarded grades for seniors due to or grade inflation instead of an indication of the

[14]For these specifications, the sample is not limited to a range of junior year GPAs. This selection criteria is not necessary for focusing the sample, and including a greater range increases the precision of the slopes' estimates.

incorporation of information of group characteristics.

The estimates of the signaling effect for the specifications that do allow for different slopes for the control group are similar. The estimate for the quadratic specification in column (5) is 0.06 grade points, and the estimate from the cubic specification in column (6) is 0.12. Figure 3b graphs the predicted senior year GPA according to the cubic specification without additional covariates and illustrates the discontinuity at the award criterion. However, the large and negative coefficient for *Qual* indicates that the estimate from the cubic specification may overstate the signaling effect. This coefficient should equal zero and suggests that the higher order specifications of the GPA mapping may be overly dynamic.

All of these estimation designs indicate a robust estimate of the signaling effect of the silver star of about 0.05 grade points on yearly GPA, or half of a letter grade in one course during a ten course academic year. This is not unrealistically large, and the fact that West Point calculates GPA using thirds of a letter grade (i.e. a B = 3.0 while a B- = 2.67) makes the suggestion that a signal might affect a course grade more plausible. Two points suggest that this estimate, as is pertains to a single graded requirement, may be understated. First, as at most institutions, at West Point some course requirements are graded "blind". For example some tests are identified only by an identification number, and for these events, there can be no information updating. However, other requirements such as papers, presentations, and class participation are not anonymous. As such, for an average signaling effect of 0.5, the signaling must be larger for those events that are not graded "blind". Second, the submissions are not graded simultaneously. Therefore, an instructor may continuously update his or her assessment of the cadet's understanding in a manner similar to the employer learning model described in Altonji and Pierret (2001). If learning by the instructor occurs, then the signaling effect of the silver star will dissipate with each graded requirement. Thus, the actual signaling effect for the first few events will be larger than the average effect.

Given this estimate of the signaling effect and the descriptive statistics for seniors' grades, I can estimate the value of some of the population parameters and verify whether they are economically reasonable. Based upon the predicted value of the grade differential in equation (3), I estimate that

$$\begin{aligned}\gamma &= 0.9,\\ Var\left(q^1\right) &= 0.2,\ \text{and}\\ Var\left(u^1\right) &= 0.022.\end{aligned}$$

This estimate of the variance of the error term indicates the precision of the grading system. A standard deviation of 0.15 implies that the two thirds of the students will have a range of one third of a letter grade deviation from their true comprehension, and 96% will have a range of two thirds of a letter grade. This calibration exercise implies that the estimates of the signaling effects are indeed reasonable.

2.4.2 Induced Effort

Just as estimating the signaling effects on wages of different levels of education are complicated by human capital effects, estimating the signaling effects of indicators of past performance on the evaluation of current performance are complicated by the effects of induced effort in the current period by such signals. As mentioned above, individuals whose true abilities are less than the award criterion should exert more effort than those whose abilities are greater than the criterion to earn the award. As such, regression discontinuity and differences-in-differences estimates of the signaling effect may be contaminated by this additional effort, which may reduce the estimates or make them negative.

While the analysis based upon senior year GPA assumes that there are no differential incentive effects induced by the silver star, this cannot be assumed for junior year GPA. Table 3 reports differences-in-differences estimates of the effect of the silver star on junior year GPA[15] and demonstrates the distorting effect of induced effort. The simple differences-in-differences estimate of -0.047 in column (1) significantly differs

[15]Estimates of the effect of the silver star on sophomore year GPA, though not reported, are similar to those reported for the juniors.

from the corresponding estimate of 0.052 for the seniors. Assuming that the size of the signaling effect is the same for juniors as it is for seniors, about 0.05 grade points, the results imply that the distorting effect of the induced effort is nearly 0.1 grade points. The specifications in columns (2) through (7) progressively include sophomore year GPA, cumulative GPA, and their square and cube as covariates. In each of these specifications the regression coefficient for *Qualified for Silver Star * Post* remains between -0.04 and -0.05 and is always significantly different from the corresponding estimate for the seniors.[16] As with the seniors, the inclusion of the previous year's GPA reduces the size of the regression coefficient for *Qualified for Silver Star,* but this estimate becomes insignificant only when cumulative GPA is included as an additional covariate. This implies that sophomore year GPA is less accurate in predicting junior year GPA than junior year GPA is in predicting senior year GPA. As such, the signaling effect of the silver star on junior year grades should be less than that on senior year grades, implying that the estimate of 0.1 grade points for the distorting effect of induced effort is an upper bound.

As discussed at the end of Section 3.1, I can compare the performances of the cadets in the first year after the implementation of the silver star with those in subsequent years to better estimate how the effort exertion can affect the estimates. Using the Classes of 1995 through 1999 and a differences-in-differences specification similar to that in equation (7) with the main effects of *Not Qualified for Silver Star* and *Not First* (the Class of 1995 being in the *First* period), the regression coefficient for *Not Qualified for Silver Star * Not First* would estimate this distorting effect. Table 4 reports the estimates from this specification. In column (1) the simple differences-in-differences estimate for the coefficient for *Not Qualified for Silver Star * Not First* is 0.077, which significantly differs from the above estimate of -0.044 for the grade differential that results from the net signaling and effort effect. Including sophomore year GPA and it square and cube results in estimates of 0.08, as reported in columns (2) through (4). This estimate implies that the difference in the effort exerted by those whose ability is just below the award criterion in the first junior class under

[16]Though not reported, estimates using the discontinuity specifications are similar.

the regime and all subsequent classes resulted in a difference in the estimates of the signaling effect of 0.08 grade points. Again, this assumes that the initial class under the regime exerted no additional effort, or significantly less, because they did not yet recognize the potential benefits that earning the silver star might provide during subsequent years. As such, this estimate is a lower bound for the effect of the additional effort.

Without any knowledge of the true value of the model's parameters and the cost of effort, it is extremely difficult to correct for this problem parametrically. A slightly less elegant method of compensating for the differential exertion of effort is to include dummy variables for individuals whose sophomore year GPA is in one of several different ranges of GPA and in the silver star regime. This specification assumes that the GPA mapping for these individuals would be the same as those not in the range if there were no additional effort exerted. Therefore, the discontinuity at the award criterion would estimate the differential that would have occurred between those that earned the silver star and those that did not had the latter not exerted additional effort.

Table 5 reports the estimates of the signaling effect of the silver star on junior year GPA using this technique with the fully integrated specification in equation (8). The specifications in columns (1) through (3) include dummy variables for those in the silver star regime whose sophomore year GPA was between 2.7 and 3.0 in groups with ranges of 0.1 grade points. For columns (4) through (6), the range of GPA is from 2.6 to 3.0. All of the estimates of the signaling effect with these specifications suffer from large standard errors, and in all cases the null hypothesis that they equal zero cannot be rejected. Interestingly, though, the pattern of estimated additional effort, as indicated by the regression coefficients for the dummy variables, suggests that additional effort is an increasing and then decreasing function of previous GPA – a similar pattern to that predicted by equation (5) and Figure 2.

2.4.3 Noise and the Size of the Signaling Effect

The Academy classifies cadets' majors or fields of study as being either math, science, and engineering (MSE) or humanities and public affairs (HPA). It can be argued that the graded requirements in the classes such as history, English, and political science are more subjective than the requirements in the engineering and mathematical classes and, therefore, may be noisier indicators of true ability. One theoretical prediction of the model discussed earlier is that the size of the signalling effect decreases with the reliability of the observed output as an indicator of true productivity. In other words, as output becomes a noisier indicator of true ability, the signaling effect becomes larger. Therefore, I test the prediction of differential signaling effects by estimating the effects separately for the MSE and HPA majors.

However, the theory also predicts that (1) a noisier observed output for the previous period results in a smaller signaling effect and that (2) a decrease in the correlation between the abilities for the two tasks results in a decrease in the size of the signaling effect. In regard to the first point, it is important to note that for the first two years, nearly all cadets take the same required set of core courses. Additionally, approximately half of the courses taken during the junior year are electives that contribute to an academic major or field of study. Considering that 65% of all the seniors who qualified for the silver star had a cumulative GPA of at least 3.0, I argue that the observed output for the previous period is not particularly noisier for HPA majors than for MSE majors. Therefore, any differences in the signaling effects between the two groups should not be diminished by differences in the noisiness of the previous output.

In regard to the second point, the required, core curriculum taken during the first two years is equally balanced between quantitatively intense courses, such as math and physics, and less quantitative courses, such as English and history. Because of this balance, a cadet's ability regarding the core curriculum should not be more or less correlated with his/her abilities in his/her senior year electives. Therefore, any differences in the signaling effects between the two groups should not be diminished by differences in the correlations between previous and current performance.

Table 6 reports estimates of the effect of the silver star on senior year GPA for the fully interacted specification (equation (8)) for both MSE and HPA majors. The linear specifications of the model, as reported in columns (1) and (4), do not yield large or significant estimates for the signaling effect. However, the estimates for the signaling effect in the quadratic specification in columns (2) and (5) are 0.061 and 0.092 for the MSE and HPA majors, respectively. For the cubic specification in columns (3) and (6), the estimates are 0.108 and 0.137, respectively. While not significantly different from each other, if the evaluations in the HPA classes are held to be more subjective than MSE classes, then these estimates suggest that when the output is a noisier indicator of true ability for the current task, the signaling effects are larger.

2.5 Conclusions

Empirical efforts to estimate the effect of a signal of individual quality on performance evaluation are challenged by the difficulty of identifying comparable treatment and control groups for analysis. Using a natural experiment resulting from the creation of an academic award that assumes the form of a prominently displayed insignia, this study offers a unique examination of the effects of criterion-based signals on both the evaluation of an individual's performance and the effort induced to earn the award.

The described statistical discrimination model provides a theoretical framework for analyzing the rather intuitive notion that individuals with higher credentials may be judged or evaluated differently. The model predicts a signaling effect for those who display the signal and perform at a certain level vs. those who do not display the signal but have the same level of performance. It also offers predictions as to how the magnitude of the signaling effect may change as other parameters in the model change, such as the correlation between the basis of the award and the current task being evaluated. Additionally, the model suggests that because there is a beneficial effect of having earned the award during subsequent evaluations, individuals will exert effort differentially in an attempt to earn the award depending upon where their natural abilities lie in relation to the award criterion.

While this paper examines the effects of signals on performance evaluation using the context of academic grading, which is relevant both to the perennial interest in grading and academic awards and to the contemporary interest in grade-inflation at elite universities and institutions, the theoretical results are highly transferable to other situations in which individual quality is assessed. These effects are also relevant to performance rewards and incentive mechanisms in either an academic or work setting. Certainly, further research that measures signaling effects and their induced incentive effects in other situations is warranted.

Figure 1a.

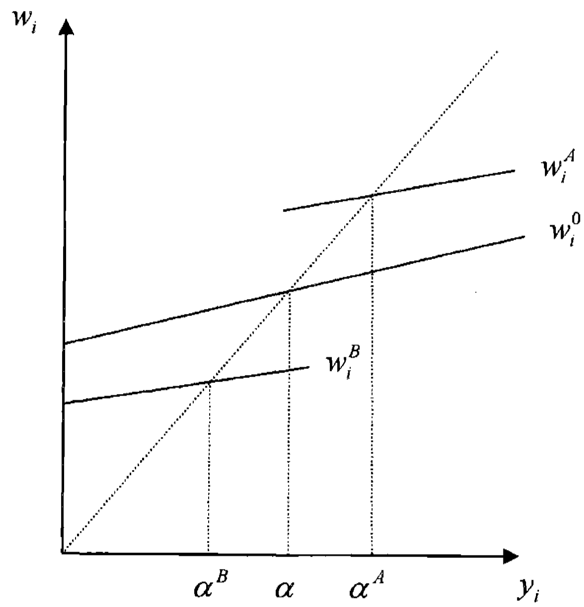


Figure 1b.

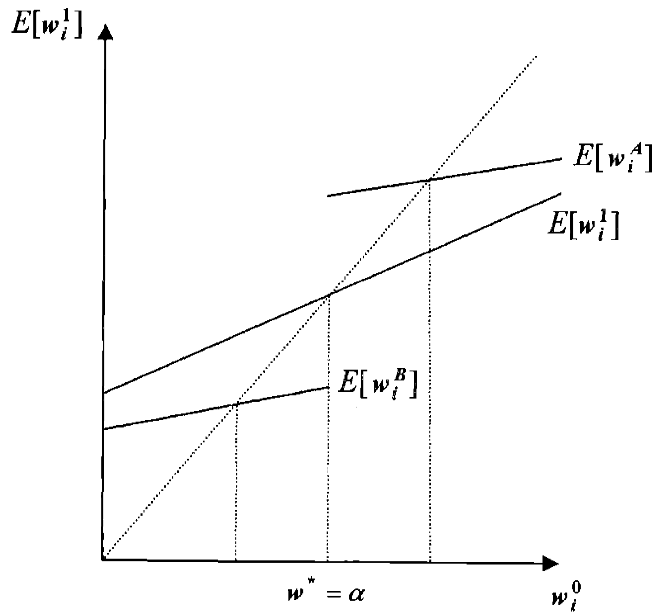
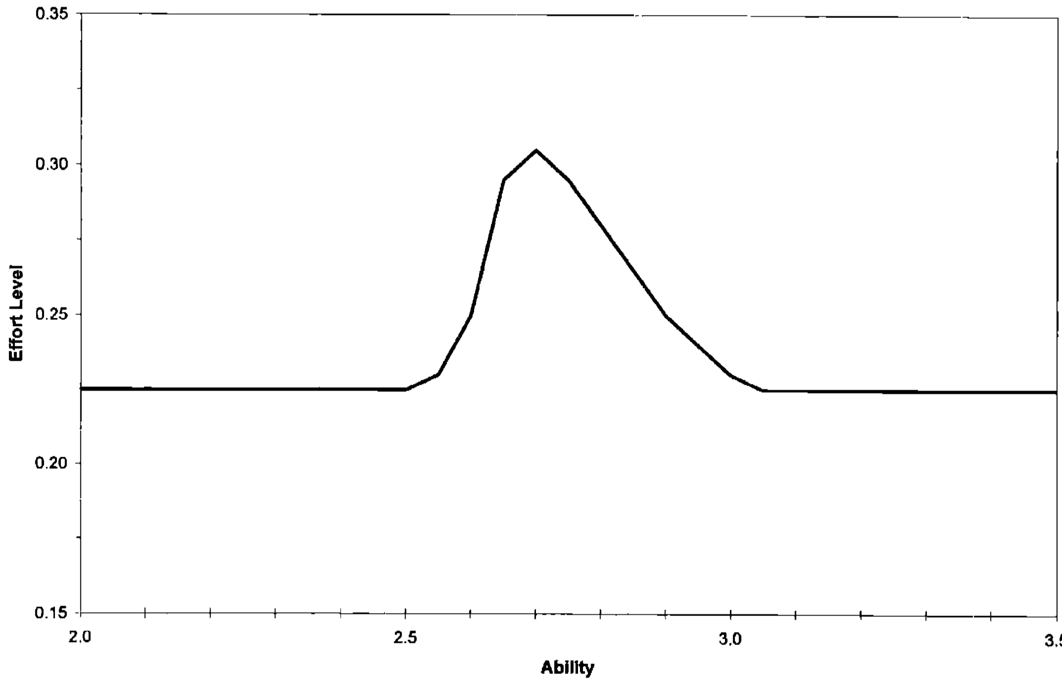
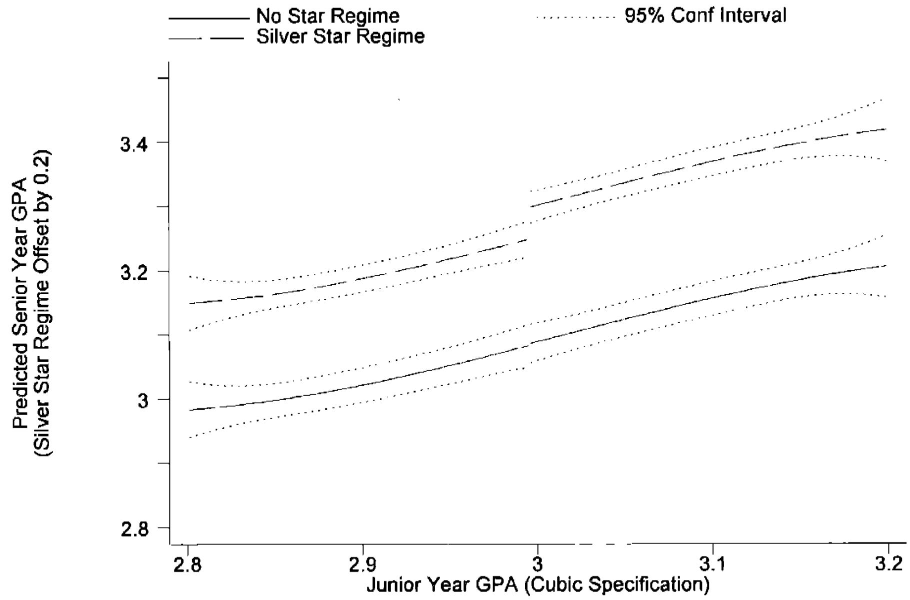


Figure 2. Optimal Level of Exerted Effort



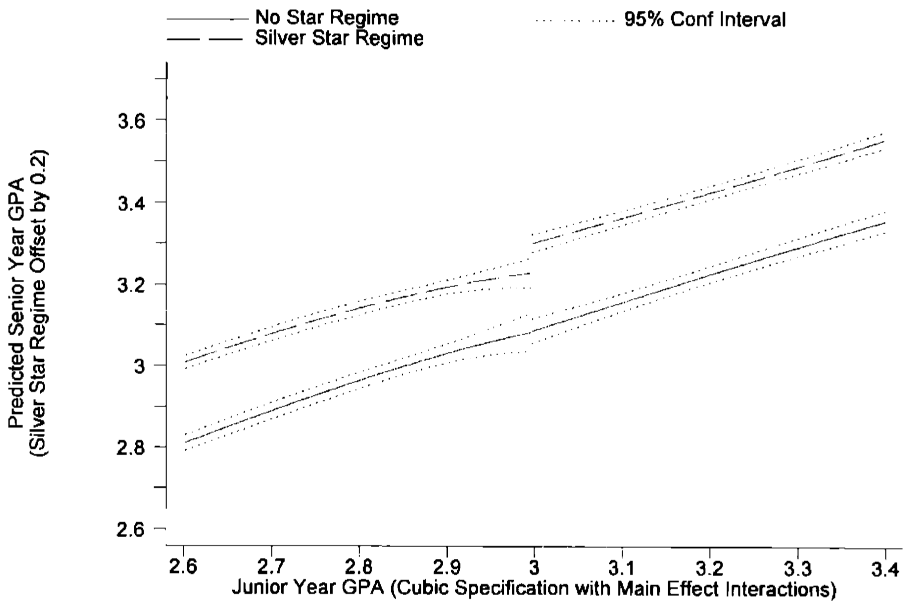
This figure graphs the optimal level of effort exerted by individuals, given their ability levels, as represented by the solution to equation (5) in the text. The parameters for this example are $w^* = 3$, $a = 2$, $b = 0.08$, $\gamma = 0.9$, and $\sigma_u = 0.1$. The figure demonstrates that while most individuals exert an equal amount of effort (in this example about 0.225), those individuals whose ability levels, or latent understanding, are just below the award criterion will exert the more effort to earn the award than those whose ability is greater than or far less than the award criterion.

Figure 3a.



This figure graphs the predicted senior year GPAs on a cubic specification of junior GPA as estimated by equation (7) in the paper and as reported in column (3) of Table 3a but without covariates. The predicted values for those in the silver star regime are offset by 0.02 grade point. For graphical clarity only, the figure does not display the predicted values for those individuals who qualified for the silver star by cumulative GPA alone.

Figure 3b.



This figure graphs the predicted senior year GPAs on a cubic specification of junior GPA as estimated by equations (8) in the paper and as reported in column column (6) of Table 3b but without covariates. The predicted values for those in the silver star regime are offset by 0.02 grade point. For graphical clarity only, the figure does not display the predicted values for those individuals who qualified for the silver star by cumulative GPA alone. Additionally, the figure graphs only the predicted values for those whose junior year GPA was between 2.6 and 3.4, although the estimates were conducted for all of those below 3.67.

Table 1a
Descriptive Statistics of Seniors

	Not Qualified for Silver Star		Qualified for Silver Star	
	Non-Silver Star Regime (1)	Silver Star Regime (2)	Non-Silver Star Regime (3)	Silver Star Regime (4)
Number of Observations	1685	2031	883	1053
Junior Year GPA	2.500	2.618	3.240	3.257
	(0.331)	(0.266)	(0.199)	(0.201)
Cumulative GPA through Junior Year	2.439	2.492	3.080	3.064
	(0.276)	(0.241)	(0.234)	(0.247)
Senior Year GPA	2.754	2.810	3.250	3.269
	(0.340)	(0.307)	(0.268)	(0.265)
Total SAT Score	1158.8	1165.0	1244.4	1233.7
	(114.0)	(108.2)	(99.1)	(105.0)
Qualified for Silver Star by:				
Junior Year GPA at least 3.0 only	-	-	35.6%	37.3%
Cumulative GPA at least 3.0 only	-	-	7.8%	5.3%
Both	-	-	56.6%	57.4%

Notes: This table reports the means and variances of the above listed measures for cadets of the graduating classes of 1991 through 1993 (non-silver star regime) and 1995 through 1999 (silver star regime) The sample includes those cadets whose junior year GPA was less than 3.67 and excludes those who earned the wreath or gold star.

Table 1b
Descriptive Statistics of Juniors

	Not Qualified for Silver Star		Qualified for Silver Star	
	Non-Silver Star Regime (1)	Silver Star Regime (2)	Non-Silver Star Regime (3)	Silver Star Regime (4)
Number of Observations	2816	2233	919	880
Sophomore Year GPA	2.400	2.465	3.213	3.229
	(0.366)	(0.333)	(0.195)	(0.196)
Cumulative GPA through Sophomore Year	2.434	2.483	3.170	3.160
	(0.314)	(0.288)	(0.176)	(0.186)
Senior Year GPA	2.651	2.795	3.287	3.320
	(0.413)	(0.391)	(0.301)	(0.300)
Total SAT Score	1168.4	1175.3	1254.9	1255.4
	(111.9)	(106.8)	(101.0)	(101.5)
Qualified for Silver Star by:				
Sophomore Year GPA at least 3.0 only	-	-	14.7%	16.8%
Cumulative GPA at least 3.0 only	-	-	10.2%	8.4%
Both	-	-	75.1%	74.8%

Notes: This table reports the means and variances of the above listed measures for cadets of the graduating classes of 1991 through 1994 (non-silver star regime) and 1996 through 1999 (silver star regime) The sample includes those cadets whose sophomore year GPA was less than 3.67 and excludes those who earned the wreath or gold star.

Table 2a
Estimates of the Effect of the Presence of the Silver Star
on Senior Year Grade Point Average – Simple Discontinuity Design

	(1)	(2)	(3)
Qualified for Silver Star	0.039	0.044	0.069
	(0.013)	(0.017)	(0.020)
Junior Year GPA	0.651	0.654	0.363
	(0.022)	(0.068)	(0.150)
Junior Year GPA^2		0.003	-0.673
		(0.070)	(0.342)
Junior Year GPA^3			-0.407
			(0.214)
Junior Year GPA * Qualified for Silver Star	0.013	-0.090	0.192
	(0.035)	(0.096)	(0.164)
Junior Year GPA^2 * Qualified for Silver Star		0.186	0.997
		(0.136)	(0.435)
Junior Year GPA^3 * Qualified for Silver Star			0.202
			(0.434)

Notes: This table reports the regression coefficients for a regression of senior year GPA on the above regressors and controls for gender, race, class year, attendance at the U.S. Military Academy Preparatory School, previous military service, and SAT score for cadets from the graduating classes of 1995-1999. The sample includes those cadets whose junior year GPA was less than 3.67 and excludes those who earned the wreath or gold star (n = 3,513). Heteroskedasticity robust standard errors are reported in parentheses.

Table 2b
Estimates of the Effect of the Presence of the Silver Star
on Senior Year Grade Point Average – Differences-in-Differences Specification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Qualified for Silver Star * Post (Received Silver Star)	0.052 (0.022)	0.048 (0.021)	0.048 (0.021)	0.048 (0.021)	0.048 (0.021)	0.047 (0.021)	0.045 (0.021)
Post	-0.020 (0.024)	-0.026 (0.024)	-0.025 (0.024)	-0.025 (0.024)	-0.021 (0.024)	-0.020 (0.024)	-0.019 (0.024)
Qualified for Silver Star	0.110 (0.017)	0.009 (0.022)	0.010 (0.022)	0.005 (0.024)	-0.017 (0.022)	-0.015 (0.023)	-0.027 (0.025)
Junior Year GPA		0.577 (0.076)	0.577 (0.075)	0.663 (0.168)	0.548 (0.076)	0.546 (0.076)	0.764 (0.169)
GPA^2			0.424 (0.442)	0.385 (0.447)		0.451 (0.446)	0.348 (0.450)
GPA^3				-2.881 (4.990)			-7.419 (5.051)
Cumulative GPA					0.143 (0.029)	0.134 (0.046)	0.141 (0.046)
Cumulative GPA^2						-0.023 (0.084)	0.073 (0.172)
Cumulative GPA^3							0.156 (0.251)

Notes: This table reports the regression coefficients for a regression of senior year GPA on the above regressors and controls for gender, race, class year, attendance at the U.S. Military Academy Preparatory School, previous military service, and SAT score for cadets from the graduating classes of 1991-1993 and 1995-1999. The sample includes those cadets whose junior year GPA was between 2.8 and 3.2 and excludes those who earned the wreath or gold star (n = 2,008). Heteroskedasticity robust standard errors are reported in parentheses.

Table 2c
Estimates of the Effect of the Presence of the Silver Star
on Senior Year Grade Point Average – Discontinuity Design with Control Group

	(1)	(2)	(3)	(4)	(5)	(6)
Qualified for Silver Star * Post	0.008	0.040	0.068	0.013	0.059	0.118
(Received Silver Star)	(0.021)	(0.025)	(0.029)	(0.022)	(0.029)	(0.032)
Post	0.006	0.000	-0.033	0.010	-0.028	-0.086
	(0.018)	(0.020)	(0.029)	(0.018)	(0.026)	(0.028)
Qualified for Silver Star	0.030	0.003	0.001	0.024	-0.016	-0.050
	(0.016)	(0.019)	(0.020)	(0.017)	(0.024)	(0.025)
Junior Year GPA	0.627	0.684	0.691	0.619	0.804	1.045
	(0.020)	(0.029)	(0.036)	(0.021)	(0.088)	(0.122)
GPA^2		0.060	0.021		0.149	0.549
		(0.028)	(0.053)		(0.075)	(0.201)
GPA^3			-0.034			0.173
			(0.058)			(0.093)
GPA * Qualified for Silver Star * Post	0.013	-0.154	0.203	-0.032	0.016	0.542
	(0.035)	(0.111)	(0.165)	(0.059)	(0.168)	(0.232)
GPA^2 * Qualified for Silver Star * Post		-0.149	0.991		0.109	1.505
		(0.103)	(0.436)		(0.246)	(0.641)
GPA^3 * Qualified for Silver Star * Post			0.225			0.600
			(0.434)			(0.737)
GPA * Post	0.022	-0.035	-0.336	0.030	-0.154	-0.691
	(0.029)	(0.074)	(0.155)	(0.030)	(0.111)	(0.194)
GPA^2 * Post		-0.059	-0.705		-0.149	-1.233
		(0.076)	(0.348)		(0.103)	(0.399)
GPA^3 * Post			-0.378			-0.585
			(0.223)			(0.235)
GPA * Qualified for Silver Star				0.046	-0.101	-0.340
				(0.047)	(0.138)	(0.164)
GPA^2 * Qualified for Silver Star					-0.226	-0.513
					(0.205)	(0.471)
GPA^3 * Qualified for Silver Star						-0.375
						(0.597)

Notes: This table reports the regression coefficients for a regression of senior year GPA on the above regressors and controls for gender, race, class year, attendance at the U.S. Military Academy Preparatory School, previous military service, and SAT score for cadets from the graduating classes of 1991-1993 and 1995-1999. The sample includes those cadets whose junior year GPA was less than 3.67 and excludes those who earned the wreath or gold star (n = 6,076). Heteroskedasticity robust standard errors are reported in parentheses.

Table 3
Estimates of the Effect of the Presence of the Silver Star
on Junior Year Grade Point Average – Differences-in-Differences Specification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Qualified for Silver Star * Post (Received Silver Star)	-0.047 (0.027)	-0.052 (0.027)	-0.052 (0.027)	-0.053 (0.027)	-0.041 (0.027)	-0.042 (0.027)	-0.041 (0.027)
Post	0.045 (0.032)	0.049 (0.032)	0.049 (0.032)	0.048 (0.032)	0.048 (0.032)	0.048 (0.032)	0.047 (0.032)
Qualified for Silver Star	0.156 (0.045)	0.063 (0.024)	0.062 (0.024)	0.067 (0.025)	0.014 (0.026)	0.014 (0.026)	0.002 (0.029)
Junior Year GPA		0.550 (0.090)	0.550 (0.090)	0.457 (0.196)	0.420 (0.093)	0.425 (0.094)	0.489 (0.195)
GPA^2			-0.100 (0.554)	-0.047 (0.567)		-0.093 (0.575)	-0.109 (0.589)
GPA^3				3.251 (5.913)			-1.925 (5.970)
Cumulative GPA					0.316 (0.064)	0.298 (0.076)	0.360 (0.092)
Cumulative GPA^2						-0.112 (0.222)	-0.292 (0.290)
Cumulative GPA^3							-0.822 (0.807)

Notes: This table reports the regression coefficients for a regression of junior year GPA on the above regressors and controls for gender, race, class year, attendance at the U.S. Military Academy Preparatory School, previous military service, and SAT score for cadets from the graduating classes of 1991-1994 and 1995-1999. The sample includes those cadets whose junior year GPA was between 2.8 and 3.2 and excludes those who earned the wreath or gold star (n = 1,707). Heteroskedasticity robust standard errors are reported in parentheses.

Table 4
Estimates of the Effect of the Presence of the Silver Star on GPA
for those Not Qualifying for the Silver Star – A Test of Effort Exertion

	(1)	(2)	(3)	(4)
Not Qualified for Silver Star * Not First	0.077	0.083	0.082	0.082
(Exerted Additional Effort)	(0.044)	(0.043)	(0.043)	(0.044)
Not First	-0.006	-0.007	-0.007	-0.006
(Not First Class in Silver Star Regime)	(0.035)	(0.034)	(0.034)	(0.034)
Not Qualified for Silver Star	-0.190	-0.112	-0.115	-0.113
	(0.040)	(0.044)	(0.045)	(0.048)
Sophomore Year GPA		0.465	0.462	0.502
		(0.123)	(0.123)	(0.264)
GPA^2			0.875	0.846
			(0.743)	(0.778)
GPA^3				-1.388
				(8.026)

Notes: This table reports the regression coefficients for a regression of junior year GPA on the above regressors and controls for gender, race, class year, attendance at the U.S. Military Academy Preparatory School, previous military service, and SAT score for cadets from the graduating classes of 1995-1999 (n = 1,037). The sample includes those cadets whose sophomore year GPA was between 2.8 and 3.2 and excludes those who earned the wreath or gold star. Heteroskedasticity robust standard errors are reported in parentheses.

Table 5
Estimates of the Effect of the Presence of the Silver Star
on Junior Year Grade Point Average – Discontinuity Design with Controls for Effort Effect

	(1)	(2)	(3)	(4)	(5)	(6)
Specification of Sophomore Year GPA Interacted with Main Effects	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
Qualified for Silver Star * Post (Received Silver Star)	-0.043 (0.032)	-0.033 (0.060)	-0.055 (0.103)	-0.024 (0.036)	0.005 (0.079)	0.018 (0.125)
Post	0.076 (0.028)	0.077 (0.058)	0.104 (0.101)	0.058 (0.032)	0.040 (0.077)	0.031 (0.124)
Qualified for Silver Star	0.058 (0.017)	0.016 (0.021)	0.044 (0.026)	0.058 (0.017)	0.016 (0.021)	0.043 (0.026)
GPA between 2.9 and 3.0 * Post	0.029 (0.024)	-0.026 (0.052)	-0.021 (0.082)	0.029 (0.034)	0.007 (0.068)	0.041 (0.102)
GPA between 2.8 and 2.9 * Post	0.039 (0.025)	0.012 (0.039)	0.015 (0.053)	0.054 (0.029)	0.038 (0.051)	0.059 (0.068)
GPA between 2.7 and 2.8 * Post	0.011 (0.030)	0.012 (0.030)	0.014 (0.034)	0.043 (0.026)	0.031 (0.039)	0.043 (0.044)
GPA between 2.6 and 2.7 * Post				0.031 (0.026)	0.024 (0.031)	0.029 (0.032)

Notes: This table reports the regression coefficients for a regression of junior year GPA on the above regressors and controls for gender, race, class year, attendance at the U.S. Military Academy Preparatory School, previous military service, and SAT score for cadets from the graduating classes of 1991-1994 and 1996-1999. The sample includes those cadets whose sophomore year GPA was less than 3.67 and excludes those who earned the wreath or gold star (n = 6,818). Heteroskedasticity robust standard errors are reported in parentheses.

Table 6
Estimates of the Effect of Presence of Silver Star
on Senior Year Grade Point Average By Major or Field of Study

	Math, Science and Engineering Majors			Humanities and Public Affairs Majors		
	(1)	(2)	(3)	(4)	(5)	(6)
Qualified for Silver Star * Post	0.010	0.061	0.108	0.008	0.092	0.137
(Received Silver Star)	(0.028)	(0.036)	(0.041)	(0.032)	(0.041)	(0.047)
Post	0.043	-0.007	-0.057	-0.032	-0.067	-0.113
	(0.025)	(0.033)	(0.037)	(0.026)	(0.034)	(0.042)
Qualified for Silver Star	0.042	0.013	-0.017	0.011	-0.081	-0.091
	(0.022)	(0.029)	(0.030)	(0.025)	(0.033)	(0.038)
Junior Year GPA	0.626	0.764	1.001	0.593	0.907	0.998
	(0.027)	(0.104)	(0.144)	(0.032)	(0.099)	(0.217)
GPA^2		0.110	0.486		0.260	0.432
		(0.089)	(0.220)		(0.081)	(0.398)
GPA^3			0.154			0.086
			(0.096)			(0.202)
GPA * Qualified for Silver Star * Post	-0.030	0.232	0.658	-0.032	-0.475	-0.065
	(0.077)	(0.203)	(0.287)	(0.088)	(0.259)	(0.379)
GPA^2 * Qualified for Silver Star * Post		0.235	1.123		1.118	2.832
		(0.301)	(0.749)		(0.413)	(1.323)
GPA^3 * Qualified for Silver Star * Post			0.261			-0.182
			(0.864)			(1.629)
GPA * Post	0.063	-0.201	-0.623	-0.007	-0.152	-0.646
	(0.041)	(0.136)	(0.248)	(0.043)	(0.133)	(0.280)
GPA^2 * Post		-0.237	-1.044		-0.087	-1.188
		(0.124)	(0.505)		(0.117)	(0.534)
GPA^3 * Post			-0.415			-0.641
			(0.293)			(0.281)
GPA * Qualified for Silver Star	-0.005	-0.177	-0.419	0.123	0.146	0.056
	(0.060)	(0.165)	(0.187)	(0.071)	(0.216)	(0.324)
GPA^2 * Qualified for Silver Star		-0.036	-0.477		-0.840	-1.014
		(0.251)	(0.509)		(0.357)	(1.229)
GPA^3 * Qualified for Silver Star			-0.026			-0.084
			(0.672)			(1.536)

Notes: This table reports the regression coefficients for a regression of senior year GPA on the above regressors and controls for gender, race, class year, attendance at the U.S. Military Academy Preparatory School, previous military service, and SAT score for a sample of cadets from the graduating classes of 1991-1993 and 1995-1999 whose junior year GPA was less than 3.67, excluding those who earned the wreath or old star. The sample size for the math, science, and engineering majors is 3,540 and for humanities and public affairs majors is 2,536. Heteroskedasticity robust standard errors are reported in parentheses.

Chapter 3

The Effect of Male Earnings on Marriage Rates: Evidence from the Texas Oil Boom and Bust

3.1 Introduction

Over the past several decades there has been a dramatic, but steady, decrease in marriage rates in the United States. From 1960 to 1999 the percentage of women age 20 to 24 who have never been married increased from 28.4% to 72.8%, while for women age 25 to 29 that percentage nearly quadrupled from 10.5% up to 38.9%.[1] It is important to understand this social phenomenon because of its many potential direct and indirect consequences. Candidates for the cause of these declines include changes in societal norms, increased labor market opportunities for women, and increased welfare generosity; and, most likely, no two explanations are mutually exclusive. In 1987, William J. Wilson proposed what has become known as the Wilson Hypothesis, that economic conditions resulting in large-scale, long-term joblessness, especially amongst inner-city black males, has been a "major underlying factor" in the large decline in marriage rates.[2] Thus, through the mechanism of "male marriageability",

[1] Author's calculations from U.S. Census Bureau data.
[2] Wilson, (1987), p. 83.

local economic conditions may affect the hazard rate into marriage.

Previous studies of the Wilson Hypothesis primarily have primarily used cross-sectional variation in local economic conditions[3] to test for corresponding variations in marriage rates or female headship rates in micro-data. Ellwood and Rodda (1991), Mare and Winship (1991), Wood (1995) use some variant of Wilson's "male marriageable pool index" (MMPI), the ratio of employed men to women within a standard metropolitan statistical area (SMSA), to measure the level of male economic status. Additionally, these studies use different identification strategies to address the problem of the endogeneity of male employment and earnings. Wood (1995) calls this the simultaneity of marital status and male employment and earnings, meaning that men may tend to work more once they marry or plan to marry as argued by Korenman and Neumark (1991). As a result, the estimates from a simple regression of male marital status on earnings may be biased upwards. In the end these studies find a positive and significant relationship between male economic conditions and marriage rates; however these relationships are not large enough to explain much of the decline in marriage rates.

A problem with these identification strategies is that, while they may instrument for local economic conditions, they do not account for the correlation between male and female labor market outcomes. If male and female earnings in a local market are positively correlated, regression estimates actually may be biased downwards because improvements in female labor market conditions may deter marriage. Another important issue is the lag structure in the relationship between economic conditions and marriage rates. The use of cross-sectional variation imposes a specific lag structure, usually a lag of one year, that may not reflect the effect of long term changes in economic conditions. This, too, may result in underestimates because of the insignificance of short-term variations in economic conditions on marriage rates.

To correct for these problems, I use the oil boom and bust in Texas between 1970 and 1990 as an experiment in which male earnings in a marriage market in-

[3]Alternatively, examples of time-series studies include Wilson (1987), Lerman (1988), and Ellwood and Crane (1990).

crease without a corresponding effect on female earnings. The 1973 Arab-Israeli War spawned an OPEC oil embargo that caused dramatic increases the price of oil in the United States. Consequently domestic oil production increased to meet demands, and in Texas, the number of operating oil rigs increased 35% between 1973 and 1974 and another 26% in 1975.[4] Because the petroleum industry is predominantly male, male employment and earnings in those counties in which petroleum extraction and refining were heavily concentrated, a geological function of the location of oil reserves, increased as production increased. However, counties with low concentrations of petroleum employment did not experience these same shocks. However, beginning in the early 1980s, as OPEC lifted the embargo and cut its oil prices, these economic conditions reversed themselves. The same counties that saw great increases in male employment and earnings during the 1970s suffered even greater declines through the 1980s. Given that the oil boom in the 1970s and the bust in the 1980s were caused by events exogenous to marriage propensities, I use it as a natural experiment for studying the effects of local economic conditions on marriage rates.

In addition to correcting for the endogeneity problem, the use of the oil boom and bust affords several other contributions. First, the oil shocks represent long-term changes in the economic prospects of a locality, changes that are more likely to affect family structure than high-frequency economic fluctuations represented in cross-sectional studies. Second, unique to this study is that I use records of marriages from the Texas Department of Health enabling me to analyze the effects on flows into marriage rather than merely on marriage stocks.[5] Black, McKinnish, and Sanders (2003) use shocks to the coal and steel industries to study the effects male economic conditions on welfare expenditures; however, using of marriage records provides a more direct measure of marriage flows. Next, unlike most cross-sectional studies, this natural experiment allows me to estimate the effects of changes in economic conditions

[4] University of Texas, "The Texas Handbook Online."

[5] Lichter et al. (1991), Sampson (1987), Mare and Winship (1991), and Wood (1995) use Census data and have to restrict the age range of women to approximate flows. Wood finds that increasing the restrictions on the age range from 20-34 to 20-29 and then to 20-24 actually reduces his estimated effect of the pool of "marriageable" men on marriage rates.

during distinct periods of fortune and decline making possible comparisons and tests of symmetry. Finally, unlike studies that equate marriageability with being employed, in this study I use male income levels to characterize the local marriage market, a characterization that Wood suggests is more informative than local male employment levels, a measure that may result in further underestimates of the effects of male economic conditions on marriage rates. As such, increases average male earnings may more directly result in increases in the probability of a good "match" in a marriage search model that result in increases in marriage rates.

Using long-differences estimates, I find a statistically significant and positive relationship between average county-level earnings and marriage rates of young women. I estimate that a 10% increase in male earnings results in a 15% increase in the hazard rate of never married young women into marriage. The instrumental variables (IV) estimates are all larger, as opposed to smaller, that ordinary least squares (OLS) estimates implying that attenuation and spurious correlation may bias the OLS estimates downward. However, like previous studies, these estimates suggest that changes in economic conditions can only explain a small portion of the overall decline in marriage rates. I estimate that the decline in economic conditions in Texas from 1980 to 1990 can explain only about 18% of the decline in marriage rates over that time. Unless the relationship between male earnings and marriage rates is significantly larger for inner-city blacks than Texans, these estimate imply that the decline in male earnings cannot explain the majority of the decline in marriage rates as Wilson hypothesizes. While the results from Texas may not be completely generalizable, they do suggest that more powerful social changes contributed to the decline in marriage rates experienced both overall during the past several decades and in specific inner-city areas.

3.2 Empirical Strategy and Data

Assuming that, in the aggregate, individual characteristics within the local market do not change significantly over time or that they change for all markets in a manner that can be captured by additive time period effects, then the hazard rate of unmarried women into marriage can be expressed as

$$Rate_{ct} = \alpha + \gamma m_{ct} + \delta_c + \lambda_t + \epsilon_{ct} \tag{3.1}$$

for marriage market c, in time t, where m_{ct} is some measure of male economic status. The coefficient γ represents the effect of local economic conditions for males on female hazard rates into marriage, and it can be estimated by weighted least squares, where the weights are the populations of unmarried women in the respective marriage markets.

Using standard time series data for these estimates prove to be problematic because current marriage rates are likely to be a function of contemporaneous and previous economic conditions requiring a lag specification, and even if one knew the appropriate lag structure to specify, these economic conditions are likely to be serially correlated. Using large time period differences with cross-sectional data avoids many of the estimation and inference issues related to the time series nature of the data. For these reasons, a strategy of long differences is computationally attractive, and for two time periods, γ can be estimated with the following simple differences equation:

$$\begin{aligned} \Delta Rate_c &= \gamma \cdot \Delta m_c + (\lambda_{t=1} - \lambda_{t=2}) + (\epsilon_{ct=1} - \epsilon_{ct=2}) \\ &= \mu + \gamma \cdot \Delta m_c + \eta_c. \end{aligned} \tag{3.2}$$

As mentioned before, researchers have typically used some form of Wilson's MMPI as a measure of the level of male economic status within a marriage market. While Wilson and others have used employment status as an indicator of marriageability, Wood argues that male income levels may be a more informative characterization of a marriage market. As such I estimate average male earnings per county as the measure of economic conditions in the local marriage market.

One problem with using average earnings as a regressor is the that of the previously mentioned endogeneity. Either due to spurious correlation with some market level characteristic that cannot be captured by market fixed effects, the simultaneity of male earnings and marital status, or the correlation between male and female earnings, OLS estimates of γ may be biased. To overcome these biases, I use as an instrument for male earnings the proportion of overall employment in the county that is in the mining industry[6] in a base period (I use 1970) interacted with time to represent the changes in economic conditions induced by the oil price shocks in the 1970s and 1980s. This IV method is similar to the use of changes in the steel and coal industries as instruments for local economics conditions by Black, Daniel, and Sanders (2002) and Black, McKinnish, and Sanders (2003).

The proportion of mining employment in a county is correlated with the change in average male earnings because the oil boom and bust represented periods of high and low demand for labor in the petroleum industry, which is predominantly male. This creates a first stage for the two-period, undifferenced model as follows:

$$m_{ct} = \pi_0 + \pi_1 \cdot \textit{1970 Mining Share of Employment}_c \cdot I(t=2) + \delta_c + \lambda_t + \varepsilon_{ct} \quad (3.3)$$

Using first differences, the first stage estimation equation becomes

$$\Delta m_c = \kappa + \pi_1 \cdot \textit{1970 Mining Share of Employment}_c + \nu_c. \quad (3.4)$$

To demonstrate how the oil boom and bust affected average earnings I first describe the effect on individual workers' earnings. Table 1 describes the characteristics of workers in Texas, highlighting those in the petroleum industry, for 1970, 1980, and 1990. The petroleum workers in Texas are disproportionately male, over 80% in all three years, compared to the overall workforce. While the skill distribution of petroleum workers is similar to that of the overall workforce, workers at all skill levels in the petroleum industry had significantly higher yearly earnings than their contemporaries in all three years.

More important than these within-year differences is how these differences grew during the oil boom and busts. Table 2 reports simple differences-in-differences esti-

[6] For Texas petroleum extraction and production represents the overwhelming majority of the mining industry in terms of both employment and earnings.

mates of the effect of the boom and bust on annual earnings of petroleum workers. As reported in column (6), the earnings of petroleum workers increased by nearly 6% more than those of other workers in Texas from 1970 to 1980. These effects are more pronounced for low skilled workers. Petroleum workers who were high school graduates experienced earnings increases of over 12%, while the earnings of with eight or less completed years of schooling increased by more than 8%. Conversely, during the oil bust these workers experienced significant decreases in earnings. From 1980 to 1990, the earnings of petroleum workers with eight or less years of schooling decrease by 19% more than those of other workers with the same level of education. The earnings of those with nine to eleven completed years of schooling decreased by 16% more.

Due to differences in the concentration of oil workers, these changes in earnings had differential effects across counties in Texas. While the overall share of mining employment in Texas in 1970 was only 2.75%, counties located near large oil and natural gas reserves had much higher concentrations of mining employment. Figure 1 indicates the relative concentration of mining employment for Texas counties in 1970. There are 254 counties in Texas, for 44 of which there are no mining employment data reported in the REIS due to confidentiality.[7] Of the remaining 210 counties, 89 have less than 2% of its workers employed in the mining industry. 87 counties have a mining share from 2% to 10%, while the remaining 78 have more than 10% of their workforce in the mining industry.

Figure 2 depicts the first stage relationship between the oil price shocks and average county earnings in Texas graphically. It plots the time series of average earnings of counties grouped by the mining industry's share of employment in 1970. While level differences between groups are attributable to numerous factors, the changes in average earnings in response to the oil price shocks are readily apparent. Beginning with the first oil price shock in 1973 and continuing through subsequent price shocks

[7]The REIS does not report employment levels for an industry in a county when there are less than ten workers employed in that industry or when all of the workers are employed by a single employer.

in the late 1970s, the proportional changes in average earnings in counties with higher concentrations of mining employment were much larger than in those counties with significantly less mining employment. Even more dramatic is the decline in average wages after 1981 in the oil counties compared to those counties with low mining employment, where there was practically no change in average earnings from 1981 to 1990.

For the long differences strategy I estimate average male earnings and marriage rates for 1970, 1980, and 1990. In addition to facilitating the use of census data, these years correspond well to the oil boom and bust. 1970 is well before the first price shocks in 1973, and 1980 corresponds to the height of the oil boom (1979 is the year of highest average as demonstrated in Figure 3) and precedes any warnings of an end to the boom. Lastly, because average earnings had stabilized by then, 1990 well represents the end of the boom.

For these years I calculate average male earnings in a county using county level earnings and employment data from the Bureau of Economic Analysis's Regional Economic Information System (REIS)[8] and Census microdata (PUMS).[9] From the microdata I calculate the relative male wage income and proportion of male employment by industry[10] for Texas in 1970, 1980, and 1990. Then, using the county level earnings and employment in each industry, I estimate average male earnings, $\overline{w}_{ct}^{m}$, for each county c and census year t based upon the following relationship:

$$\overline{w}_{ct}^{m} = \frac{\sum_{j=1}^{J} \left(\frac{\overline{w}_{jt}^{m}}{\overline{w}_{jt}} \right) \overline{w}_{cjt} \cdot \rho_{jt} \cdot Employment_{cjt}}{\sum_{j=1}^{J} \rho_{jt} \cdot Employment_{cjt}} \tag{3.5}$$

for $\overline{w}_{j}^{m}$ and $\overline{w}_{j}$ represent the average male earnings and overall earnings, respectively,

[8]The REIS data are collected from state-level unemployment insurance filings from businesses.

[9]Because of limitations in the geographic coding of PUMS data, it is not possible to simply use this microdata. The PUMS identifies county groups rather than individual counties. These groups can be geographically large and do not correspond very well between census periods. Therefore, I use the aggregate data afforded by the REIS.

[10]The REIS data reports employment and earnings data by nine major industrial categories: mining, construction, manufacturing, transportation and utilities, wholesale trade, retail trade, financial services, other services, and public administration.

in industry j in Texas and ρ_{jt} is the proportion males employed in the industry. Thus, given these industry characteristics from the PUMS and the relative sizes of the industries in each county from the REIS, I can estimate average male earnings for each county.

I calculate the hazard rate into marriage using marriage records from the Texas Department of Health from 1969 to 1991[11] and county level estimates of female population and marital status from the 1970 Census Fourth Count and the 1980 and 1990 Summary Tape Files. Unlike most cross-sectional data, which indicate only whether a women is married, the vital statistics records show contemporaneous marriage flows for a county, and the hazard rates into marriage can be expressed as the number of new marriages[12] per 1000 unmarried women. However, this metric may not precisely capture the hazard rate of unmarried women into marriage because it does account for divorced women and their marriages. This should be of minimal significance for the 15 to 24 year age group; however, the same should not be assumed for the 25 to 34 year age group.

Table 3 reports the average county earnings and the marriage rates of women age 15 to 24 and 35 to 34 in Texas for1970, 1980, and 1990. Before the oil boom average male earnings were \$37,700 (in 2000 dollars), while at the end the average was \$40,740. For each of these years, the large standard deviations indicate significant variation in average earnings across counties. During this period the proportion of women age 25 to 34 that were ever married decreased from 92% to 81%. Besides an overall decrease in marriage rates of women, the decrease in the proportion of women age 15 to 24 ever married from 42% to 27%, a decrease by nearly one third, indicates that women also waited longer to marry. In terms of hazard rates into marriage, marriages of 15 to 24 year old women decrease from 170 marriages per 1000 never married women to 88 marriages. Similarly, the marriages of women age 25 to 34 decreased from 355 to

[11] I excluded the years 1972 and 1986 because the marriage records for September through December of 1972 are missing from the data and the 1986 records as missing nearly 20,000 records for which I have found no systematic explanation.

[12] I tabulate the number of marriages per year per county by the age of the bride and average these over a three year period to reduce any high-frequency "noise" in marriage rates.

199. The large standard deviations of these marriage rates indicate large cross-county differences which may be explained in part by differing economic conditions.

3.3 Findings and Discussion

3.3.1 Empirical Results

Table 4 reports the OLS and IV estimates of the effect of average male earnings on the marriage rates of women age 15 to 24 and 25 to 34 for the boom and bust periods, columns (1) through (4) and (5) through (8) respectively. Estimates in panel A are for the simple specifications described by equations (2) through (4). In Panels B and C, I include controls for county population and the male-to-female ratio. Including these additional covariates have important implications for the estimations and provide some information regarding the social forces behind the change in marriage rates, which I discuss later.

The OLS estimates of a regression of the log difference in the number of marriages on the log difference in earnings as specified by equation (2) are reported in columns (1) and (5). All of these estimates are weighted by the county's population of never married women in the respective age group, and heteroskedasticity robust standard are also reported. As discussed before, these estimates may be biased downwards due to normal measurement error attenuation or due to bias from the positive correlation between male and female earnings. In nearly all cases they are small and insignificant. For example, the OLS point estimate in the simple specification for women age 15 to 24 in the boom period is 0.423 indicating that a 10% increase in the average earnings per job corresponds to a 4.2% increase in the number of marriages. The corresponding OLS estimate for the bust period is smaller, 0.288.

Column (2) and (6) report the reduced form estimates, and they indicate a strong correlation between the proportion of a county's employment that was in the mining industry in 1970 and the change in marriage rates as specified by equation (2). For example the reduced form estimates for women age 15 to 24 for the boom and bust period are 1.555 and -1.186, indicating that a county with a 10 percentage point higher proportion of employment in mining than other counties had a 15.6% increase in marriages associate with the oil boom, while that difference in mining employment corresponds to an 11.9% decrease in marriages during the bust period.

The mechanism through which the concentration of mining employment and the timing of the oil boom and bust could have affected female marriage rates is through male earnings. Columns (3) and (7) report the first stage estimates of the effect of the oil boom and bust on the average male earnings of counties with higher proportions of mining employment as specified by equation (3). The regression coefficient of 0.571 indicates that a 10 percentage point higher proportion of employment in mining corresponds to a 5.7% higher average earnings per job at the height of the oil boom. The corresponding estimate for the bust period is -1.121 indicating an 11.2% decrease in male earnings for a county with a 10 percentage point higher proportion of employment in mining than other counties.[13] These point estimates demonstrate that the oil price shocks had a significant impact on male earnings in counties where there were higher concentrations of mining employment. They also indicate that there is a strong enough correlation between the 1970 proportion of mining employment in a county and the change in the county's average earnings during the oil boom and bust for the proportion of mining employment to serve as an instrument for average male earnings.

Finally, the IV estimates of the effect of average male earnings on marriage rates are reported in columns (4) and (8). For the younger women in the boom period the IV estimate is 2.725, and is statistically significant. The estimate implies that a 10% increase in average male earnings during the boom corresponds to 27% increase in marriage rates for women age 15 to 24. The estimate for the bust period over half of that – a point estimate of 1.058 implying that a 10% decrease in average earnings during the bust resulted in a 10.6% decrease in marriage rates for the younger women.

An assumption of the simple first differences specification is that the additive effect of time on marriage rates is uniform across all counties, except as explained

[13]Differences in these estimates for the two age groups is a function of the weights.

by other covariates. Thus, while the level of marriage rates for any specific county are accounted for by differencing out the individual fixed effects, differences in the change in marriage rates appear as stochastic errors. However, if these changes are spuriously correlated with the instrument, the IV estimates will be biased. While the location of oil reserves and consequently the location of petroleum employment are *prima facia* exogenous to marriages rates, two factors prove worthy of consideration: county population and the male-to-female ratio in a county.

Much of the petroleum production in Texas occurs in rural counties, and even if mining employment were uniform across all counties, the proportion of mining employment in counties with larger populations would be smaller. As such the proportion of mining employment is negatively correlated with county population. If the decline in marriage rates were larger in urban areas relative to the rural areas, regardless of economic conditions, the IV estimates will be biased upwards. To correct for this I add as a covariate the natural log of the number of women in the respective age group in the county. Panel B of Table 4 reports the estimates for the specifications which include controls for population. In all cases the reduced form estimates are smaller in magnitude implying that population may have affected the change in marriage rates. Consequently, the IV estimates of the effect of male earnings on marriage rates are smaller – 1.617 for 15 to 24 year old women in the boom period and 1.451 for the bust period – though they are still significant. They imply that a 10% increase in average male earnings corresponds to 15% increase in marriage rates for young women.

The regression coefficient for population for women age 15 to 24 during the boom period is -0.05 (standard error is 0.01). Thus, a county that was twice as large in terms of population as another county experienced a 5% decrease in marriage rates relative to the smaller county during the 1970s. The coefficient for women age 25 to 34 is -0.12 (standard error is 0.02). Therefore, the social causes of the decline in marriage, such as changes in societal norms and increased opportunities for women, were more pronounced in urban areas than in rural areas.

Another potential correlation between the concentration of mining employment and marriage rates results from the fact that the oil shocks may have induced the migration of males to and from these oil counties changing the male-to-female ratios in the local marriage markets. Other marriage market research has demonstrated a positive, although small, relationship between the ratios of adult males and females to the propensity for a women to be married. Angrist (2002) exploits variations in the sex ratios of different immigrant populations in the United States in the early 1900s and estimates that an increase in the ratio of males to females in a marriage market from 1.00 to 1.25 will increase the probability that a women age 18 to 33 will be married by 3 percentage points, or 6% of the average marriage rate. If prospects of higher earnings induced males to migrate to the oil counties during the oil boom, then the IV estimates in this study might be biased upwards.

While the oil counties did experience increases in the ratios of men to women, these increases were small. The regression coefficient for the effect of the proportion of mining employment in the county in 1970 on the log difference in the ratio or 15 to 29 year old males to females (weighted by the population of 15 to 29 year old females) for the boom period is 0.508. The implies that a county with a 10 percentage point higher proportion of employment in mining than other counties had a 5.1 percentage point increase in the sex ratio associated with the oil boom. The corresponding estimate for the bust period is -0.198, implying a 2.0 percentage point decrease in the sex ratio. Although the estimates are statistically significant, based upon Angrist's estimates, these changes in sex ratio should not large enough to induce significant changes in the marriage rates of a county. Ignoring some of its second order biases due to the endogeneity of the sex ratios, controlling for the ratio of men to women in the respective age group in a county in the regression eliminates much of the bias due to induced migration and changes in sex ratios. These estimates are reported in Panel C and are similar to those when sex ratio is not included as a regressor. Except for women age 25 to 34 in the boom period, all of the IV estimates are well within a standard error of the corresponding estimate in Panel B, implying that changes in marriage rates is not due to male migration and changes in sex ratios in a local marriage market.

Overall, I conclude that by including controls for population, the specifications in Panel B produce accurate estimates of the effects of changes in male earnings on marriages rates, a 15% increase in marriage rates for a 10% increase in average male earnings. The IV estimates for these specifications are higher than the OLS estimates – the IV estimate for women age 15 to 24 for the boom period is significantly different than the OLS estimate, while the difference for the bust period is marginally significant – implying that the OLS estimates indeed are biased downwards. As previously discussed, this downward bias may be due to normal measurement error attenuation or due to bias from the positive correlation between male and female earnings.

Interestingly, these estimates are similar for the boom and the bust periods. This symmetry suggests that individuals are not overly optimistic compared to being overly pessimistic regarding future earnings opportunities in their decisions to marry. Additionally, the estimates for the older age group of women are slightly smaller than those for the younger group, although these differences are insignificant. However, because of the higher standard errors, these estimates are either insignificant or marginally significant. That older women are less affected by changes in male earnings may simply be a result of the "marginal' woman marrying earlier, changing the composition of single older women relative to single younger women.

3.3.2 Different Measures of Marriage Rates

A feature unique to this study is the use of marriage records to measure the hazard rate, or flows, into marriage rather than the traditional use of marital status, a stock measure. Using the same empirical strategy used for marriages rates, I can estimate the effects of male earnings on the marital status, whether a woman has ever married, and compare the estimates with those for the flows into marriage. For these estimates I regress the proportion of women ever married on average male earnings weighted by the total population of women in the respective age group for the boom and bust periods. These results are reported in Table 5.

The IV estimate for women age 15 to 24 for the boom period is 0.721 implying that a 10% increase in male earnings during the boom resulted in a 7.2 percentage point increase in the proportion of women ever married. The corresponding estimate for the bust period is 0.516, though it is not significantly, implying fairly symmetric effects of increases and decreases in earnings. For both periods the OLS estimates for women age 15 to 24 are smaller than the IV estimate; it is significantly different for the bust period. Again, this may result from attenuation bias or the downward bias due to the correlation between male and female earnings.

However, because these estimates measure the stock of marriage, they are not directly comparable to the hazard rate measures. However, again applying these estimates to the summary data I can compare the contributions of the decrease in male earnings from 1980 to 1990 to the decrease in marriage rates as measured by these estimates to those above. The average of the two estimates for the 15 to 24 year old women, 0.620, predicts that a 3.6% decrease in average male earnings will result in a 2.2 percentage point decrease in the number of women ever married. This explains approximately 22% of the decrease in the proportion of women aged 15 to 24 ever married from 37.2% to 27.0%. This is comparable to the estimate using the marriage flows.

The estimates for the older age group, however, are nearly equal to zero, because there are many different factors over time which may affect whether a woman age 25 to 34 is has married. The fact that these estimates are negative is likely to be a function of business cycles; a contemporaneous period of high (low) male earnings may have been preceded by a corresponding period of lower (higher) earnings, which nay have affected a woman's decision to marry at the time. These estimates highlight the caution that should be exercised in using stock measures of marriage for older women and the usefulness of the flow measures of this study.

3.4 Implications and Conclusions

Using the oil boom and bust in Texas from 1970 to 1990, which represent shocks to earnings in the predominantly male petroleum industry, to instrument for average male earnings in Texas counties, I estimate that a 10% increase in male earnings results in a 15% increase in the hazard rate of never married young women into marriage. While these estimates do indicate a downward bias, they do not indicate a significantly large contribution of changes in male economic status to the decline in marriage rates over the past several decades.

It is a simple exercise to apply these estimates to the summary data in Table 3 to determine the contributions of changes in male earnings on the overall decline in marriage rates. For example, from 1980 to 1990 male earnings in Texas decreased by 3.6% from $42,280 to $40,740. During this period the hazard rate into marriage of women age 15 to 24 decreased 31.3% from 127.8 marriage per 100 never married women to 87.8. The decrease in male earnings predicts a decrease in marriage rates of approximately 5.5%, which would explain about 18% of the overall decrease in marriage rates in Texas. This is larger than the estimated effects of changes in male economic conditions found in other studies, such as Ellwood and Rodda (1991), Mare and Winship (1991), and Wood (1995). However, even if the IV estimates in these studies are biased downward, the changes in male earnings still cannot account for the majority of the decline in marriage rates.

Yet, questions of transferability are justified. Are these results generalizable? Because the instrument works through the wages of petroleum workers, the local average treatment effect of changes in male earnings (despite some spillover to other industries) applies to women who are likely to marry the class of men to which petroleum workers in Texas belong. As discussed before, Texas petroleum workers are predominantly white males, and although their educational attainment is very similar to that of the entire workforce, those with twelve or fewer grades of education experienced the greatest change in earnings during the oil boom and especially during the bust. However, these workers even the less education enjoyed higher overall wages compared to others with the same levels of education.

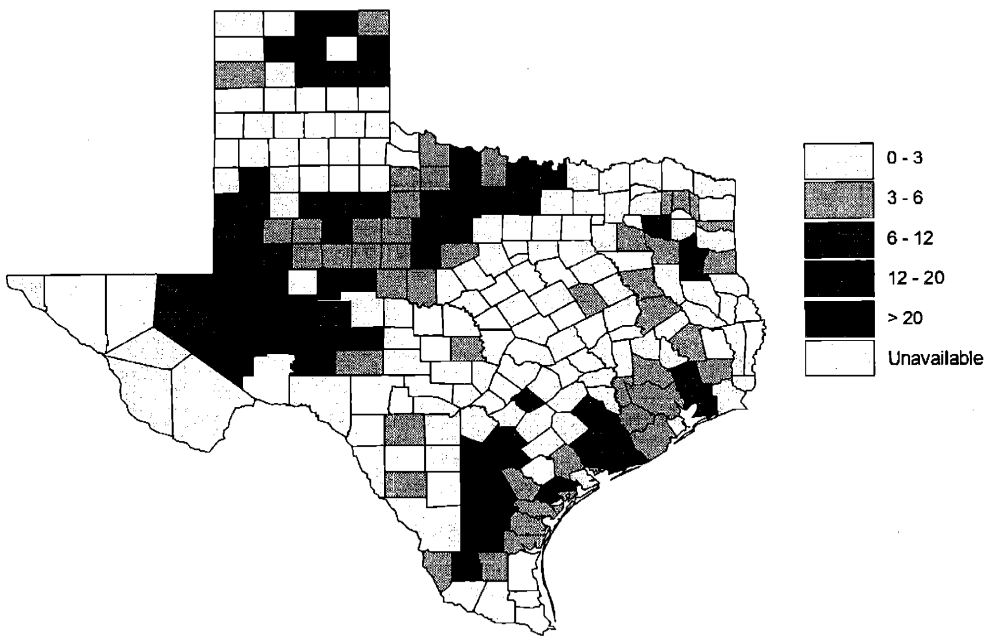
So while these results may apply to many regions it is questionable as to whether they reflect the potential response from inner-city black males who are generally less educated and earn very low wages. Wilson (1996) estimates that real earnings decreased by 30% for the bottom fifth of the income distribution of males from 1970 to 1989,[14] implying a 22.5% decrease (15% x 1.5) in marriage rates for the lowest income groups due to changes in male earnings for each decade. Even if all blacks fell into this income bracket (which they obviously do not), this decline in male earnings could explain only half of the decline in marriage rates of black women age 15 to 24 from a 1970 married rate of 21.3% to and 1980 rate of 13.1%.[15] The relationship between male earnings and marriage rates for inner-city blacks must be significantly higher than these estimates if decreases in male economic status are to explain the majority of the decline in marriage rates as Wilson claims.

Thus, while these estimate indicate are stronger relationship between male economic status and marriage rates than previous studies, certainly much stronger forces have been the impetus of the significant declines in marriage rates over the past few decades. Understanding and measuring the effects of the forces may provide greater insight into the plight of family dissolution in inner cities.

---

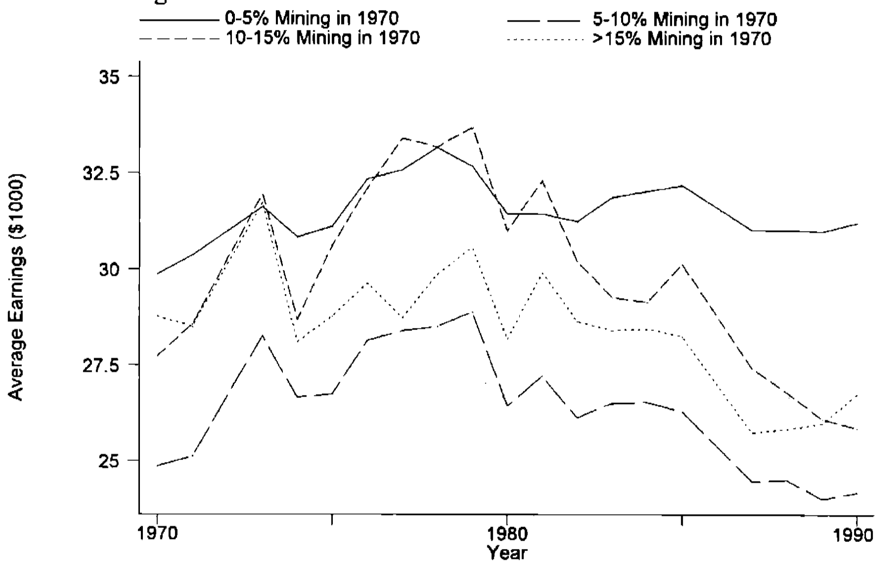
[14] Wilson (1996), p. 25.
[15] Wilson (1987), p. 69.

Figure 1. Percentage of County Workers Employed in Mining Industry in 1970



Source: BEA Regional Economic Information System

Figure 2.



0-5% Mining in 1970
5-10% Mining in 1970
10-15% Mining in 1970
>15% Mining in 1970
35
32.5
30
27.5
25
Average Earnings ($1000)
1970
1980
1990
Year

Table 1
Characteristics of Texas Workers

	1970		1980		1990	
	Petroleum Workers (1)	All Workers (2)	Petroleum Workers (3)	All Workers (4)	Petroleum Workers (5)	All Workers (6)
	A. Make-up of Labor Force					
N	4,809 (0.62%)	773,356	7,471 (0.75%)	1,002,124	6,206 (0.52%)	1,188,973
Gender						
Male	86.7%	62.8%	83.6%	58.0%	81.1%	54.8%
Female	13.3	37.2	16.4	42.0	18.9	45.2
Race						
White	94.8%	88.9%	91.4%	87.1%	91.3%	87.2%
Black	4.3	9.8	6.8	10.3	6.0	9.2
Other	0.9	1.3	1.8	2.6	2.7	3.6
Years of Schooling						
Less than 8 years	15.5%	18.0%	8.1%	9.1%	4.6%	4.9%
9 to 11 years	17.2	18.7	13.6	13.1	8.6	8.2
12 years	36.1	36.1	38.1	39.3	35.4	34.8
1 to 3 years of college	14.8	14.1	20.0	19.9	27.2	29.8
4 years of college	16.5	13.1	20.1	18.7	24.2	22.4
	B. Average Yearly Earnings (2000$)					
Years of School						
Less than 8 years	32,385	21,147	35,138	19,922	24,312	16,241
9 to 11 years	36,898	24,784	36,025	22,101	25,079	18,041
12 years	37,254	27,137	40,135	25,298	35,042	22,897
1 to 3 years of college	40,066	29,101	41,526	27,366	40,622	27,474
4 years of college	65,414	45,194	62,939	40,719	69,646	45,869
Total	41,493	28,253	44,141	27,682	43,591	28,685

Notes: This table reports characteristics of workers in Texas from the 1970, 1980, and 1990 5% PUMS Census data.

Table 2
Differences-in-Differences Estimates of the Effects of the Oil Boom and Bust
on Employment Outcomes for Petroleum Industry Workers

	Boom Period – 1970-1980				Bust Period – 1980-1990			
	All Adults		Working Adults		All Adults		Working Adults	
	Wage Income (1)	Log Income (2)	Wage Income (5)	Log Income (6)	Wage Income (5)	Log Income (6)	Wage Income (7)	Log Income (8)
Total	4,338	0.017	3,059	0.058	-2,805	-0.048	-1,079	-0.033
	(408)	(0.021)	(479)	(0.018)	(424)	(0.020)	(504)	(0.018)
Grades 0-8	6,385	0.029	3,743	0.083	-7,324	-0.205	-6,547	-0.191
	(754)	(0.062)	(1,067)	(0.056)	(917)	(0.084)	(1,443)	(0.077)
Grade 9-11	3,805	0.040	1,568	0.052	-8,553	-0.231	-6,533	-0.160
	(829)	(0.053)	(981)	(0.047)	(864)	(0.063)	(1,126)	(0.057)
Grade 12	5,150	0.089	4,781	0.125	-3,368	-0.037	-2,170	-0.023
	(603)	(0.032)	(658)	(0.028)	(550)	(0.030)	(618)	(0.027)
1-3 Years of College	3,756	-0.007	3,133	0.040	-1,612	0.015	-810	-0.008
	(988)	(0.051)	(1,172)	(0.047)	(816)	(0.040)	(926)	(0.036)
4+ Years of College	1,010	-0.031	1,867	0.030	-1,223	-0.010	1,878	0.009
	(1,497)	(0.047)	(1,558)	(0.042)	(1,389)	(0.039)	(1,492)	(0.035)

Notes: This table reports simple differences-in-differences estimates of the change in previous year's earnings of workers in the petroleum industry compared to all other workers for the periods of 1970-1980 and 1980-1990. Earnings data are from the 1970, 1980, and 1990 5% PUMS Census data. Heteroskedasticity robust standard errors are reported in parentheses.

Table 3
Average County Earnings and Marriage Rates

	Average Earnings		Marriages per 1000 Never Married Women		Proportion Ever Married	
	All Workers	Males	Age 15-24	Age 25-34	Age 15-24	Age 25-34
Year	(1)	(2)	(3)	(4)	(5)	(6)
1970	29.81	37.70	169.9	355.2	0.420	0.922
	(5.11)	(5.43)	(71.1)	(263.1)	(0.070)	(0.029)
1980	32.47	42.28	127.8	325.5	0.373	0.877
	(6.17)	(7.17)	(70.4)	(267.9)	(0.062)	(0.042)
1990	31.80	40.74	87.8	198.8	0.270	0.809
	(6.81)	(8.26)	(41.6)	(130.1)	(0.056)	(0.059)

Notes: This table reports average county earnings and marriage rates for Texas counties in 1970, 1980, and 1990. Earnings data are from the Bureau of Economic Analysis' Regional Economic Information System, and the averages are weighted by the number of workers per county. Female population and marital status are from the Bureau of the Census summary tape files. Marriages are counted from the Texas Department of Health vital statistics records. The averages for the number of marriages are weighted by the number of never married women per county in the respective age groups, and the proportions of ever married women are weighted by the number of women per county in the respective age groups. Standard deviations are reported in parentheses.

Table 4
First Differences and Instrumental Variables Estimates
of the Effect of Average Male Earnings on Marriage Rates

	Boom Period – 1970-1980				Bust Period – 1980-1990			
	OLS	Reduced Form	First Stage	IV	OLS	Reduced Form	First Stage	IV
Age Group	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. No Additional Controls								
15-24	0.423 (0.441)	1.555 (0.367)	0.571 (0.103)	2.725 (1.070)	0.288 (0.192)	-1.186 (0.356)	-1.121 (0.181)	1.058 (0.333)
25-34	0.008 (0.746)	2.128 (0.830)	0.720 (0.116)	2.957 (1.671)	0.152 (0.292)	-0.936 (0.567)	-1.217 (0.200)	0.770 (0.507)
B. Controlling for Population								
15-24	0.649 (0.340)	1.126 (0.344)	0.696 (0.102)	1.617 (0.723)	0.208 (0.222)	-1.069 (0.348)	-0.737 (0.155)	1.451 (0.591)
25-34	1.040 (0.372)	1.074 (0.604)	0.846 (0.108)	1.269 (0.706)	0.087 (0.359)	-0.871 (0.605)	-0.957 (0.166)	0.910 (0.787)
C. Controlling for Population and Male-to-Female Ratio								
15-24	0.523 (0.312)	0.940 (0.293)	0.658 (0.102)	1.429 (0.697)	0.190 (0.225)	-1.073 (0.347)	-0.738 (0.155)	1.455 (0.587)
25-34	0.979 (0.384)	0.811 (0.612)	0.647 (0.110)	1.254 (0.954)	0.037 (0.339)	-0.820 (0.584)	-0.760 (0.162)	1.080 (0.909)

Notes: This table reports OLS and instrumental variables regression coefficients for a regression of the log difference in the number of marriages per 1000 never married women in the respective age group on the log difference in average male earnings as described by equations (2) through (4) in the text. Estimates are weighted by the county level mean number of never married women in the respective age group. Heteroskedasticity robust standard errors are reported in parentheses.

Table 5
First Differences and Instrumental Variables Estimates
of the Effect of Average Earnings on Marital Status

Age Group	OLS (1)	Reduced Form (2)	First Stage (3)	IV (4)
		A. Boom Period – 1970-1980		
15-24	0.370	0.460	0.638	0.721
	(0.089)	(0.110)	(0.097)	(0.185)
25-34	-0.007	-0.020	0.621	-0.032
	(0.038)	(0.046)	(0.094)	(0.076)
		B. Bust Period – 1980-1990		
15-24	0.119	-0.351	-0.681	0.516
	(0.050)	(0.103)	(0.146)	(0.133)
25-34	-0.060	0.041	-0.721	-0.056
	(0.034)	(0.033)	(0.147)	(0.042)

Notes: This table reports OLS and instrumental variables regression coefficients for a regression of the log difference in the proportion of women ever married in the respective age group on the log difference in average male earnings as described by equations (2) through (4) in the text. Estimates are weighted by the county level mean population of women in the respective age group. Heteroskedasticity robust standard errors are reported in parentheses.

Chapter 4

The Effect of Pay Spreads on Tournament Selection: Evidence from Enlisted Personnel in the U.S. Army

4.1 Introduction

The tournament model literature describes the many different incentives affecting individual performance that paying individuals differently based upon their relative levels of output may create.[1] In structuring a tournament-based compensation scheme, the principle's primary tool is the spread of pay between different levels of relative performance. In addition to motivating effort, these pay spreads may induce the self-selection of individuals into different tournaments or "leagues" of relative ability.

[1]Examples of the groundbreaking theoretical work on tournament models and effort incentives under compensation schemes based upon relative performance are Lazear and Rosen (1981), Nalebuff and Stiglitz (1983), Green and Stokey (1983), O'Keeffe et al. (1984), and Rosen (1986), while McLaughlin (1988) is a good survey of this literature. Carmichael (1983) and Malcomson (1984) provide theoretical arguments for the existence of optimal effort-inducing contracts using promotion ladders. Asch (1993) and Asch and Warner (1994a) summarize the theoretical literature as is pertains to effort incentives and ability sorting in hierarchical organizations with specific applications to the military.

Such an implication is particularly important for organizations with internal labor markets that, unlike typical firms, face many different problems in placing high quality individuals into the organizations' highest ranks. An organization with a vertical hierarchy may have skill requirements for supervisory roles specific to that organization thereby placing restrictions on lateral entry. Because of these limitations, large hierarchical firms, such as the military, are forced to promote from within. In order to place high ability individuals into more important, higher-level positions, such a firm must first attract a sufficient number of high ability individuals into its entry-level positions. Second, the firm must provide sufficient incentives to retain enough high ability individuals throughout their careers with the firm to fill the higher-ranking positions. The firm's compensation scheme may be one of its most important tools for affecting the retention of individuals of different abilities. In this paper, I examine the question of how pay spreads in a hierarchical organization's tournament-like compensation scheme affect ability sorting, in terms of the retention of individuals in the organization, using reenlistment decisions of enlisted personnel in the U.S. Army.

Lazear and Rosen (1981) first discussed the sorting of individuals into different leagues and the issues of adverse selection affecting rank-order tournaments involving heterogeneous contestants. They showed that with a sufficiently large pay spread workers will sort themselves based upon relative ability, though the pay spread will induce inefficiently high effort from the high ability workers. O'Keeffe et al. (1984) argue that given proper monitoring precision, first-best marginal incentives can be maintained and that proper self-selection can be induced by increasing or decreasing the pay spreads in the contests. Specifically, decreasing the pay spread in contests designed for low ability workers can induce high ability workers to self-select into the higher league, and increasing the pay spread in the high ability contest can induce the low ability workers to self-select into the lower league. Thus, larger pay spreads in compensation schemes attract higher ability workers.

Empirical tests of the predictions of tournament models, though not sparse, have been primarily limited to analyses of sporting events with mixed findings.[2] Addi-

[2]Examples include Ehrenberg and Bognanno (1990a and 1990b) which found that higher prize

tionally, most of these have only examined the effects of prize spreads in contests on player performance without examining the ability sorting induced by pay spreads. One exception to this is Ehrenberg and Boganno (1990b), in which the authors find that higher prize money increases that probability that exempt (high ability) golfers enter a tournament and decreases the probability that non-exempt players enter. Another is Lynch and Zax (1997), which uses data from the USA Track and Field Road Running Information Center and finds that in repeated-play tournaments when there are more tournaments than an individual can possibly enter, forcing him to select amongst several different tournaments, prize spreads will induce ability sorting.

In this study, I analyze the retention decisions of enlisted personnel in the U.S. Army to examine the effects of pay spreads on ability sorting in hierarchical organizations. At the macro-level, personnel management in the all-volunteer military is not unlike that of a typical firm or organization – individuals are hired and promoted, some choose to leave or are dismissed creating turnover, and others stay until they retire. Due to the sheer size of the military – some 1.4 million active duty military personnel and another 1.3 million National Guard and Reserve volunteers in 2000 – personnel policies within the services are extremely standardized, which greatly simplifies the process of analyzing the effects of personnel policy decisions.[3]

When an individual joins the Army, or enlists, he or she signs a service agreement for a specific length of service, normally from two to six years. The length of the agreement depends upon both personal preference and the availability of positions and programs at the time. Near the completion of his respective service agreement, an individual who is not disqualified due to medical or disciplinary reasons may agree to an additional one-time extension of his service agreement of up to one year, or he may sign a completely new service agreement for another two to six years.

spreads in golf tournaments may increase player performance using data from the 1984 Men's Professional Golfers Association (PGA) Tour and the 1987 European Men's PGA Tour. Orszag (1994) refutes this finding using data from the 1992 PGA Tour. Becker and Huselid (1992) uses data from the National Association of Stock Car Auto Racing and the International Motor Sports Association to show that higher prize spreads increased driver performance, though at a diminishing rate.

[3]For an excellent summary of enlisted personnel management policies and issues see Kirby and Thie (1996).

There is a reenlistment "window" extending from one year to three months prior to the termination of a service agreement in which an individual may sign completely new agreements, and those who do are considered to have reenlisted. Reenlistment decisions can be viewed as tournament choice decisions in which an individual decides to participate in the tournament by reenlisting or to not participate by not reenlisting.

The military has conducted extensive research on the enlistment and retention decisions of service members since the end of the military draft and the creation of the all-volunteer force in 1974.[4] Yet, despite the extensiveness of the retention literature, the majority of the research has focused mainly on aggregate supply elasticities without paying much attention to individuals' ability levels. In these studies, ability, as quantified by score on the Armed Forces Qualification Test (AFQT), is usually added to the regression equation with other demographic characteristics simply as an additional control or "taste" variable with inconclusive implications. In one exception, a study by Buddin et al. (1992) on the effects of promotion rates on retention found that not only are retention rates sensitive to promotion rates but that when promotion rates are added as an additional covariate, the effect of AFQT scores on retention are statistically significant and negative. They suggest that promotion rates could be used as a tool for encouraging and rewarding high achievers by increasing their effective relative pay.

In this study, I provide greater empirical support for this conclusion by looking at the retention rates of different ability levels of individuals, not simply by using ability as an additional "taste" variable but as it is related to pay spreads created by different promotion rates.[5] Instead of using promotion rates as a covariate, I

[4]Hogan and Gilroy (1991) provide a survey of the literature and the different methodologies used.

[5]Promotions in the Army are heavily managed in order to maintain the force's desired composition in terms of rank and experience distributions. Promotions at the lowest levels are relatively decentralized and dependent upon an individual's time served in each grade. However, promotions are increasingly centralized in the higher ranks, which allows the Army headquarters to appropriately shape the composition of the force to its desired structure. Beyond the lowest levels, selection for promotion is a tournament competition based upon relative performance. Eligible individuals are compared in terms of their potential for continued service and increased responsibility, and the most qualified are selected for promotion into increasingly fewer higher-ranking positions. Essentially, these competitions only occur within the separate military occupational specialties (MOS) of the Army. Thereby, cooks only compete with other cooks for promotion, and infantrymen only

quantify the effect of promotion rates in terms of expected total accumulated career pay. Differences in promotion rates across levels of performance lead to pay spreads between different levels of relative performance. The downsizing of the U.S. Army from an enlisted strength of over 650,000 in the late 1980s to nearly 400,000 by 2000 caused great turbulence in the promotion rates of different occupational specialties in the military. I exploit the variations in promotion rates during this period, which created differences in expected future career paths, and thus different pay spreads faced by individuals at the times of the completion of their enlistment contracts. I then analyze the effects of these pay spreads on the decisions individuals of different ability levels whether or not to reenlist in the Army. Additionally, I account for potential endogeneity in the pay using instrumental variables techniques.

My findings indicate that higher ability individuals are less likely to leave the Army than lower ability individuals when there is an increase in the spreads in the expected career pay. In other words, not only is an individual of higher ability more likely to be retained when there is an increase in the average level of pay, he is more likely to be retained when there is an increase in the pay of the higher relative performers. Conversely, a lower ability individual is less likely to be retained when there is greater spread in the compensation scheme. Thus, by adjusting the effective pay spreads in the Army's compensation scheme, policy makers can use personnel and pay policies in conjunction to affect not only overall retention levels but also the quality of the retained force.

4.2 Empirical Strategy and Identification

Given an ability metric, η, and an expected earnings distribution for the Army's compensation scheme with respect to that ability metric, we can determine the effects of pay spreads on the probabilities of reenlistment of individuals of different ability

---

compete with other infantrymen. The result is that promotion rates may vary from year to year and from MOS to MOS depending upon personnel flows and force restructuring. These promotion rates are widely disseminated amongst service members through active measures such as regular career counseling and by passive means such as the monthly publication of promotion rates in occupational periodicals.

levels by estimating the following conditional expectation function:

$$E[RETAINED_i] = E\left[\begin{array}{c} \beta_0 + \beta_1 \cdot MED\_PAY_i + \beta_2 \cdot \eta_i + \beta_3 \cdot SPREAD_i + \\ \beta_4 \cdot SPREAD_i \cdot \eta_i + \beta_5 \cdot A_i + \beta_6 \cdot X_i \end{array}\right] \tag{4.1}$$

where $RETAINED_i$ is a binary indicator of an individual's retention, MED_PAY_i is the median expected career pay, $SPREAD_i$ is a measure of the pay spread, A_i is a vector of year dummies with interacted ability quantile to account for changes in the distribution of compensation in the best occupational alternative over time absent any explicit quantifications of outside opportunities, and is X_i a vector of individual characteristics that include gender, race, occupational specialty, age, age squared, marital status, number of children, and the unemployment rate at the home of record.

Using a linear probability model, ordinary least squares estimates of the parameter β_4 describe the interacted effect of ability quantile and the pay spread in an occupation's compensation scheme on the probability that an individual will be remain in that occupation. If a relatively higher ability individual is more likely to be retained than a lower ability individual when there is an increase in the expected income of the higher relative performers, then β_4 should be positive.

Other studies on retention in the military have predominantly used two econometric models, the annualized cost of leaving (ACOL) model and the dynamic programming model.[6] The main difference between the two models is that the ACOL model assumes that individuals calculate the value of staying in the military on the basis of a single optimal future date for leaving the military, while the dynamic programming model allows for several decisions throughout an individual's career and models the multiple associated error terms in the stochastic specification of the model. The dynamic programming model is well suited for the analysis of policies that deviate

[6]Gotz and McCall (1984) originally developed the dynamic programming model to study the retention of Air Force officers. However, several others, including Fernandez et al. (1985) and Daula and Moffitt (1991), have extended and modified the model to apply to other military retention issues.

significantly from current policy, while the ACOL model allows for a better examination of the implications of marginal policy changes.[7] While the parameterization described by equation (1) more closely resembles the ACOL model, neither of these models address differential incentives across ability levels.

In this study, I estimate the above parameters using the reenlistment decisions of Army enlisted personnel nearing the completion of their second enlistment contract. The decision to remain in the Army can be thought of as a decision to enter a tournament, one whose payouts for relative levels of performance are characterized by the differences in the career progressions and career pay earned by different individuals. The use of the second reenlistment decision affords several theoretical and practical conveniences. First, because these individuals have served two terms, their decisions to reenlist are less affected by their ability to adapt to the military lifestyle, or "taste", than are the decisions for potential first term reenlistees. As such, these individuals differ very little in terms of the utility that they derive from mere participation in the tournament, which Lynch and Zax (1997) argues can muddle the theoretical predictions of ability sorting. Secondly, because the continuation rates for those who reenlist a second time are well over 90% through twenty years of service,[8] at which point they become eligible for retirement, it is reasonable to assume that when considering to reenlist a second time, individuals consider the choice as a commitment through twenty years, and the calculation of expected career payouts need not include an optimization of expected career pay regarding any subsequent decisions to reenlist.

There are two potential sources of variation in the pay spreads in career earnings that would allow for these estimations. The first is simply changes in the military's pay schedule that would create differences in the earnings for different relative rates of career progression. The second is variations in promotion rates over time that would cause variations in the relative rates of career progression and subsequently in the distribution of earnings. In regards to the first source, there has been almost no

[7]Asch and Warner (1994b), p. 2.
[8]Asch and Warner (1994a), p. 11.

change in the military's pay schedule over the past fifteen years. Except for some targeted pay raises for the year 2000, all pay changes for the military have been universal raises equal to or less than the rate of inflation. Because of this, the ratio of pay between two individuals of different ranks and years of service (YOS) has remained the same.

However, due to the downsizing of the military following the end of the Cold War, there have been significant variations in promotion rates. Figure 1 shows the progression of the drawdown in the Army endstrength during this time frame. Beginning in 1987, Congress ordered the Army to reduce its active duty strength by 31% by 1997. By 1992 the number of active duty members was reduced by 22%.[9] Nearly two thirds of that reduction occurred through voluntary separations after the end of the Gulf War in 1991.[10] By 1993 the Army's combat force shrank from 18 to 14 divisions and then to 10 divisions by 1996. The 1991 National Defense Authorization Act directed that a major target for the reduction should be those with more the twenty years of service. Additionally, the 1993 National Defense Authorization Act allowed services to offer retirement from service with full benefits, though reduced annuities, at fifteen rather than twenty years of service, until 1995. The results of the downsizing, organizational restructuring, and personnel policies included significant variations in personnel requirements and the fulfillment of those requirements at each grade of each occupational specialty, or career management field (CMF), over time.[11] As such, this resulted in significant variations in the promotion rates both amongst CMFs and over time.

Table 1 reports the median promotion rates, in terms of the number of months served in a grade before promotion to the next grade by those promoted, for various CMFs representing both combat arms and support occupations for each fiscal year

[9] Defense Conversion Commission (1993), p. 1-13.

[10] Ibid.

[11] Currently, there are 212 different MOSs in the Army. However, these MOSs are grouped into 32 CMFs. The significance of the CMF is that competitions for job assignments and promotions into the higher enlisted ranks are increasingly managed by CMF as opposed to by MOS. Thus, a convenient method of estimating individuals' career paths is by using the promotion rates within the CMF rather than within each MOS.

during the drawdown.[12] This sample of promotion rates demonstrates that the variations in promotion rates occurred both between years and between CMFs during this time. Table 2 reports the average number of months served in a grade before promotion to the next grade by those promoted, for all of the CMFs. The large variances in the promotion rates for each grade and year indicate that the times served by individuals in a grade can substantially differ depending upon how quickly the individuals are promoted relative to their peers. Additionally, the average promotion rates vary significantly across time. Within this short time period the average promotion rates for E-6s and E-8s differ by as much as a year, and for E-7s by as much as 18 months. Thus, there were significant variations in the shapes of individuals' expected career paths, in terms of time served in each grade over the course of a career, within this short time period. I use these variations in promotion rates as the main source of variation in career earnings and spreads in career pay for different levels of relative performance expected by eligible reenlistees.

For the purposes of this model, the notion of ability does not necessarily have to refer to innate ability. The measure of ability may simply represent a level of human capital or productivity. However, the metric used must be correlated to productivity in the occupation in question and must be comparable across individuals. For this, I use two different measures of ability, AFQT score and the time spent in grade E-4 before promotion to grade E-5.

The AFQT is taken as a part of the recruiting processes prior to each individual's initial entry into the Army and is scored from 1 to 99, where an individual's score represents his rank amongst the percentage of other test-takers. Despite arguments as to the effectiveness of the AFQT in measuring innate ability, this measure is used routinely in research as such a measure, and it is still used by the military in assessing prospective recruits. Furthermore, there have been several studies that have found positive correlations between AFQT and performance on complex individual and team tasks. Fernandez (1992) finds that increases in the AFQT scores of radio operators

[12]The ranks for enlisted personnel range from E-1 to E-9, with E-9 being the highest grade attainable by an enlisted soldier.

are correlated with increases in their performance on various individual and team tasks associated with their MOS, and Scribner et al. (1986) find that increases in each of the AFQT scores of the commander and gunner of an M1A1 tank crew are correlated with increases in the crew's overall performance on qualification courses.

However, because individuals must meet certain skill criteria in order to qualify for specific occupational specialties, the raw AFQT score may not be a comparable measure of relative ability across specialties. For instance, an intelligence analyst with an AFQT score of 85 might not be as high in ability relative to his peers as an infantryman with a similar score. As such, I compare individuals' AFQT scores within each CMF and cohort to obtain a relative ability quantile for each individual. However, for purposes of controlling for year effects and interacted year and ability effects, representing changes in outside opportunities, I use the standard AFQT score as the measure of ability because it remains comparable for the entire sample.

The second ability measure is the time served in grade E-4 before promotion to E-5. The promotion between these grades is the first promotion that is not based mainly on time served in the Army. Before being promoted to E-5, a soldier must be approved for promotion by a board of senior enlisted personnel, and is thereafter selected for promotion based upon an Army-wide comparison of performance measures. Those who are higher performers are therefore promoted more quickly. Because these promotion rates have also varied across time and CMF, I normalize the promotion rates for each CMF and cohort to obtain a comparable, relative promotion rate measured in standard deviations from the mean.

Using these measures, one could estimate the effects of pay spreads on the retention rates of individuals by an ordinary least squares (OLS) estimate of the parameters in equation (1). However, the estimates of the parameters pertaining to expected career pay and pay spreads could potentially be biased due to the correlation of the calculation of expected career pay with an omitted variable that is also correlated with individuals' retention decision. In fact, any omitted factor, such as morale within a CMF, that affects retention is correlated with expected career pay and the pay spreads because promotion rates are affected by the number of individuals competing for selection in the next higher grade. If morale causes individuals to be more likely to be retained, it will cause average promotion rates to be slower and will, therefore, cause expected career pay to be lower. Slower promotion rates for both high and low levels of relative performance would also result in smaller pay spreads. Thus, the OLS estimates for these parameters are potentially biased downwards.

To properly identify the system, I use the changes in the personnel strength levels authorized by the Army Headquarters during the drawdown as instrumental variables for expected career pay and the pay spreads. These changes in authorization levels differed across CMFs and time due to the organizational changes ordered during the drawdown. Some CMFs faced greater reduction than others. Additionally, the sizes of these reductions changed over time because of political changes at the national level. For instance the 1989 Base Force Plan envisioned a reduction in Army endstrength from 771,00 to 536,000 by 1997. Midway through this reduction, the 1993 Clinton Bottom-Up Review ordered the final endstrength to be 495,000 by 1999. The results were numerous changes in the authorization levels for each CMF and for each grade within the CMFs. These authorization levels affect the promotion rates, and thus career pay, because any change in the number of positions available at a grade affects the rate of promotion into that grade; fewer positions mean slower promotion rates. The changes in authorization levels are exogenous to an individual's retention decision in that Congress and the Army Headquarters dictated the changes, and they affect retention only through the changes in expected career pay and pay spreads. The use of these changes as instrumental variables allows for the unbiased estimation of the model's parameters.

4.3 The Data

4.3.1 Data Sources

The data for this study are from several datasets maintained by the Department of Defense's Defense Manpower Data Center (DMDC) and the Army. The first is the

DMDC's yearly Gains and Losses File. This dataset reports information on all individuals who are enlisted into or discharged from the Army during each fiscal year ending the 30th of September. It includes extensive biographical and professional information on each individual. To contractually complete the terms of an individual's service agreement, every individual who reenlists in the Army is honorably discharged and is then reenlisted the following day. This procedure creates two entries in the Gains and Losses File for reenlisting personnel – one for the discharge and another for the reenlistment. As such, I use the Gains and Losses File to create a set of eligible reenlistees and determine whether or not they have reenlisted. Those individuals who are discharged and are not barred from reenlistment due to medical or disciplinary reasons are considered to have been eligible, and if those individuals have a corresponding reenlistment entry in the dataset, they are also considered to have been retained.

The second dataset used in this study is the DMDC's Enlisted Master File, which is a dataset of all enlisted personnel in the Army for each fiscal year. It, too, includes extensive biographical and professional information. I use this dataset for calculating promotion rates. The length of time an individual served in a grade before being promoted to the next higher grade can be calculated from an individual's dates of promotion using two subsequent years' files.

Changes in the authorized personnel strength levels come from the Personnel Management Authorization Document (PMAD), which is changed and published several times during each fiscal year by the Army Headquarters. It consists of the authorized number of personnel in each MOS and grade, and it projects these figures out for each year six years into the future. I use the last PMAD published for the 1988 to 1998 fiscal years and aggregate the authorizations for each MOS into their respective CMF totals to obtain each year's authorized strength level for each CMF.

Other data used in this study include seasonally adjusted unemployment rates, by state, from the Bureau of Labor Statistics (BLS) and annual Consumer Price Index (CPI) rates, also from the BLS, which were used to discount all pay calculations to 1982-1984 dollars.

4.3.2 Eligible Reenlistees and Retention Rates

Table 3 contains the descriptive statistics for those personnel who were eligible for reenlistment at the completion of their second enlistment for fiscal years 1993-1998, a total sample of 154,002 individuals. The demographics of these cohorts are roughly the same for each year. There is a modest increase over the time period in the proportion that has a high-school diploma, and this increase accompanies corresponding decreases in the proportions that have GEDs and no college education and those that have some college education. These cohorts are also fairly similar in terms of individual AFQT scores.[13] Though the percentages of each cohort scoring in each quintile are similar for each year, there is a slight increase in the percentages scoring in the fourth and middle quintiles. This corresponds to a decrease in the percentage scoring in the second quintile. These compositional changes of the yearly cohorts in ability, as measured by the AFQT, and education primarily reflect changes in the demographics of individuals entering the service – through the late 1980s and early 1990s there was a significant increase in the proportion of those entering military service who were high school graduates as opposed to non-graduates or GED holders.

The average number of months served by those eligible for reenlistment is 86.7 months (just over seven years) with a fairly high standard deviation of 20.7 months. This variance is a result of differences in the terms of individuals' enlistment contracts. Service contracts range from two to six years in length. Therefore, the time an individual has served by the time he becomes eligible for reenlistment near the end of a second contract may vary from three to twelve years. Variations in contract lengths are the result of both individual preferences and different service agreements offered by the Army at the times of an individual's initial enlistment and subsequent reenlistments. These differences in time served and in individuals' rates of advancement lead to variations in individuals' grades at the time of eligibility. The majority of the eligible reenlistees is in the grade E-5. Those who are in the grades E-1 through E-3

[13]The number scoring in the bottom quintile is small because federal law prohibits the enlistment of an individual who scores below 10 on the AFQT and limits the number of those scoring below 20 who are allowed to enlist to 10% of each year's total number of enlistees.

are generally individuals who have been reduced in rank for disciplinary reasons but have not committed acts so egregious as to bar them from reenlistment and continued service.

Table 4 reports the proportion of the sample of individuals eligible for reenlistment that actually reenlisted, or was retained. The overall retention rate of individuals completing their second term of service from 1993 to 1998 was 74%. Blacks and Hispanics were retained at a higher rate than whites by an average of six to seven percentage points, while females were retained at a rate of about five percentage points higher than males. The higher rates for minorities and females are most likely the result of a relatively lower number of opportunities and lower levels of equality, whether perceived or real, for women and minorities in the outside labor market compared to those for men and whites. Retention rates tended to be higher for those with more education, although over time the difference in these rates between those with low and those with high levels of education decreased. An examination of the retention rates by individuals' AFQT scores reveals a similar picture. This trend reflects the increases in the education premium through the 1990s, which provided greater incentives for the more educated individuals to leave the military.

Table 5 breaks down the sample of eligible reenlistees by grade and the number of years served and reports each group's retention rate. One major observation that can be made from these statistics is that the retention rates of those individuals with the same number of YOS increase with grade. This can be explained by the fact that an individual is more likely to remain in an occupation if he can expect greater overall compensation. Those individuals who have attained higher ranks can expect to receive more pay over the course of the remainder of their Army careers, if they chose to reenlist, than their lower ranking peers for two reasons. The first is simply because they are higher ranking, which results in a higher monthly salary. The second is because they can expect that in the future they will be promoted sooner to even higher ranks than their lower ranking peers in which case they would receive even higher salaries than their peers. Quantifying any psychic or non-pecuniary benefits such as prestige, perks, power, or sense of accomplishment associated with promotion would further increase this difference in expected compensation.

One must be careful when comparing retention rates of groups with different lengths of time in service because of potential selection bias. The length of an individual's service contracts is not exogenous. An individual who signs two four-year contracts may have a greater commitment to military service or less reservation about future service, due to various reasons such as family situation or personal preferences, than an individual who only signs two two-year contracts. However, the retention rates for the different years of service groups indicate no discernible pattern, with the possible exception of the below average rates of lowest and highest YOS groups.

In this study, I examine the retention decisions of second term eligible reenlistees in grade E-5 and E-6 with seven to ten YOS. This group constitutes the vast majority of the second term eligible reenlistees and is representative of the typical soldier. The demographics of this group is not unlike those of the whole sample as reported in Table 3 through Table 5. Table 6 reports the average proportion of YOS groups seven through ten that is in each grade for the entire Army, giving a representation of the relative status of the reenlistees in terms of their career progression. For example, for those with eight YOS, E-5s comprise approximately the lower two thirds of the soldiers while E-6s represent the upper third. This ratios becomes more even with time, as for those with ten YOS, of whom roughly the bottom half are E-5s and the top half are E-6s. If faster promotion is to be considered an indication of relative ability, these differences must be kept in mind when comparing the results for the different groups.

Table 7 reports the average ability levels in these groups as measured by relative AFQT quantile and standard deviations of the time to promotion to E-5 from the mean. These averages generally conform to what one would expect given the distribution of grades from Table 6. The group averages generally decrease with YOS. That the mean promotion rate for E-5s with seven YOS is so low can only be explained by the endogeneity of the length the service agreements.

4.3.3 Calculating Expected Career Pay

Using the promotion rates calculated for Tables 1 and 2, we can construct expected career paths for each individual given his grade, time served in that grade, total length of service, and CMF, if he chooses to reenlist. Additionally, by using different quantiles of promotion rates within a year and CMF, we can construct different expected career paths for different levels of performance – a higher performing individual is promoted at the faster promotion rates, while the lower performing individual is promoted at a slower rate. Using a few assumptions, we can then calculate the total pay an individual would receive over the course of these different career paths.

The simplicity of the Army's compensation system makes it conducive to making such calculations. With the exception of certain special skill bonuses, which are few and relatively small, an individual's monthly salary is dependent solely upon his rank and YOS. Because of this structure, the terms rank and pay grade are generally equivalent. In addition to a monthly base salary, individuals in the military receive a monthly sustenance allowance, and those who do not live in government-provided housing receive an additional housing allowance, which increases with rank. For calculations of pay in this study I include both of these allowances with the monthly base pay.

In calculating expected career pay I make the following assumptions: The first is that individuals who decide to reenlist after their second term will decide to remain in the Army until they are eligible for retirement after twenty YOS. This is justified by the extremely high continuation rates after ten YOS.[14] Additionally, I assume that individuals will plan to retire after twenty YOS because by then most enlisted personnel have reach their terminal grade and thereafter face fewer promotions and in-grade longevity pay raises.[15] The next assumption is that individuals will not change CMF during the remainder of their career. MOS changes in the Army are rare, and when an individual is allowed to change his MOS, it usually occurs prior to his second reenlistment. Lastly, I assume that individuals discount future income at

[14]Kirby and Thie (1996), pp. 11 and 112-114.
[15]Asch and Warner (1994b), p. 22.

a rate of 0.06 annually. I later check the sensitivity of this assumption by using an alternate discount rate of 0.18 suggested by Warner and Pleeter (2001) based upon their analysis of decisions made by military personnel during this same period of downsizing.

For each individual I calculate the expected career pay for each decile level of performance as follows: Θ^d_{gjt} is the expected number of months of time in grade (TIG) served before promotion from grade g to grade $g+1$ for the d^{th} decile performer in CMF j, at time t, based upon an average of the promotion rates of the current and previous years. Therefore,

$$\alpha^{nd}_{igjt} = \left(\sum_{l=0}^{n} \Theta^d_{g+l\ jt}\right) - TIG_{igjt} \tag{4.2}$$

is the number of months of time in service (TIS) for the d^{th} decile performer in CMF j, at time t, after which individual i can expect to be promoted to grade $g+n+1$. In other words, an individual expects that if he consistently performs at the d^{th} decile of relative performance, he will be promoted to the grade $g+n+1$ after having served a total of α^{nd} months. G^d_{igjt} is the expected grade at retirement for the d^{th} decile performer at grade g, in CMF j, given individual i's current TIG based upon the expected promotion rates. From these expectations, we can calculate an individual's total expected career pay for each level of performance by summing up the discounted monthly salaries over the projected career paths for the different rates of promotion. Λ^d_{igjt} is the expected total discounted career income for the d^{th} decile performer at grade g, in CMF j, given individual i's current TIG.

$$\begin{aligned}
\Lambda^d_{igjt} \quad = \quad & \sum_{s=0}^{\alpha^{0\ d}_{igjt}} \left(\frac{PAY_{gt\ s+TIS_{igjt}}}{(1+\rho)^{\frac{s}{12}}}\right) + \\
& \sum_{l=0}^{G^d_{igjt}-g-1} \left(\sum_{s=\alpha^{l+1\ d}_{igjt}+1}^{\alpha^{l+1\ d}_{igjt}} \frac{PAY_{g+l+1\ t\ s+TIS_{igjt}}}{(1+\rho)^{\frac{s}{12}}} \right) + \\
& \sum_{s=\alpha^{G^d_{igjt}-g-1\ d}_{igjt}+1}^{240} \left(\frac{PAY_{G^d_{igjt}\ t\ s+TIS_{igjt}}}{(1+\rho)^{\frac{s}{12}}}\right)
\end{aligned} \tag{4.3}$$

where PAY_{gts} is the monthly service pay for an individual at grade g with a cumulative TIS of s, under the pay chart effective at time t and where ρ is the discount rate. This expected career income is adjusted to 1982-84 dollars.

In order to make these calculations comparable across individuals, I use slight modifications to the calculations of career pay described above. These modifications involve the starting point in time for calculating each individual's expected career pay. For the expectations of career pay to be comparable across individuals, the length in time of each career path must be the same. Therefore, these calculations must contain the same number of months from the time of decision to the point of retirement. Additionally, I use as the start date for calculating career pay the point at which an individual's time in grade is equal to the average time in grade of the group. This allows for the comparison of expected career pay for those who would enter the same "tournament." These adjustments, although simple, require the reasonable assumption that an individual's reenlistment decision would not change if, all other things held constant, the individual's decision date and retirement date differed from their true dates by a few months.

The pay spread in an organization's compensation scheme can be measured as the difference in pay between two different levels of performance. For this study, I measure the pay spread as the difference in expected career pay between the 8th and median deciles of relative performance, or $\Lambda^{8}_{igjt} - \Lambda^{5}_{igjt}$, measured in 10,000 1982-84 dollars. Figures 2a and 2b show the distribution of expected career earnings for E-5s and E-6s, respectively, with seven to ten YOS. The values for these figures and the average pay spread between the 8th and median deciles of relative performance are reported in Table 8. Two important observations can be made from these figures. The first is that the pay spreads are fairly small, especially considering that the amount is spread out over a period of eleven to fourteen years. The largest average pay spread for these groups is under \$18,000. This indicates that the pay spreads in the Army's compensation scheme are extremely small relative to the general labor market in which there are significant differences in income levels. The second observation is that the pay spreads decrease with YOS and increase with grade. That the size of the spreads decreases with YOS is not surprising; it is simply a function of shorter lengths of time before retirement and less accumulation of the pay differences at the end of two different careers paths. However, that the pay spreads increase with grade indicates that the difference in pay between different levels of relative performance increases more than linearly. This means that the benefit to those who advance early in their careers is not just the difference in pay at the present time but also the potential for increasingly larger differences in pay at higher ranks.

4.4 Results

Using the calculations discussed above I estimate the parameters of equation (1) using ordinary least squares (OLS) and instrumental variables (two-stage least squares) techniques. I divide the sample into groups by grade and YOS. Separating the sample by YOS reduces the differences between the individual's true decision and retirement dates and those necessary to make the pay calculations comparable and preserves the reasonableness of the assumption that these deviations would not affect an individual's reenlistment decision. I separate the groups by grade because those in different grades are deciding between different tournaments. The tournaments are distinctly different because the individuals would be competing for promotion against different peer groups for the remainder of their careers, and the prize structures are different, as discussed earlier.

For the first stage of the instrumental variables (IV) estimates, I use the percent changes in the authorized strength levels for each grade, CMF, and year as instruments for the median expected career pay and the pay spreads. Because the these expectations are calculated using an average of the current and previous years' promotion rates, I use as instruments the previous three years changes in authorization levels. For E-5s I use the changes in authorization levels for E-6s through E-9s, and for E-6s I only use the changes for E-7s through E-9s. Including interactions of these instruments with the ability measure, for the potentially endogenous variables of median pay, pay spread, and pay spread interacted with ability, there are 24 instruments (three years times four grades plus another twelve including the interacted instruments) for the E-5s and 18 instruments for the E-6s.

Tables 9 and 10 report the OLS and IV estimates of the coefficients for equation (1) for E-5s and E-6s, respectively, by YOS using the alternate measures of ability. For proper inference, the reported standard errors are heteroskedasticty robust and are corrected for clustering at the CMF by year level, because, while one might expected correlation between the reenlistment decisions of individuals in the same CMF and year cohort, one cannot control for these fixed effects due to the promotion rates and career pay which vary at these levels. A positive coefficient for the interaction of the pay spread and ability indicates that a relatively higher ability individual is more likely to be retained than a lower ability individual when there is an increase in the expected pay spread.

The OLS results for the interaction term are generally very small, near zero, and all but one are statistically insignificant. However, nearly all of the IV estimates are positive and larger than the OLS results, though due to the large standard errors induced by instrumenting, only a few of the IV estimates are significantly different that their OLS counterparts. The IV estimates for the E-6s suffer the most from higher standard errors, and none of them achieve statistical significance, though all but one are positive. The tables report the F-value for a test of the joint significance of the interacted instruments in the first stage estimation of the interaction between the pay spread and ability. The extremely high F-values imply that the differences between the OLS and IV estimates are not the result of a weak instruments problem. Unfortunately, because the standard errors are corrected for clustering, a Hausman specification or over-identification test cannot be conducted to test for bias in the OLS estimates. However, because instrumenting increases nearly all of the point estimates, the OLS estimates are most likely biased downwards due to an errors in variables problem.

The IV estimates for the effects of the interaction between pay spread and ability on retention for E-5s with eight and nine YOS using the relative promotion rate to E-5 as the ability measure are 0.032 (with a standard error of 0.013) and 0.035

(0.013), respectively. This indicates that an increase of $10,000 in the pay spread will, on average, make an individual slightly more than 3% more likely to reenlist than another individual who is one standard deviation lower in ability. The same coefficient for E-5s with eight YOS using relative AFQT quantile as the ability measure is 0.080 (0.041), indicating that a similar increase in the pay spread will make an individual about 8% more likely to reenlist than another individual who is a whole quantile lower in ability. Since one standard deviation is about a third of a quantile, this 8% percent reduces to about 2.7% if one tries to make the coefficients comparable and in terms of standard deviations. This comparison, admittedly, does not hold as nicely for all of the YOS groups, but most of estimates using relative AFQT are larger than their counterparts, and the reason for this difference can be thought of as being the results of differences in the units of measure for ability.

One alternate specification for these estimates is to use a different personal discount rate. Warner and Pleeter (2001) uses military personnel's career decisions during the drawdown to estimate a personal discount rate of about 18%. Tables 11 and 12 reports the OLS and IV estimates using an 18% discount rate in the calculation of expected career pay and the corresponding pay spreads. The results in this case are very similar to those above. The OLS estimates for the effects of the interaction between pay spread and ability are small and insignificant, while the IV estimates are positive and larger than the OLS estimates but in many cases are statistically insignificant, especially for E-6s. Many of the estimates are about twice the size as the corresponding specification with a 6% discount rate. A twelve-year, $10,000 annuity with an interest rate of 18% is worth approximately the same as a $16,000 annuity with an interest rate of 6%. Therefore, we can speculate that the individuals discount hyperbolically and place more emphasis on promotions and pay implications in the near future than on the distant future.

Though the results are not reported for the sake of brevity, other alternative specifications, which include using the natural log of the promotion rate and the difference in expected career pay between the 8th and 2nd deciles of relative performance as the pay spread, yield similar results to those above. Attempts to use each of the ability measures as an instrument for the other to alleviate any attenuation bias in the estimates due to mismeasured ability produced excessively large standard errors that prevented making any relevant inference about the point estimates.

Though these estimates do not produce a "smoking gun", there enough evidence to suggest that increases in the pay spreads can induce ability sorting. There may be several reasons for the shortcomings in the evidence in producing more definitive results, other than the possibility that the theory is incorrect.

One reason may be that the calculations of career pay may not accurately reflect the expected remuneration in terms of individual utility. Absent from the calculations of career pay are quantifications of any psychic or non-pecuniary benefits to promotion such as prestige, perks, power, or sense of accomplishment. If the monetary characterization were proportional to the overall remuneration, then the point estimates would be equally proportional to the effect of true expected utility gains and spreads in the tournament. If the money is negligibly important to individuals and the calculations of career pay represent imperfect measures of the remuneration in terms of utility, then the reason that the IV estimates are larger than the OLS estimates may be due to measurement error bias as opposed to endogeneity problems. The OLS point estimates could possibly be biased downwards for both reasons.

Another reason for a lack of extremely compelling evidence may be that the pay spreads are too small to induce a differential reaction. While the model assumes that individuals have complete information, this may not necessarily be true. Individuals may be able to notice large fluctuations in promotions rates but possibly not smaller ones. Additionally, Lazear and Rosen (1981) argues that the prize spread must be "sufficiently large" to induce a separating equilibrium. A $10,140 eleven-year annuity (at a 6% interest rate), the average value of the pay spread for E-5s with nine YOS, equates to a spread of about $1,140 per year or about $95 per month. It is unclear whether is amount is sufficiently large enough. In fact, without a large enough spread, Rosen and Lazear predict that high payouts may attract low ability individuals "climb" into tournaments designed for high ability individuals while high ability individuals may "slum" into tournaments designed for low ability individuals. The combination of the effects may obscure the detection of any ability sorting because the pay spreads in the Army are not large enough.

Lastly, the estimations for career pay and the instruments for them are backwards-looking, in that they use an average of the previous two years promotion rates to predict future promotion rates, where, in fact, individuals might be more forward-looking and use future forecasts of promotion rates, based upon expected changes in force structure, to generate their expectations of future career paths.

4.5 Conclusion

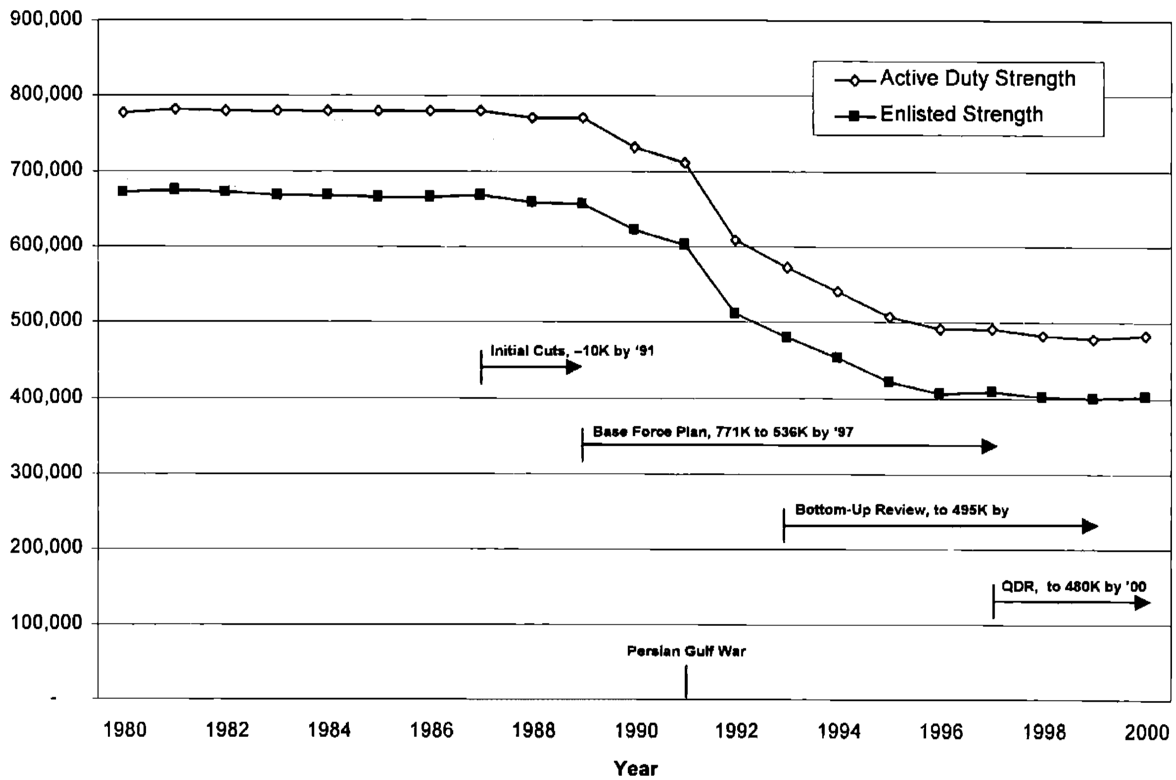
The main result gained from this study is that pay spreads in an organization's tournament-based compensation scheme can induce ability sorting. I estimate that for second term eligible reenlistees of grade E-5 in the Army, an increase in the spread of expected career pay of 10,000 1982-84 dollars will increase the probability of a individual being retained by approximately 3% points more than it increases the probability of an individual of one standard deviation lower in ability being retained. The estimates for E-6s, though positive, are not statistically significant enough for any definitive estimation of the sorting power of the pay spread.

That the pay spreads can induce ability sorting provides hierarchical firms with a powerful tool for inducing the self-selection of individuals with relatively higher abilities to stay with a firm and compete for higher-ranking positions. Such firms can manipulate their personnel and pay policies to attract higher ability individuals for placement into high-level positions, while inducing the lower ability individuals to leave the firm. However, these findings also imply that using the compensation scheme simply to adjust aggregate manpower levels may induce adverse affects of the quality mix on the resulting population of employees. In fact, as a sign that policy makers have begun to take notice of these, Congress, in 1999, unprecedentedly approved specific, targeted pay raises for different grades with different lengths of service for the 2000 military pay schedule and subsequently made further targeted pay raises for 2001 and 2002.

While this study is one of the first to examine the affects of pay spreads in tournaments on ability sorting in non-sports setting, the question still remains as to the implications of these pay spreads on worker effort and performance. The studies that have explored these effects in sporting events have found mixed results, lending credence to the notion ability sorting may be more important for an organization's productivity that attempts to improve worker effort. These may become even more relevant as the literature on worker externalities and peer effects grows and we learn more about how the quality mix of an organization affects individual and overall productivity.

Figure 1.

Drawdown Effects on Army Endstrength



Source: Department of Defense (2000).

Figure 2a. Distribution of Expected Career Earnings -- Grade E-5

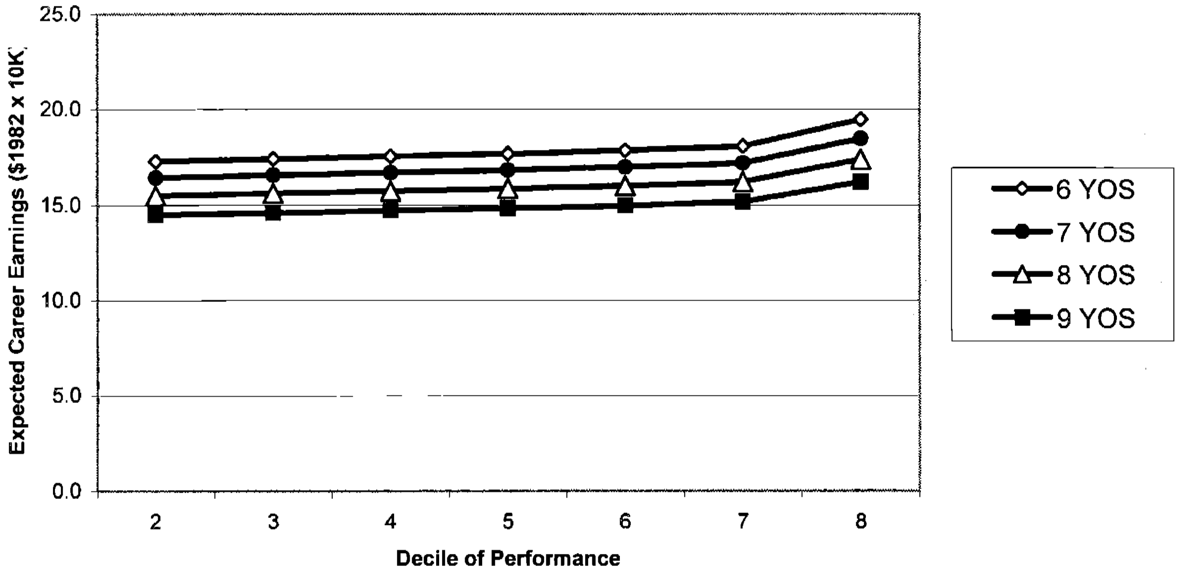
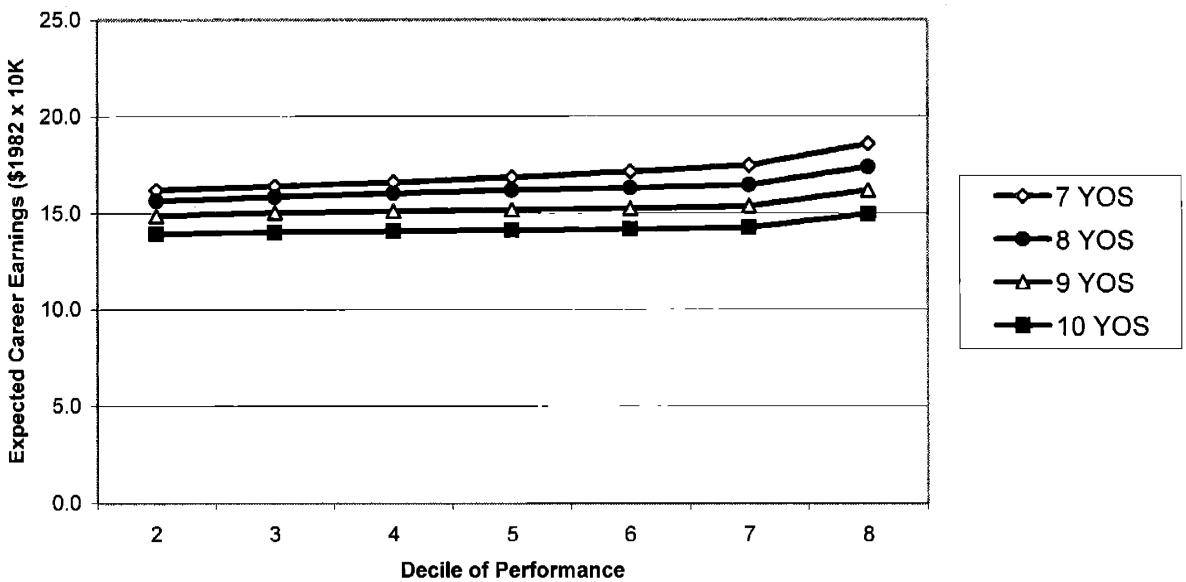


Figure 2b. Distribution of Expected Career Earnings -- Grade E-6



Notes: These figures graph the average expected career pay for different deciles of relative performance for reenlistment eligible E-5s and E-6s in the U.S. Army from 1993-1998 based upon averages of the current and the previous two years' promotion rates for each career management field.

Table 1
Median Number of Months Served in Grade Before Promotion
for Selected Career Management Fields

	Fiscal Year							
	1991	1992	1993	1994	1995	1996	1997	1998
Career Management Field	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				A. Grade E-5				
Infantry	29.8	36.7	37.6	31.4	32.4	36.7	36.1	31.2
Field Artillery	39.8	42.6	44.7	44.6	36.5	41.6	49.7	46.7
Mechanical Maintenance	47.5	43.6	48.7	50.8	55.9	53.8	54.7	54.8
Aviation Operations	28.3	24.3	37.6	31.4	34.5	33.4	48.7	40.5
Military Police	31.8	38.9	43.6	49.7	50.8	49.8	46.7	38.6
				B. Grade E-6				
Infantry	62.8	72.4	84.1	62.1	73.0	82.2	62.9	66.9
Field Artillery	70.1	71.8	93.6	75.6	76.0	81.2	68.0	68.9
Mechanical Maintenance	68.0	63.8	81.0	69.8	59.8	65.9	54.7	56.9
Aviation Operations	49.5	62.6	69.9	51.7	59.8	64.0	54.8	80.2
Military Police	95.3	66.9	84.7	86.1	69.0	85.2	58.9	77.1
				C. Grade E-7				
Infantry	83.7	85.3	83.0	75.7	76.4	73.0	75.1	59.8
Field Artillery	84.2	81.7	81.8	71.6	86.8	76.1	74.1	59.9
Mechanical Maintenance	78.1	79.2	81.7	63.7	61.9	74.1	72.1	65.0
Aviation Operations	72.1	77.3	82.6	85.9	75.1	68.0	71.0	72.0
Military Police	78.0	83.3	75.8	61.7	69.0	58.9	60.9	61.8
				D. Grade E-8				
Infantry	62.1	64.0	60.1	61.9	62.8	59.9	60.9	58.8
Field Artillery	57.0	64.6	79.3	54.7	72.0	60.9	56.9	53.7
Mechanical Maintenance	56.1	64.3	57.7	54.7	59.8	61.9	56.8	50.7
Aviation Operations	55.2	64.2	54.7	48.7	52.8	56.8	56.8	55.8
Military Police	60.1	63.9	60.0	55.8	39.6	58.9	45.6	52.8

Notes: This table reports the median number of months served in each grades by those individuals who were promoted to the next higher grade for the selected career management fields in the U.S. Army for fiscal years 1991 through 1998.

Table 2
Number of Months Time-in-Grade Before Promotion

	Fiscal Year							
	1991	1992	1993	1994	1995	1996	1997	1998
Grade	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
E-5	51.2	53.3	55.4	51.9	56.6	55.9	56.2	55.4
	(29.9)	(30.1)	(29.6)	(30.4)	(29.6)	(29.3)	(29.5)	(29.5)
E-6	75.3	72.8	81.7	76.1	73.6	80.7	69.1	71.4
	(21.9)	(21.4)	(25.9)	(24.6)	(24.4)	(26.8)	(23.8)	(26.2)
E-7	83.8	84.0	83.9	76.6	77.0	75.3	72.7	66.0
	(14.3)	(14.6)	(16.7)	(19.1)	(18.0)	(16.8)	(16.1)	(14.7)
E-8	64.0	63.6	64.3	59.5	66.0	65.0	59.3	54.5
	(14.4)	(9.3)	(15.4)	(13.7)	(16.4)	(15.4)	(13.1)	(12.0)

Notes: This tables reports the average number of months served in each grade by those individuals who were promoted to the next higher grade in the U.S. Army for fiscal years 1991 through 1998. Standard deviations are reported in parenthesis.

Table 3 (cont.)
Descriptive Statistics of 2nd Term Eligible Reenlistees
In the U.S. Army from 1993 to 1998

	Fiscal Year						Total Sample
	1993	1994	1995	1996	1997	1998	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rank							
E-1 – E-3	3.0	3.9	4.9	6.1	5.4	4.0	4.6
E-4	23.7	26.3	25.2	28.9	30.2	32.9	27.7
E-5	56.9	54.8	54.7	50.8	51.8	51.1	53.4
E-6	15.8	14.4	14.6	13.7	12.1	11.6	13.8
E-7	0.6	0.7	0.6	0.5	0.5	0.5	0.6
Months of Service	88.1 (19.9)	87.2 (20.2)	87.6 (21.4)	86.3 (22.5)	86.1 (20.5)	84.6 (19.1)	86.7 (20.7)
AFQT Quintile							
81-100%	15.5	15.1	16.3	16.6	15.9	16.5	16.0
61-80%	25.0	24.7	26.1	26.9	26.3	27.1	26.0
41-60%	35.2	35.5	34.2	34.8	36.6	38.1	35.7
21-40%	23.8	24.4	21.7	20.2	20.4	18.1	21.6
1-20%	0.5	0.3	1.7	1.5	0.8	0.3	0.9

Table 3
Descriptive Statistics of 2nd Term Eligible Reenlistees
in the U.S. Army from 1993 to 1998

	Fiscal Year						Total Sample
	1993	1994	1995	1996	1997	1998	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
N	27,904	26,004	25,121	27,159	26,147	21,667	154,002
Age	28.4 (3.6)	28.2 (3.6)	28.2 (3.6)	28.3 (3.6)	28.1 (3.5)	28.1 (3.4)	28.2 (3.6)
Race							
White	56.4	54.2	56.8	57.6	57.2	57.8	56.6
Black	36.1	37.7	34.7	33.2	32.8	32.3	34.4
Asian	1.0	1.0	1.0	1.1	1.3	1.4	1.1
Native American	0.4	0.5	0.5	0.5	0.6	0.6	0.5
Other [a]	6.1	6.5	6.9	7.5	8.0	8.7	7.2
Unknown	0.1	0.1	0.1	0.1	0.2	0.2	0.1
Gender							
Male	88.3	87.9	87.9	87.9	87.6	87.2	87.8
Female	11.7	12.1	12.1	12.1	12.4	12.8	12.2
Education Level							
B.S. Degree +	2.7	2.3	2.6	2.9	2.2	2.4	2.5
Some College	17.5	16.9	18.2	18.3	17.0	16.5	17.4
HS Diploma	73.5	74.8	74.1	74.2	76.8	78.0	75.1
GED	5.9	5.6	4.8	4.3	3.7	2.9	4.6
No HS Diploma	0.4	0.4	0.4	0.4	0.3	0.3	0.4

[a] Those in the racial classification "Other" are predominantly Hispanic.

Table 4
Retention Rates for 2nd Term Eligible Reenlistees
in the U.S. Army from 1993 to 1998

	Fiscal Year						Total Sample
	1993	1994	1995	1996	1997	1998	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Overall	0.75	0.75	0.75	0.70	0.73	0.75	0.74
Race							
White	0.73	0.73	0.72	0.66	0.69	0.72	0.71
Black	0.77	0.77	0.78	0.75	0.78	0.80	0.77
Asian	0.82	0.81	0.85	0.79	0.83	0.81	0.82
Native American	0.65	0.70	0.74	0.63	0.67	0.75	0.69
Other	0.79	0.79	0.78	0.78	0.79	0.78	0.78
Unknown	0.83	0.76	0.60	0.73	0.74	0.88	0.76
Gender							
Male	0.74	0.74	0.74	0.69	0.72	0.75	0.73
Female	0.80	0.78	0.80	0.76	0.77	0.78	0.78
Education Level							
B.S. Degree +	0.82	0.83	0.79	0.71	0.70	0.73	0.77
Some College	0.85	0.84	0.83	0.81	0.80	0.78	0.82
HS Diploma	0.73	0.73	0.73	0.68	0.72	0.75	0.72
GED	0.67	0.67	0.67	0.57	0.64	0.71	0.65
No HS Diploma	0.61	0.59	0.63	0.63	0.59	0.60	0.61
AFQT Quintile							
81-100%	0.77	0.77	0.75	0.68	0.70	0.71	0.73
61-80%	0.76	0.75	0.74	0.69	0.71	0.74	0.73
41-60%	0.75	0.75	0.75	0.70	0.73	0.76	0.74
21-40%	0.75	0.74	0.76	0.72	0.76	0.79	0.75
1-21%	0.78	0.64	0.85	0.81	0.71	0.67	0.79

Table 5
Distribution of 2nd Term Eligible Reenlistees by Years of Service and Grade with Respective Retention Rates

Years of Service	Pay Grade					
	E-1 – E-3 (1)	E-4 (2)	E-5 (3)	E-6 (4)	E-7 (5)	Total (6)
3	0.02	0.71	0.82	0.91	-	0.53
4	0.03	0.78	0.89	0.93	-	0.73
5	0.04	0.79	0.90	0.93	-	0.80
6	0.03	0.55	0.89	0.94	-	0.75
7	0.02	0.26	0.84	0.88	0.87	0.66
8	0.01	0.20	0.83	0.89	0.85	0.79
9	0.03	0.37	0.85	0.91	0.91	0.85
10	0.03	0.13	0.83	0.90	0.86	0.84
11	0.00	0.16	0.69	0.83	0.87	0.73
12	0.00	0.13	0.63	0.83	0.73	0.68
Total	0.03	0.54	0.86	0.90	0.86	0.74

Table 6
Average Distribution of Grades
by Years of Service from 1993 to 1998

Years of Service	Pay Grade				
	E-1 – E-3	E-4	E-5	E-6	E-7
	(1)	(2)	(3)	(4)	(5)
7	0.01	0.20	0.62	0.18	0.00
8	0.00	0.04	0.64	0.31	0.01
9	0.00	0.01	0.55	0.41	0.02
10	0.00	0.00	0.45	0.49	0.06

Notes: This table reports the average yearly proportion of Army enlisted personnel with the specified number of years of service that are in each grade level.

Table 7
Average Ability of 2nd Term Eligible Reenlistees
by Years of Service and Grade

	Years of Service			
	7	8	9	10
	(1)	(2)	(3)	(4)
	A. Grade E-5			
Relative AFQT Quantile	0.510 (0.276)	0.512 (0.277)	0.525 (0.278)	0.504 (0.280)
N	25,692	8,968	7,044	2,793
Promotion Rate to E-5	-0.915 (1.447)	-0.292 (1.188)	-0.227 (0.109)	-0.322 (1.124)
N	24,625	8,618	6,779	2,672
	B. *Grade E-6*			
Relative AFQT Quantile	0.578 (0.284)	0.571 (0.286)	0.572 (0.285)	0.550 (0.286)
N	3,700	3,876	4,979	2,348
Promotion Rate to E-5	0.371 (0.852)	0.378 (0.798)	0.333 (0.817)	0.217 (0.839)
N	3,416	3,607	4,594	2,175

Notes: This table reports the mean and standard deviation of the ability measures for second term eligible reenlistees for fiscal years 1993 through 1998. The AFQT measure represents the quantile of AFQT score relative to an soldier's grade and career management field in the year the soldier is eligible for reenlistement (1 is highest; 0 is lowest). The promotion rate represents the soldier's rate of promotion to grade E-5 in terms of the standard deviation of the number of days served in grade E-4 before promotion to E-5 for a soldier's career management field in the year the soldier is eligible for reenlistment.

Table 8
Distribution of Estimated Career Pay for Different Levels of Relative Performance
(Measured in $1982 x 10,000)

Years of Service	Decile of Performance: 2nd (1)	3rd (2)	4th (3)	5th (4)	6th (5)	7th (6)	8th (7)	50-80 Pay Spread (8)
				A. Grade E-5				
7	16.19	16.40	16.61	16.85	17.15	17.48	18.56	1.730
	(0.38)	(0.039)	(0.40)	(0.42)	(0.43)	(0.44)	(0.72)	(0.590)
8	15.64	15.85	16.04	16.19	16.31	16.46	17.39	1.204
	(0.38)	(0.37)	(0.37)	(0.38)	(0.39)	(0.41)	(0.66)	(0.53)
9	14.86	15.03	15.12	15.18	15.26	15.38	16.20	1.014
	(0.37)	(0.35)	(0.35)	(0.35)	(0.37)	(0.39)	(0.60)	(0.461)
10	13.92	14.02	14.08	14.13	14.19	14.27	14.97	0.845
	(0.34)	(0.34)	(0.33)	(0.33)	(0.34)	(0.35)	(0.54)	(0.404)
				B. Grade E-6				
7	17.29	17.43	17.55	17.69	17.85	18.09	19.46	1.774
	(0.40)	(0.40)	(0.40)	(0.41)	(0.43)	(0.47)	(0.76)	(0.728)
8	16.44	16.58	16.70	16.84	16.99	17.21	18.47	1.635
	(0.41)	(0.42)	(0.42)	(0.42)	(0.43)	(0.46)	(0.72)	(0.656)
9	15.50	15.63	15.74	15.88	16.03	16.23	17.40	1.522
	(0.39)	(0.42)	(0.42)	(0.42)	(0.42)	(0.44)	(0.68)	(0.600)
10	14.50	14.61	14.72	14.85	14.99	15.18	16.23	1.380
	(0.39)	(0.42)	(0.43)	(0.43)	(0.43)	(0.44)	(0.63)	(0.557)

Table 9
Estimates of the Effects of Pay Spreads in Expected Career Earnings
On the Retention of 2nd-Term Eligible Reenlistees of Grade E-5

	7 Years of Service		8 Years of Service		9 Years of Service		10 Years of Service	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			A. Relative AFQT Quantile as Ability Measure					
Median Career Pay	0.102	-0.050	0.112	0.074	0.072	0.305	0.051	0.010
	(0.009)	(0.046)	(0.012)	(0.093)	(0.012)	(0.115)	(0.013)	(0.101)
Ability	-0.030	-0.007	-0.103	-0.130	-0.027	-0.048	-0.032	-0.113
	(0.045)	(0.062)	(0.053)	(0.068)	(0.077)	(0.104)	(0.120)	(0.162)
Pay Spread	0.026	0.029	-0.024	-0.026	0.015	0.002	-0.030	-0.063
	(0.011)	(0.024)	(0.014)	(0.040)	(0.025)	(0.061)	(0.035)	(0.093)
Pay Spread * Ability	-0.003	-0.016	0.059	0.080	0.024	0.048	-0.053	0.042
	(0.014)	(0.027)	(0.021)	(0.041)	(0.035)	(0.071)	(0.063)	(0.125)
F-Test for IVs in First Stage	-	154.93	-	73.32	-	65.38	-	19.34
N	25,237		8,869		6,995		2,772	
			B. Promotion Rate to E-5 as Ability Measure					
Median Career Pay	0.109	0.013	0.104	0.039	0.057	0.153	0.045	0.144
	(0.012)	(0.039)	(0.016)	(0.077)	(0.015)	(0.104)	(0.014)	(0.075)
Ability	-0.007	-0.038	-0.003	-0.058	-0.016	-0.029	-0.010	-0.030
	(0.006)	(0.013)	(0.010)	(0.022)	(0.010)	(0.028)	(0.016)	(0.037)
Pay Spread	0.023	0.025	0.007	0.026	0.027	0.051	-0.061	-0.026
	(0.008)	(0.016)	(0.010)	(0.029)	(0.012)	(0.041)	(0.025)	(0.058)
Pay Spread * Ability	-0.001	0.007	-0.002	0.032	0.003	0.035	-0.008	0.041
	(0.003)	(0.006)	(0.006)	(0.013)	(0.008)	(0.013)	(0.015)	(0.034)
F-Test for IVs in First Stage	-	298.75	-	109.67	-	113.64	-	21.78
N	24,208		8,534		6,735		2,659	

Notes: This table reports the regression coefficients for a regression of retention on the above regressors and controls for gender, race, career management field, age, age squared, marital status, number of children, and the unemployment rate at the home of record.

Table 10
Estimates of the Effects of Pay Spreads in Expected Career Earnings
On the Retention of 2nd-Term Eligible Reenlistees of Grade E-6

	7 Years of Service		8 Years of Service		9 Years of Service		10 Years of Service	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Relative AFQT Quantile as Ability Measure								
Median Career Pay	0.038	-0.092	0.049	0.154	0.065	0.110	0.051	0.164
	(0.019)	(0.099)	(0.017)	(0.156)	(0.014)	(0.122)	(0.018)	(0.117)
Ability	0.074	-0.091	-0.053	-0.055	-0.090	-0.196	0.130	-0.038
	(0.080)	(0.151)	(0.093)	(0.128)	(0.070)	(0.101)	(0.135)	(0.201)
Pay Spread	-0.018	-0.037	0.002	0.048	-0.003	0.014	-0.002	0.024
	(0.018)	(0.081)	(0.018)	(0.066)	(0.014)	(0.043)	(0.031)	(0.088)
Pay Spread * Ability	0.031	0.137	0.014	0.013	0.007	0.080	-0.015	0.101
	(0.021)	(0.090)	(0.026)	(0.077)	(0.024)	(0.063)	(0.037)	(0.118)
F-Test for IVs in First Stage	-	9.72	-	15.51	-	26.21	-	12.51
N	3,624		3,838		4,943		2,725	
B. Promotion Rate to E-5 as Ability Measure								
Median Career Pay	0.039	-0.117	0.040	0.013	0.057	0.050	0.046	0.140
	(0.022)	(0.171)	(0.018)	(0.151)	(0.014)	(0.115)	(0.019)	(0.128)
Ability	-0.012	-0.055	-0.031	-0.005	-0.015	-0.025	-0.002	-0.098
	(0.015)	(0.068)	(0.016)	(0.050)	(0.011)	(0.032)	(0.021)	(0.062)
Pay Spread	-0.004	0.035	0.010	0.076	0.002	0.054	-0.010	0.021
	(0.012)	(0.047)	(0.013)	(0.039)	(0.011)	(0.032)	(0.023)	(0.070)
Pay Spread * Ability	0.003	0.020	0.011	-0.006	0.003	0.010	-0.009	0.072
	(0.007)	(0.034)	(0.009)	(0.031)	(0.007)	(0.022)	(0.015)	(0.041)
F-Test for IVs in First Stage	-	30.48	-	39.50	-	80.82	-	30.96
N	3,348		3,578		4,563		2,169	

Notes: This table reports the regression coefficients for a regression of retention on the above regressors and controls for gender, race, career management field, age, age squared, marital status, number of children, and the unemployment rate at the home of record.

Table 11
Estimates of the Effects of Pay Spreads in Expected Career Earnings
On the Retention of 2nd-Term Eligible Reenlistees of Grade E-5 – 18% Discount Rate

	7 Years of Service		8 Years of Service		9 Years of Service		10 Years of Service	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			A. Relative AFQT Quantile as Ability Measure					
Median Career Pay	0.176 (0.015)	-0.052 (0.079)	0.191 (0.020)	0.129 (0.167)	0.131 (0.022)	0.679 (0.227)	0.114 (0.033)	0.315 (0.230)
Ability	-0.031 (0.045)	-0.015 (0.063)	-0.101 (0.053)	-0.130 (0.067)	-0.029 (0.077)	-0.062 (0.104)	-0.031 (0.120)	-0.044 (0.164)
Pay Spread	0.051 (0.021)	0.064 (0.047)	-0.047 (0.029)	-0.063 (0.079)	0.043 (0.048)	0.024 (0.121)	-0.024 (0.066)	-0.041 (0.174)
Pay Spread * Ability	-0.008 (0.030)	-0.030 (0.059)	0.123 (0.045)	0.174 (0.086)	0.037 (0.069)	0.057 (0.144)	-0.101 (0.121)	-0.098 (0.252)
F-Test for IVs in First Stage	-	98.74	-	27.69	-	18.29	-	4.93
N	25,217		8,869		6,995		2,774	
			B. Promotion Rate to E-5 as Ability Measure					
Median Career Pay	0.195 (0.020)	0.060 (0.068)	0.183 (0.026)	0.050 (0.126)	0.112 (0.032)	0.212 (0.180)	0.115 (0.044)	0.317 (0.173)
Ability	-0.006 (0.006)	-0.034 (0.012)	-0.001 (0.010)	-0.066 (0.023)	-0.013 (0.010)	-0.033 (0.028)	-0.001 (0.017)	-0.032 (0.038)
Pay Spread	0.044 (0.016)	0.058 (0.032)	0.015 (0.020)	0.053 (0.054)	0.060 (0.022)	0.102 (0.077)	-0.085 (0.045)	-0.043 (0.110)
Pay Spread * Ability	-0.003 (0.007)	0.014 (0.013)	-0.006 (0.013)	0.077 (0.029)	0.006 (0.016)	0.068 (0.026)	-0.023 (0.030)	0.099 (0.066)
F-Test for IVs in First Stage	-	308.53	-	97.21	-	101.87	-	22.25
N	24,189		8,534		6,735		2,660	

Notes: This table reports the regression coefficients for a regression of retention on the above regressors and controls for gender, race, career management field, age, age squared, marital status, number of children, and the unemployment rate at the home of record.

Table 12
Estimates of the Effects of Pay Spreads in Expected Career Earnings
On the Retention of 2nd-Term Eligible Reenlistees of Grade E-6 – 18% Discount Rate

	7 Years of Service		8 Years of Service		9 Years of Service		10 Years of Service	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				A. Relative AFQT Quantile as Ability Measure				
Median Career Pay	0.094	-0.147	0.105	0.232	0.118	0.167	0.111	0.337
	(0.040)	(0.240)	(0.032)	(0.268)	(0.027)	(0.200)	(0.035)	(0.212)
Ability	0.091	-0.104	-0.049	-0.037	-0.083	-0.194	0.125	-0.039
	(0.081)	(0.150)	(0.091)	(0.119)	(0.070)	(0.094)	(0.135)	(0.198)
Pay Spread	-0.027	-0.071	0.011	0.084	0.005	0.011	0.004	0.059
	(0.034)	(0.144)	(0.033)	(0.108)	(0.024)	(0.070)	(0.054)	(0.146)
Pay Spread * Ability	0.046	0.283	0.021	0.001	0.007	0.148	-0.023	0.149
	(0.040)	(0.169)	(0.047)	(0.132)	(0.043)	(0.102)	(0.063)	(0.197)
F-Test for IVs in First Stage	-	5.52	-	11.96	-	19.33	-	9.27
N	3,622		3,837		4,946		2,338	
				B. Promotion Rate to E-5 as Ability Measure				
Median Career Pay	0.087	-0.198	0.086	-0.056	0.106	0.080	0.098	0.329
	(0.046)	(0.333)	(0.033)	(0.261)	(0.025)	(0.187)	(0.035)	(0.257)
Ability	-0.011	-0.055	-0.027	-0.013	-0.013	-0.023	-0.003	-0.086
	(0.014)	(0.062)	(0.015)	(0.046)	(0.011)	(0.031)	(0.021)	(0.062)
Pay Spread	-0.009	0.080	0.024	0.116	0.011	0.088	-0.012	0.054
	(0.021)	(0.086)	(0.021)	(0.064)	(0.019)	(0.054)	(0.039)	(0.128)
Pay Spread * Ability	0.007	0.040	0.018	-0.007	0.005	0.015	-0.012	0.117
	(0.013)	(0.061)	(0.015)	(0.053)	(0.012)	(0.037)	(0.026)	(0.067)
F-Test for IVs in First Stage	-	29.57	-	40.46	-	84.15	-	34.04
N	3,347		3,577		4,566		2,169	

Notes: This table reports the regression coefficients for a regression of retention on the above regressors and controls for gender, race, career management field, age, age squared, marital status, number of children, and the unemployment rate at the home of record.

Appendix A

Proof for Chapter 2

The partial derivative of $\frac{\sigma_{01}}{\sigma_\nu}$ with respect to σ_{01} is positive.

$$\begin{aligned}
\frac{\partial\left(\frac{\sigma_{01}}{\sigma_\nu}\right)}{\partial\sigma_{01}} &= \frac{\partial\left(\frac{\sigma_{01}}{\left(Var(u_i^0)+Var(\eta_i)+\left(\frac{\sigma_{01}}{\sigma_1^2}\right)^2 Var(q_i^1)\right)^{1/2}}\right)}{\partial\sigma_{01}} \\
&= \frac{\partial\left(\sigma_{01}\left(\sigma_{u^0}^2+\sigma_\eta^2+\left(\frac{\sigma_{01}}{\sigma_1^2}\right)^2\sigma_1^2\right)^{-1/2}\right)}{\partial\sigma_{01}} \\
&= \sigma_{01}\frac{\partial\left(\sigma_{u^0}^2+\sigma_\eta^2+\left(\frac{\sigma_{01}}{\sigma_1}\right)^2\right)^{-1/2}}{\partial\sigma_{01}} + \left(\sigma_{u^0}^2+\sigma_\eta^2+\left(\frac{\sigma_{01}}{\sigma_1}\right)^2\right)^{-1/2}\frac{\partial\sigma_{01}}{\partial\sigma_{01}} \\
&= -\frac{\sigma_{01}}{2}\left(\sigma_{u^0}^2+\sigma_\eta^2+\left(\frac{\sigma_{01}}{\sigma_1}\right)^2\sigma_1^2\right)^{-3/2}\left(\frac{2\sigma_{01}}{\sigma_1^2}\right) + \left(\sigma_{u^0}^2+\sigma_\eta^2+\left(\frac{\sigma_{01}}{\sigma_1}\right)^2\right)^{-1/2} \\
&= \frac{1}{\sigma_\nu} - \left(\frac{\sigma_{01}}{\sigma_1}\right)^2\frac{1}{\sigma_\nu^3} \\
&= \left(\sigma_{u^0}^2+\sigma_\eta^2+\left(\frac{\sigma_{01}}{\sigma_1}\right)^2\right)\frac{1}{\sigma_\nu^3} - \left(\frac{\sigma_{01}}{\sigma_1}\right)^2\frac{1}{\sigma_\nu^3} \\
&= \frac{\sigma_{u^0}^2+\sigma_\eta^2}{\sigma_\nu^3} \\
&> 0 \qquad \text{(Q.E.D.)}
\end{aligned}$$

Bibliography

[1] Asch, Beth J. (1993), *Designing Military Pay: Contributions and Implications of the Economic Literature*, Santa Monica, California: RAND.

[2] Asch, Beth J., and John T. Warner (1994), *A Theory of Military Compensation and Personnel Policy*, Santa Monica, California: RAND.

[3] Asch, Beth J., and John T. Warner (1994), *A Policy Analysis of Alternative Military Retirement Systems*, Santa Monica, California: RAND.

[4] Becker, Brian E., and Mark A. Huselid (1992), "The Incentive Effects of Tournament Compensation Systems," *Administrative Science Quarterly*, 37, 336-350.

[5] Buddin, Richard J., Daniel S. Levy, Janet M. Hanley, and Donald Mark Waldman (1992), *Promotion Tempo and Enlisted Retention*, Santa Monica, California: RAND.

[6] Carmichael, Lorne (1983), "Firm-Specific Human Capital and Promotion Ladders," *The Bell Journal of Economics*, 14, 251-258.

[7] Daula, Thomas V., and Robert A. Moffitt (1991), "Estimating a Dynamic Programming Model of Army Reenlistment Behavior," in Curtis L. Gilroy, David K. Horne, and D. Alton Smith (eds.), *Military Compensation and Personnel Retention: Models and Evidence*, Alexandria, Virginia: United States Army Research Institute for the Behavioral and Social Sciences, 181-201.

[8] Defense Conversion Commission (1993), *The DoD Drawdown: Planned Spending and Employment Cuts – Annex A to Adjusting to the Drawdown*, Washington, D.C.: Logistics Management Institute.

[9] Department of Defense (2000), *Selected Manpower Statistics: Fiscal Year 2000*, Washington, D.C.: Directorate for Information Operations and Reports.

[10] Ehrenberg, Ronald G., and Michael L. Bognanno (1990a), "The Incentive Effects of Tournaments Revisited: Evidence from the European PGA Tour," *Industrial and Labor Relations Review*, 43, 74-S - 88-S.

[11] Ehrenberg, Ronald G., and Michael L. Bognanno (1990b), "Do Tournaments Have Incentive Effects?" *Journal of Political Economy,* 98, 1307-1324.

[12] Fernandez, Judith C. (1992), "Soldier Quality and Job Performance in Team Tasks," *Social Science Quarterly*, 73, 253- 265.

[13] Fernandez, Richard L., Glenn A. Gotz, and Robert M. Bell (1985), *The Dynamic Retention Model*, Santa Monica, California: RAND.

[14] Gotz, Glenn A., and John J. McCall (1984), *A Dynamic Retention Model for Air Force Officers: Theory and Estimates*, Santa Monica, California: RAND.

[15] Green, Jerry R., and Nancy L. Stokey (1983), "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 91, 349-364.

[16] Hogan, Paul F., and Curtis L. Gilroy (1991), "Reenlistment Models: A Methodological Review," in Curtis L. Gilroy, David K. Horne, and D. Alton Smith (eds.), *Military Compensation and Personnel Retention: Models and Evidence*, Alexandria, Virginia: United States Army Research Institute for the Behavioral and Social Sciences, 19-42.

[17] Kirby, Shelia Nataraj, and Harry Thie (1996), *Enlisted Personnel Management: A Historical Perspective*, Santa Monica, California: RAND.

[18] Lazear, Edward P., and Sherwin Rosen (1981), "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89, 841-864.

[19] Lynch, Jim and Jeffrey Zax (1997), "The Rewards to Running: Prize Structure and Performance in Professional Road Racing," University of Colorado at Boulder.

[20] Malcomson, James M. (1984), "Work Incentives, Hierarchy, and Internal Labor Markets," *Journal of Political Economy*, 92, 486-507.

[21] McLaughlin, Kenneth J. (1988), "Aspects of Tournament Models: A Survey," in Ronald Ehrenberg (ed.), *Research in Labor Economics,* 9, 225-256.

[22] Nalebuff, Barry J., and Joseph E. Stiglitz (1983), "Prizes and Incentives: Towards a Theory of Compensation and Competition," *The Bell Journal of Economics*, 14, 21-43.

[23] O'Keeffe, Mary, W. Kip Viscusi, and Richard J. Zeckhauser (1984), "Economic Contests: Comparative Reward Schemes," *Journal of Labor Economics*, 2, 27-56

[24] Rosen, Sherwin (1986), "Prizes and Incentives in Elimination Tournaments," *American Economics Review*, 76, 701-715.

[25] Scribner, Barry L., D. Alton Smith, Robert H. Baldwin, and Robert L. Phillips (1986), "Are Smart Tankers Better Tankers? AFQT and Military Productivity," *Armed Forces and Society*, Vol. 12, 193-206.

[26] Warner, John T., and Saul Pleeter (2001), "The Personal Discount Rate: Evidence from Military Downsizing Programs," *American Economic Review*, 91, 33-53.

[27] Aigner, Dennis J., and Glen G. Cain (1977), "Statistical Theories of Discrimination in Labor Markets," *Industrial and Labor Relations Review*, 30, 175-187.

[28] Altonji, Joseph G., and Charles R. Pierret (2001), "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 116, 313-350.

[29] Anderson, N. (1970), "Functional Measurement and Psychophysical Judgement," *Psychological Review*, 77, 153-170.

[30] Arrow, Kenneth (1972), "Models of Job Discrimination," in Anthony H. Pascal, ed., *Racial Discrimination in Economic Life*, Lexington, Massachusetts: D.C. Heath and Co., 83-102.

[31] Balzer, William K., and Lorne M. Sulsky (1992), "Halo and Performance Appraisal Research: A Critical Examination," *Journal of Applied Psychology*, 77, 975-985.

[32] Cook, Thomas D., and Donald T. Campbell (1979), *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Chicago: Rand McNally College Publishing Company.

[33] Feldman, Jack M. (1981), "Beyond Attribution Theory: Cognitive Processes in Performance Appraisal," *Journal of Applied Psychology*, 66, 127-148.

[34] Goldin, Claudia, and Cecilia Rouse (2000), "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians," *The American Economic Review*, 90, 715-741.

[35] Ilgen, Daniel R., Janet L. Barnes-Farrell, and David B. McKellin (1993), "Performance Appraisal Process Research in the 1980s: What Has It Contributed to Appraisals in Use?" *Organizational Behavior and Human Decision Processes*, 54, 321-368.

[36] Murphy, Kevin R., William K. Balzer, Maura C. Lockhart, and Elaine J. Eisenman (1985), "Effects of Previous Performance on Evaluations of Present Performance," *Journal of Applied Psychology*, 70, 72-84.

[37] Phelps, Edmund S. (1972), "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62, 659-661.

[38] Seaver, W. Burleigh, and Richard J. Quarton (1976), "Regression Discontinuity Analysis of Dean's List Effects," *Journal of Educational Psychology*, 68, 459-465.

[39] Slovic, P., and S. Lichtenstein (1971), "Comparison of Bayesian and Regression Approaches to the study of Information Processing in Judgement," *Organizational Behavior and Human Performance*, 6, 649-744.

[40] Spence, Michael (1973), "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355-374.

[41] Stiglitz, Joseph E. (1975), "The Theory of 'Signaling,' Education and the Distribution of Income," *American Economic Review*, 65, 283-300.

[42] Thistlethwaite, D.L. and Donald T. Campbell (1960), "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," *Journal of Education Psychology*, 51, 309-317.

[43] Thorndike, E.L. (1920), "A Constant Error in Psychological Ratings," *Journal of Applied Psychology*, 4, 25-29.

[44] Tversky, Amos, and Daniel Kahneman (1974), "Judgement under Uncertainty: Heuristics and Biases," *Science*, 185, 1124-1131.

[45] Tyler, John H., Richard J. Murnane, and John B. Willett (2000), "Estimating the Labor Market Signaling Value of the GED," *Quarterly Journal of Economics*, 115, 431-468.

[46] Angrist, Joshua D. (2002), "How Do Sex Ratios Affect Marriage and Labor Markets? Evidence from America's Second Generation," *Quarterly Journal of Economics*, 117, 997-1038.

[47] Becker, Gary (1973), "Theory of Marriage, Part I," *Journal of Political Economy*, 81, 814-46.

[48] Black, Dan, Daniel Kermit, and Seth Sanders (2002), "The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust," *American Economic Review*, 92, 27-50.

[49] Black, Dan, Terra McKinnish, and Seth Sanders (2003), "Does the Availability of High-Wage Jobs for Low-Skilled Men Affect AFDC Expenditures? Evidence from Shocks to the Coal and Steel Industries," *Journal of Public Economics*, forthcoming.

[50] Ellwood, David, and Jonathan Crane (1990), "Family Change Among Black Americans: What Do We Know?" *Journal of Economic Perspectives*, 4, 65-84.

[51] Ellwood, David, and David Rodda (1991), "The Hazards of Work and Marriage: The Influence of Male Employment on Marriage Rates," Harvard University mimeo.

[52] Imbens, Guido W., and Joshua D. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-475.

[53] Korenman, Sanders, and David Neumark (1991), "Does Marriage Really Make Men More Productive?" *The Journal of Human Resources*, 26, 282-307.

[54] Lerman, Robert (1988), "Employment Opportunities of Young Men and Family Formation," Brandeis University mimeo.

[55] Lichter, Daniel, Felicia LeClere, and Diane McLaughlin (1991), "Local Marriage Markets and the Marital Behavior of Black and White Women," *American Journal of Sociology*, 96, 843-67.

[56] Mare, Robert, and Christopher Winship (1991), "Socioeconomic Change and the Decline of Marriage for Blacks and Whites," in The Urban Underclass, eds. Christopher Jencks and Paul Peterson, Washington: The Brookings Institute, 175-202.

[57] Sampson, Robert (1987), "Urban Black Violence: The Effect of Male Joblessness and Family Disruption," *American Journal of Sociology*, 93, 348-82.

[58] University of Texas, "The Handbook of Texas Online," http://www.tsha.utexas.edu/handbook/online/.

[59] Wilson, William J. (1987), *The Truly Disadvantaged*, Chicago: University of Chicago Press.

[60] Wilson, William J. (1996), *When Work Disappears*, New York: Alfred A. Knopf, Inc.

[61] Wood, Robert G. (1995), "Marriage Rates and Marriageable Men: A Test of the Wilson Hypothesis," *The Journal of Human Resources*, 30, 163-193.