*Econometrica*, Vol. 66, No. 2 (March, 1998), 249–288

# ESTIMATING THE LABOR MARKET IMPACT OF VOLUNTARY MILITARY SERVICE USING SOCIAL SECURITY DATA ON MILITARY APPLICANTS

BY JOSHUA D. ANGRIST[1]

The volunteer armed forces play a major role in the American youth labor market, but little is known about the effects of voluntary military service on earnings. The effects of military service are difficult to measure because veterans are both self-selected and screened by the military. This study uses two strategies to reduce selection bias in estimates of the effects of military service on the earnings of veterans. Both approaches involve the analysis of a special match of Social Security earning records to administrative data on applicants to the armed forces. The first strategy compares applicants who enlisted with applicants who did not enlist, while controlling for most of the characteristics used by the military to select soldiers from the applicant pool. This is implemented using matching methods and regression. The second strategy uses instrumental variables that were generated by an error in the scoring of the exams that screen military applicants. Estimates from both strategies are interpreted using models with heterogeneous potential outcomes. The empirical results suggest that soldiers who served in the early 1980s were paid considerably more than comparable civilians while in the military, and that military service is associated with higher employment rates for veterans after service. In spite of this employment gain, however, military service led to only a modest long-run increase in the civilian earnings of nonwhite veterans while actually reducing the civilian earnings of white veterans.

KEYWORDS: Earnings, instrumental variables, matching estimators, program evaluation, selection bias.

## 1. INTRODUCTION

THE VOLUNTEER ARMED FORCES is the largest single employer of young men and women in the United States. Each year, military recruiters screen hundreds of thousands of applicants, including large numbers of minorities who disproportionately apply to and serve in the active duty armed forces (Cooper (1977); Orvis and Gahart (1990)). However, after a period of rapid growth beginning in the early 1980's, the size of the military declined sharply, starting in 1987. Between 1989 and 1992, the number of yearly enlistments in the Armed Forces

[1] This research was supported by Grant SBR-9122627 from the National Science Foundation. The Institute for Research on Poverty at the University of Wisconsin also provided financial support. Special thanks go to Warren Buckler, Russell Hudson, and the staff of the Office of Research and Statistics at the Social Security Administration's Baltimore Headquarters for providing and assisting with Social Security earnings data. Thanks also go to Mike Dove and the staff at the Defense Manpower Data Center for providing and assisting with data on military applicants and to Michael Sarel, Dan Nabot, and Jon Guryan for research assistance. Finally, thanks go to the referees, the editor, seminar participants at Hebrew University, Brown, Columbia, CREST, and the MIT-Harvard Econometrics workshop for comments, and especially to Guido Imbens for helpful discussions. The author bears sole responsibility for the contents of this paper. The paper does not reflect the views of any US government agency or employee.

by men and women without prior military service fell by 27 percent. Enlistments by white men declined by 25 percent while enlistments by black men, the group hardest hit by military downsizing, declined by 47 percent (Angrist (1993a)). The major avenue used to effect these declines was an increase in applicant test-score cutoffs and other entry standards.

The social and economic consequences of recent reductions in the size of the armed forces have been a major cause of public concern and debate (US Congress (1989); Defense Department (1992)). Some commentators presume that the recent reductions in force size constitute a significant loss of employment opportunities, especially for blacks (e.g., Laurence (1992); Business Week (1992); Moskos and Butler (1996)). On the other hand, the labor-market value of service in the volunteer military has also been questioned (e.g., Mahoney (1991); USA Today (1990)). Attempting to resolve this conflict, many econometric studies have compared the earnings of veterans and nonveterans.[2] The proper interpretation of results from such studies is unclear, however, because veterans are both self-selected and screened by the military. Indeed, the fact that World War II veterans live longer than nonveterans the same age is a classic example of selection bias induced by pre-screening on health characteristics (Seltzer and Jablon (1974)).

The problem of selection bias plagues almost all evaluation research outside of randomized trials. In some cases, however, natural experiments are available as a partial substitute for randomization. In papers on the effects of military service in the Vietnam and World War II eras, for example, Angrist (1990) and Angrist and Krueger (1994) corrected for selection bias by using instrumental variables that affected the likelihood of military service. Imbens and van der Klaauw (1995) provide a similar analysis of the effects of compulsory military service in the Netherlands. These studies, which consistently show that being drafted can hurt future earnings, exploit the fact that the military draft induced some exogenous variation in the process of selection for military service. Until now, however, researchers have not isolated similarly convincing strategies for the identification of effects of *voluntary* military service on earnings.[3]

This paper presents evidence from two new strategies for estimating the effect of voluntary military service on the earnings and employment status of veterans.

[2] Papers comparing the earnings of volunteers with those of nonveterans include Bryant, Samaranayake, and Wilhite (1993), Magnum and Ball (1989), and Phillips, Andrisani, Daymont, and Gilroy (1992). Most of these studies report higher earnings for veterans, though Stafford (1991) finds that the economic returns to years of military experience are lower than the returns to years of civilian experience.

[3] Other studies of the effects of compulsory military service include Cutright (1974), who linked 1964 Social Security earnings to data on conscripts, and Laurence, Ramsberger, and Gribben (1989), who examined low-ability volunteers from the late 1970's (a group examined here as well) and low-ability veterans from the Vietnam-era "Project 100,000." Both of these studies found negative effects in comparisons to partially matched control groups. In contrast, Beusse's (1974) earlier findings using Social Security data suggest that low-ability veterans from "Project 100,000" did earn more than comparable civilians. Regression studies of Vietnam veterans by Berger and Hirsch (1983) and De Tray (1982) report both positive and negative effects.

First, comparisons by veteran status are restricted to a sample of applicants to the military, only about half of whom actually enlist. Nonenlisting applicants probably provide a better control group for veterans than conventional cross-section samples because, like veterans, applicants have indicated a strong interest in military service. Moreover, the data analyzed here contain information on most of the characteristics used by the military to screen applicants. The selection bias induced by military screening can therefore be eliminated using regression techniques or by matching on the covariates used in the screening process.

The second strategy relies on instrumental variables generated by changes in the score scale used to grade the military's entrance exam. When the Armed Forces Vocational Aptitude Battery (ASVAB) was introduced as an applicant screening test in 1976, the score scales unintentionally allowed large numbers of previously unqualified applicants to enter the military. This episode has come to be known as "ASVAB misnorming" because score scales are determined by a process known as test-norming (see, e.g., Maier and Sims (1986)). After policy makers became aware of the severity of the scoring error, and new score scales were developed and checked for validity, the ASVAB score scale was corrected in late 1980. This correction reduced the probability of acceptance to the military by as much as 30 percent for some low-scoring applicants. Dummy variables indicating the cohorts that applied before the correction therefore provide instruments for the relationship between veteran status and earnings.[4]

In addition to estimating the effects of military service, the paper also makes a contribution by illustrating and comparing a number of techniques for nonexperimental evaluation research in models with heterogeneous treatment effects. Section 2 begins by outlining a conceptual framework for evaluation research with heterogeneous potential outcomes. This framework, based on the work of Rubin (1974), is used to clarify the relationship between regression estimates of average treatment effects and the matching estimator discussed in Rubin (1977). Instrumental variables (IV) estimates of treatment effects in models with heterogeneous potential outcomes are also discussed. As in Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996), IV estimates are interpreted as local average treatment effects (LATE) that are tied to a specific intervention. In this case, IV is shown to estimate the effect of military service on low-scoring applicants who got into the military solely because they applied before errors in the ASVAB score scale were corrected. The discussion of IV methods in this section extends the original LATE result because it applies to an IV estimator that involves interaction terms in a two-way contrast.[5]

[4] The military ordinarily adjusts entrance standards in light of recruiting conditions, but the ASVAB misnorming was a watershed event in the history of the AVF. It has been estimated that one out of three male Army and Air Force recruits between 1976 and 1980 would not have been enlisted under conventional entrance standards (Eitelberg (1988); see also Maier and Truss (1983)).

[5] See Imbens, Leibman, and Eissa (1997) for a related discussion of differences-in-differences estimators in models with heterogeneous potential outcomes.

After describing the data in Section 3, the matching and regression estimates are presented in Section 4, and the IV estimates are presented in Section 5. The juxtaposition of results from alternate estimation strategies highlights the fact that matching, regression, and IV estimators potentially uncover different parameters even when the assumptions justifying each approach are valid. In this case, however, the results of all the estimation methods are broadly consistent, suggesting that the mechanisms which link military service to earnings are similar for different groups of veterans. The empirical findings indicate that soldiers who served in the early 1980's were paid considerably more than comparable civilians while in the military and that veterans were more likely to have positive earnings in each year after service. But in spite of this employment gain, military service led to only a modest long-run increase in the civilian earnings of nonwhite veterans while actually reducing the civilian earnings of white veterans, at least in the years following their discharge from service. Another important finding is that for both whites and nonwhites, simple comparisons by veteran status generate misleading overestimates of the effect of military service on civilian earnings.

## 2. CONCEPTUAL FRAMEWORK

### 2.1 *Matching Estimators*

The primary purpose of this paper is to estimate the impact of military service on the earnings of veterans. The parameter of interest can be formally described as follows. For any applicant observed after application, define random variables representing what the applicant would earn had he served in the military and what the applicant would earn had he not served in the military. Denote these two potential outcomes by $Y_0$ and $Y_1$ and denote veteran status by a dummy variable, $D$. For each applicant, we observe only $Y = Y_0 + (Y_1 - Y_0)D$, so $Y_0$ is not observed for veterans and $Y_1$ is not observed for nonveterans. We might nevertheless still hope to identify certain averages of $Y_1 - Y_0$. The effect of treatment on the treated (Rubin (1977)) is one such parameter:

$$E[Y_1 - Y_0 \mid D = 1] = E[Y_1 \mid D = 1] - E[Y_0 \mid D = 1].$$

This tells us whether, on average, veterans benefited or suffered from military service.

Simple comparisons by veteran status can be used to estimate $E[Y_1 - Y_0 \mid D = 1]$. Because the sample used here includes only applicants to the military, these comparisons control for differences between veterans and nonveterans that originate in the decision to apply to the military. But such comparisons do not control for most of the criteria used by the military to choose which applicants to accept. The earnings of nonenlisting applicants might therefore provide a poor indicator of what enlisting applicants would have earned if they had not served in the military. To explore this point further, note that the comparisons by veteran status can be decomposed as follows:

$$(1) \qquad E[Y_1 \mid D=1] - E[Y_0 \mid D=0] = E[Y_1 - Y_0 \mid D=1] + \{E[Y_0 \mid D=1] - E[Y_0 \mid D=0]\}.$$

This shows that comparisons of earnings by veteran status are equal to $E[Y_1 - Y_0 \mid D=1]$ plus a bias term attributable to the fact that, even in the applicant population, the earnings of nonveterans are not necessarily representative of what veterans would have earned had they not served in the military.

If veteran status were randomly assigned, then $D$ would be independent of $Y_0$ and $Y_1$, implying $E[Y_0 \mid D=0] = E[Y_0]$ and $E[Y_1 \mid D=1] = E[Y_1]$. In this case, the effect of treatment on the treated is also the average treatment effect in the population subject to randomization and can be estimated by simple comparisons. Rubin (1978) refers to treatment status that is independent of potential outcomes as an *ignorable* treatment assignment. Although claims for ignorable treatment assignment are usually implausible outside of an experimental setting, it is more plausible that veteran status among applicants is ignorable conditional on a set of observed covariates. In particular, it might be the case that in the pool of military applicants, all of whom have indicated a strong interest in military service by taking the time to undergo a preliminary physical and complete the ASVAB, the principle remaining sources of bias in veteran/nonveteran comparisons are differences in the variables used by the military to screen and select applicants. These variables are age, schooling, and test scores, and they are available in the data set analyzed here.

The assumption that veteran status is ignorable conditional on predetermined covariates, denoted by $X$, can be expressed using conditional independence notation as follows:

Assumption 1 (Rosenbaum and Rubin (1983)): $(Y_1, Y_0) \amalg D \mid X$.

This assumption implies that

$$(2) \qquad E[Y_0 \mid D, X] = E[Y_0 \mid X].$$

The effect of treatment on the treated can therefore be estimated using the sample analog of the following expression:[6]

$$(3) \qquad E[Y_1 - Y_0 \mid D=1] = E\{E[Y_1 \mid X, D=1] + E[Y_0 \mid X, D=1] \mid D=1\}$$
$$= \int \{E[Y_1 \mid X, D=1] - E[Y_0 \mid X, D=0]\} dF(X \mid D=1),$$

[6] Assumption 1 has been called "selection on observables" in a regression context (Goldberger (1972), Barnow, Cain, and Goldberger (1981)). Regression estimates are typically discussed in a context where $Y_1 - Y_0$ is assumed to be constant, in which case (2) implies the same restriction on conditional means as Assumption 1.

where $dF(X \mid D=1)$ is the density or probability mass function for $X$ given $D=1$. In other words, $E[Y_1 - Y_0 \mid D=1]$ is obtainable as a weighted average of contrasts between veteran and nonveteran earnings at each value of $X$. The weights are given by the density or distribution function of $X$ among veterans. Values of $X$ for which there are no veterans, i.e., for which the probability $P(D=1 \mid X)=0$, are given zero weight.

A problem that sometimes arises with this approach, noted by Rubin (1977), is that even if the covariates are discrete, there can be values of $X$ where only treated observations appear. At such values, the conditional expectation $E[Y_0 \mid X, D=0]$ is naturally undefined. This problem is handled here by defining the parameter of interest to be

$$\alpha_c \equiv E[Y_1 - Y_0 \mid D=1, P(D=1 \mid X) < 1],$$

where the restriction $P(D=1 \mid X) < 1$ refers to the population. In practice, it can happen that some population cells where both treatment and control observations are available nevertheless remain unrepresented in a random sample. In this study, however, the sample was drawn conditional on $X$. Therefore, sample observations on both veterans and nonveterans are necessarily available wherever the population probability of treatment is neither zero nor one (subject to certain confidentiality restrictions described in the Appendix).

To construct an estimator of $\alpha_c$ based on (3), suppose that $X$ is discrete, as it is here, taking on values $\{x_1, \ldots, x_k, \ldots, x_K\}$. Suppose also that there are $N_{1k}$ observations in the population of veterans with $X=x_k$ and $N_{0k}$ observations in the population of nonveterans with $X=x_k$. The corresponding sample sizes are denoted $n_{1k}$ and $n_{0k}$. Let $\delta_k = 1[n_{1k}>0, n_{0k}>0]$.[7] Finally, let $\bar{y}_{1k}$ denote the average earnings of veterans with $X=x_k$ and let $\bar{y}_{0k}$ denote the average earnings of nonveterans with $X=x_k$. Then the following estimator is consistent for the population parameter, $\alpha_c = \{E[Y_1 - Y_0 \mid D=1, P(D=1 \mid X) < 1\}$, when either Assumption 1 or equation (2) holds:

$$\hat{\alpha}_c \equiv \sum_k \delta_k N_{1k}[\bar{y}_{1k} - \bar{y}_{0k}] \Big/ \sum_k \delta_k N_{1k}. \tag{4}$$

Note that the weighting function in (4), $\delta_k N_{1k}/\sum_k \delta_k N_{1k}$, is the population distribution function of $X$ among veterans at values where $[\bar{y}_{1k} - \bar{y}_{0k}]$ is defined.[8]

In what follows, I refer to $\hat{\alpha}_c$ as a controlled contrast because it is an estimator that controls for all observed differences between veterans and nonveterans at the time of application. In practice, the observed covariates take on values in the set of all possible combinations of race, application year,

[7] As noted above, the sample design implies that $\delta_k = 1[n_{1k}>0, n_{0k}>0]$ equals the population indicator $1[N_{1k}>0, N_{0k}>0]$. In practice, however, a few cells are missing because of the confidentiality edit.

[8] Since the sampling scheme conditions on $X$, the estimator $\hat{\alpha}_c$ is unbiased as well as consistent. This estimator differs slightly from the estimators discussed by Rubin (1977) in that population cell sizes, $N_{1k}$, are used as weights. Rubin (1977) assumes random sampling and uses $n_{1k}$ to weight.

schooling at the time of application, the Armed Forces Qualification Test (AFQT) score group,[9] and year of birth. To put this approach in context, note that a sample-weighted version of $\hat{\alpha}_c$ was used by Card and Sullivan (1988, Table III) to estimate the effect of a government training program on employment rates, controlling for employment histories. More recently, Dehejia and Wahba (1995) estimated the effect of a training program on trainee earning using Rosenbaum and Rubin's (1983, 1984) modified version of $\hat{\alpha}_c$ incorporating the propensity score.[10] These examples are somewhat unusual because econometric evaluations typically rely on regression models to control for covariates (see, e.g., Barnow, Cain, and Goldberger (1981), Ashenfelter and Card (1985), Heckman and Robb (1985)). The next subsection discusses the relationship between $\hat{\alpha}_c$ and regression estimates of treatment effects.

## 2.2 *Regression Estimators*

Differences between regression and matching strategies for the estimation of treatment effects are partly cosmetic. While matching methods are often more transparent to nonspecialists, regression estimation is more straightforward to implement when covariates are continuously distributed because matching on continuous covariates requires stratification or pairing (Cochran (1968)). Note, however, that both methods require a similar sort of approximation since regression on continuous covariates in any finite sample requires functional form restrictions. The fact that both stratification and functional form approximations can be made increasingly accurate as the sample size grows suggests that the manner in which continuous covariates are accommodated is not the most important difference between the two methods.

The essential difference between regression and matching in evaluation research is the weighting scheme used to pool estimates at different values of the covariates. This can be seen by analyzing the matching and regression estimands in a simple example where there is a single binary covariate, $x$, and the probability of treatment is positive at both values of $x$. When restriction (2) holds, we can write

$$E[Y_1 - Y_0 \mid x = 1, D = 1] = E[Y_1 - Y_0 \mid x = 1] = \alpha_1$$

and

$$E[Y_1 - Y_0 \mid x = 0, D = 1] = E[Y_1 - Y_0 \mid x = 0] = \alpha_0.$$

[9] The AFQT is a composite constructed from some of the ASVAB subtests. AFQT test results are usually reported as one of 5 score-group categories that indicate the test-taker's position relative to a reference population. Score group I denotes percentiles 93–100 in the reference population and score group V denotes the lower decile of the reference population.

[10] The propensity score is the function $e(X) \equiv P[D = 1 \mid X]$. Methods incorporating the propensity score exploit the fact that if $D$ is ignorable given $X$, it is also ignorable given $e(X)$.

The effect of treatment on the treated is therefore,

$$(5)\quad \begin{aligned} &E[Y_1 - Y_0 \mid D=1] \\ &\quad = \alpha_0 P[x=0 \mid D=1] + \alpha_1 P[x=1 \mid D=1] \\ &\quad = \frac{\alpha_0 P[D=1 \mid x=0]P[x=0] + \alpha_1 P[D=1 \mid x=1]P[x=1]}{P[D=1]}. \end{aligned}$$

This is just equation (3) for the simple example.

Now consider the regression model

$$(6)\quad Y = \beta_0 + \beta_1 x + \alpha_r D + \mu,$$

where $\mu$ is a (population) residual that satisfies $E[\mu x] = E[\mu D] = 0$, so that

$$(7)\quad \alpha_r = E\{(D - E[D \mid x])Y\}/E\{(D - E[D \mid x])D\}.$$

By definition, $\alpha_r$ is free of "omitted variables bias" from the covariate $x$. To simplify $\alpha_r$, note that under Assumption 1, the conditional expectation function for $Y$ given $D$ and $x$ is

$$(8)\quad Y \equiv E[Y \mid D, x] + \epsilon = E[Y_0 \mid x] + E[Y_1 - Y_0 \mid x]D + \epsilon.$$

Using (8) to substitute for $Y$ in (7), we have:

$$(9)\quad \alpha_r = \frac{\begin{array}{l}\alpha_0 P[D=1 \mid x=0](1 - P[D=1 \mid x=0])P[x=0] \\ \quad + \alpha_1 P[D=1 \mid x=1](1 - P[D=1 \mid x=1])P[x=1]\end{array}}{E\{P[D=1 \mid x](1 - P[D=1 \mid x])\}}.$$

Thus, $\alpha_r$ is also a weighted average of the underlying covariate-specific treatment effects, $\alpha_0$ and $\alpha_1$. Moreover, a key feature common to both $\alpha_r$ and $\alpha_c$ is that values of $X$ where $P(D=1 \mid X)$ is equal to either 1 or 0 are given zero weight.

The difference between $\alpha_r$ and $\alpha_c$ is in the nature of the weights at values of $x$ where both veterans and nonveterans are observed. The parameter $\alpha_c$ weights each of the underlying treatment effects by $P[D=1 \mid X]P[X]$, whereas the regression parameter weights each of the underlying treatment effects by $P[D=1 \mid X](1 - P[D=1 \mid X])P[X]$. In other words, the weights underlying $\alpha_c$ are proportional to the probability of veteran status at each value of the covariates while the weights underlying $\alpha_r$ are proportional to the variance of veteran status at each value of the covariates.[11]

The question of whether regression results will differ from matching results depends on how much treatment effects vary with $X$ and on the range of values for $P[D=1 \mid X]$. If, for example, the values of $P[D=1 \mid X]$ are all less than .5, then the matching and regression weights are roughly proportional. On the

[11] When $X$ takes on more than two values and the regression includes a saturated model for $X$, the regression estimand is

$$E\{E[Y_1 - Y_0 \mid X]P[D=1 \mid X](1 - P[D=1 \mid X])\}/E\{P[D=1 \mid X](1 - P[D=1 \mid X])\}.$$

other hand, the weights are negatively correlated if the probability of treatment is always $\frac{1}{2}$ or greater. Suppose that in the previous example $P[D=1 \mid x=0]=.9$, $P[D=1 \mid x=1]=\frac{1}{2}$, and $P(x=1)=\frac{1}{2}$. Applying equations (5) and (9), we have $E[Y_1 - Y_0 \mid D=1] = .64\alpha_0 + .36\alpha_1$ and $\alpha_r = .26\alpha_0 + .74\alpha_1$. Thus, while $E[Y_1 - Y_0 \mid D=1]$ reflects the fact that veterans are much more likely to have $x=0$, the regression parameter $\alpha_r$ puts more weight on the treatment effect for those with $x=1$ because the variance of $D$ is much larger for that group.

### 2.3 *Instrumental Variables Estimators*

The IV estimates in this paper exploit changes in the probability of enlistment caused by the ASVAB misnorming. When the Armed Forces Vocational Aptitude Battery (ASVAB) was first introduced in 1976, incorrectly normed score scales allowed large numbers of applicants whose true AFQT scores should have disqualified them to enter the military. When the ASVAB score scale was corrected in October, 1980, the probability of enlistment dropped dramatically for applicants whose correctly normed AFQT scores put them in category IV but previously appeared to be in category III. This is because the military tries to minimize the number of category-IV enlistments.[12] True high-scorers, however, were largely unaffected by the scoring change because errors in the score scale were concentrated at the low end of the AFQT score distribution.

The simplest way to use the ASVAB misnorming to construct IV estimates is to use application-year dummies as instruments. Conditional on race, year of birth, schooling, and AFQT scores, any remaining effects of application year on earnings are arguably attributable to the large differences in the probability of being accepted for military service on different years caused by misnorming. A problem with this approach is that application year is the same as time-since-application. This variable is very likely related to earnings for reasons other than veteran status, since application to the military typically signals a labor market transition like school completion.

Two approaches are used to control for possible application-year effects. The first exploits the fact that, because three application-year dummies plus interaction terms are available as instruments, a linear trend for time-since-application can be included in the model. The second strategy exploits the fact that application year had almost no effect on applicants with true AFQT scores in score category III. Because of the interaction between application year and AFQT scores, additive application-year effects can also be included as regressors. As in most IV strategies, however, potential outcomes (or the stochastic part of a model for potential outcomes) must be assumed independent of the instruments, conditional on regressors. This independence assumption is the

[12] Partly in response to the misnorming episode, a ceiling of 80 percent of category IV enlistments was imposed in October, 1981. Also, military, regulations required that no applicant with an AFQT score below the 31st percentile (category IV–V) who is not a high school graduate be accepted for enlistment after October 1981 (Defense Department (1988, p. II-23)). Category V enlistments had already been outlawed in 1951.

major difference between IV estimation and regression or matching strategies. For the latter two methods to produce valid causal effects, treatment status itself must be independent of potential outcomes conditional on covariates, as described in Assumption 1.

Notation for the IV models is as follows. Let $W$ be a covariate taking on values in the set $\{w_1,\dots,w_k,\dots,w_K\}$ to indicate all possible combinations of year of birth and schooling ($W$ is $X$ without application year and AFQT scores). Denote AFQT score group category by $S$, application year by $A$, and time since application by $t-A$, where, for the purposes of this general discussion, $t$ is fixed. The analysis begins with a constant-treatment-effects model where $Y_1 - Y_0$ equals a constant, $\alpha^*$. Given this assumption, identification turns on restrictions imposed on $Y_0$. The first model for $Y_0$ is

$$(10) \qquad E[Y_0 \mid W = w_k, S, A] = \beta_{kS} + (t-A)\gamma,$$

where the term $\beta_{kS}$ denotes a separate effect for each combination of $W$ and $S$ and the term $(t-A)\gamma$ is a linear trend. The key identifying assumption here is that, except for the linear term, application year is excluded from the conditional mean of nonveteran earnings. Constant-treatment-effects is also a strong simplifying assumption, but provides a natural starting place for discussion of IV estimates since attention is initially focused on the exclusion restrictions. Given the assumption of constant treatment effects and equation (10), we have

$$(11) \qquad E[Y \mid W = w_k, S, A] = \beta_{kS} + (t-A)\gamma + \alpha^* E[D \mid W = w_k, S, A].$$

Estimates of $\alpha^*$ can therefore be constructed by replacing population means with sample means and fitting (11) by ordinary or weighted least squares.

The strong restrictions embodied in equation (11) can be partly relaxed. In particular, the term $(t-A)\gamma$ in (10) can be replaced by $\delta_{kA}$, an additive application-year effect for each value of $W$. The equation motivating IV estimation in this case is

$$(12) \qquad E[Y \mid W = w_k, S, A] = \beta_{kS} + \delta_{kA} + \alpha^* E[D \mid W = w_k, S, A].$$

The parameter $\alpha^*$ is identified in equation (12) because, although the equation includes a full set of effects for each combination of $W$ and $S$ and a full set of effects for each combination of $W$ and $A$, the conditional probability of veteran status also varies as a function of interactions between $A$ and $S$.

The interactions between $A$ and $S$ are highlighted in Table III in Section 5 below, which reports enlistment probabilities by year of application and AFQT score group. The table shows a sharp decline in enlistment probabilities for men with AFQT scores in AFQT category IV, but little decline for men with AFQT scores in category IIIb, and no change at all for men with scores in category IIIa.[13] The easiest way to see how (12) exploits this information is to suppose

[13] Category IV scores are percentile scores in the 10–30 range, category IIIb scores are percentile scores in the 31–49 range, and category IIIa scores are percentile scores in the 50–64 range. The AFQT score data tabulated in Table III refer to a corrected score scale.

that $W$ is constant. Equation (12) then becomes

$$E[Y \mid S, A] = \beta_S + \delta_A + \alpha^* E[D \mid S, A]. \tag{13}$$

Because $E[D \mid S, A]$ does not vary with $A$ for men with AFQT scores in category IIIa, differences in $E[Y \mid S, A]$ by application year for group IIIa applicants identify the term $\delta_A$. Estimation based on (12) or (13) can therefore be interpreted as using AFQT-IIIa applicants as a control group to estimate differences in average earnings by application year that are not attributable to differences in acceptance probabilities.

The IV estimates have been discussed so far in the context of a model where $Y_1 - Y_0$ is fixed. However, these estimates can also be interpreted as average causal effects for a particular subpopulation, similar to estimates of the effect of treatment on the treated produced by matching estimators. The causal interpretation of IV estimates is outlined here for a simple example where there is a single binary covariate, $s$, indicating applicants with AFQT scores in category IV ("low"), and a single binary instrument, $a$, indicating applicants who applied before 1981 ("early").

Just as potential earnings have been indexed against veteran status, it is possible to define potential veteran status, $D_1$ and $D_0$, indexed against the binary instrument, $a$. The random variable $D_1$ indicates whether a given applicant would have served in the military if he applied early, while $D_0$ indicates whether a given applicant would have served in the military if he applied late. In practice, we observe only $D = D_0 + (D_1 - D_0)a$ for any given applicant. Following Imbens and Angrist (1994), an interpretation of the IV estimates can be based on the assumptions that potential outcomes and potential treatment status are jointly independent of $a$, and that potential treatment status satisfies a monotonicity condition, i.e., that $D_1 \geq D_0$ with probability one. Monotonicity is implied by most economic and econometric models for treatment assignment. Given these two assumptions, simple IV estimators derived from comparisons of average earnings and the probability of veteran status by early/late application year would be equal to the local average treatment effect (LATE), in this case, $E[Y_1 - Y_0 \mid D_1 \geq D_0]$. This is the effect of treatment on the treated for veterans who got into the military solely by virtue of applying before admission standards were raised.

The problem with this interpretation of IV estimates based on the ASVAB misnorming is, as noted above, that potential earnings almost certainly vary by application year for reasons other than veteran status. This is the motivation for including additive application-year effects in equations (11)–(13). The joint distribution of $Y_1$, $Y_0$, $D_1$, and $D_0$ is therefore restricted as follows. First, the conditional mean function of $Y_0$ allows for an effect of $a$ but is restricted to be additive:

Assumption 2: $E[Y_0 \mid s, a] = \beta_0 + \beta_1 s + \delta a.$

Second, individual-level treatment effects $(Y_1 - Y_0)$ and potential treatment status $(D_1, D_0)$ are assumed to be jointly independent of the instrument $a$ given the covariate, $s$:

ASSUMPTION 3: $\{(Y_1 - Y_0), D_1, D_0\} \amalg a \mid s$.

This is an ignorability assumption for the instrument, similar to Assumption 1, but cast in terms of $Y_1 - Y_0$ because of the secular effect of $a$ introduced in Assumption 2. Note that independence of $Y_1 - Y_0$ and $a$ is a strong assumption that almost certainly fails to hold during the years when veterans were entering and leaving the military. On the other hand, the matching estimates reported below suggest that this may be a reasonable assumption for earnings after 1985.

The last assumption in this setup imposes a monotonicity restriction involving both potential treatment assignments and the covariate $s$, which plays a special role here because men with $s = 0$ are used to identify the effect of $a$ in the model for $Y_0$:

ASSUMPTION 4: $P[s(D_1 - D_0) \geq (1 - s)(D_1 - D_0) \geq 0] = 1$.

This assumption requires the interaction between application year and test scores to operate in a unidirectional manner. When $s = 1$, i.e., for low-scoring men, Assumption 4 implies $D_1 \geq D_0$ with probability one as in the usual case. But when $s = 0$, i.e., for high-scoring men, Assumption 4 implies that application year has no effect on veteran status, or $D_1 = D_0$. It is this restriction that allows the use of application year contrasts for high scoring men to control for application year differences that are not attributable to military service. The monotonicity captured in Assumption 4 is clearly stronger than univariate monotonicity. On the other hand, it has testable implications that are easily checked. For example, Table III shows no application-year differences in enlistment probability for men with AFQT scores in category IIIa for both races and category IIIb for nonwhites.

Now, consider the population analog of an IV estimate of the effect of veteran status on $Y$ in an equation that controls for additive $s$ and $a$ effects, while using the interaction term $(a \cdot s)$ as an instrument. This estimand solves the conditional expectation function,

$$E[Y \mid s, a] = \beta_0 + \beta_1 s + \delta a + \alpha_{IV} E[D \mid s, a], \tag{14}$$

for the parameter $\alpha_{IV}$. The solution is

$$\alpha_{IV} = \frac{\begin{array}{c}\{E[Y \mid s=1, a=1] - E[Y \mid s=1, a=0]\} \\ -\{E[Y \mid s=0, a=1] - E[Y \mid s=0, a=0]\}\end{array}}{\begin{array}{c}\{E[D \mid s=1, a=1] - E[D \mid s=1, a=0]\} \\ -\{E[D \mid s=0, a=1] - E[D \mid s=0, a=0]\}\end{array}}. \tag{15}$$

The causal interpretation of this ratio is given in the following theorem:

THEOREM: *Suppose that Assumptions* 2–4 *hold. Then*

$$\alpha_{IV} = E[Y_1 - Y_0 \mid s = 1, D_1 > D_0]. \tag{16}$$

This modifies LATE for the case where a two-way contrast is required to control for possible secular effects of the instrument. Proof of the theorem is given in an Appendix, along with a slightly more general result that allows $P[s(D_1 - D_0) \geq (1-s)(D_1 - D_0)] = 1$, instead of Assumption 4.

The meaning of (16) is that $\alpha_{IV}$ can be interpreted as an average treatment effect for the subpopulation affected by the "experiment" embodied in application year. In particular, IV estimates capture the effect of military service on these low-scoring men whose treatment status was potentially affected by application year (i.e., $D_1 > D_0$). Note that the population with $s = 1$ and $D_1 > D_0$ does not include all veterans ($D = 1$), all low scorers ($s = 1$), or even all veterans with $s = 1$, because there are many veterans with low AFQT scores who would have served regardless of when they applied. On the other hand, like the controlled contrast, $\alpha_c$, the parameter $\alpha_{IV}$ captures the effect of treatment on a well-defined subpopulation that was subjected to treatment.

Finally, note that the IV estimates reported below are actually more complicated than the sample analog of $\alpha_{IV}$ in (15) in two ways. First, the reported estimates implicitly use more than one instrument because estimation of equations (11) and (12) involve multiple application-year/score-group contrasts.[14] Second, there are exogenous covariates other than AFQT scores. A version of (16) can easily be developed to handle these modifications, however. IV estimators with many covariates implicitly combine simple estimators like $\alpha_{IV}$ in a variance-weighted average, in the same way that regression produces a variance-weighted average of covariate-specific estimates. Similarly, IV estimators that use more than one instrument pool the full set of underlying single-instrument estimators with weights proportional to the effect of each instrument on the treatment dummy. See Angrist and Imbens (1995) for details.

## 3. DATA DESCRIPTION AND COMPARISONS BY VETERAN STATUS

To estimate the effect of voluntary military service on earnings, I combined administrative data from the US military with earnings data from the Social Security Administration (SSA). The military data come from Defense Manpower Data Center (DMDC) files containing information on applicants and entrants to the military for each fiscal year.[15] The applicant records report information collected at the time of application, including basic demographic variables, physical examination results, and test scores. Applicant records do not indicate whether an applicant eventually enlists and enters the military. Instead, the act of enlistment generates a new record in the DMDC's computerized filing

[14] Estimates using a single instrument were reported in an earlier version of this paper (Angrist (1995)). An illustrative calculation for the 1988–91 average employment rates of nonwhites is

$$\alpha_{IV} = [(.193 - .151) - (.146 - .113)]/[(.448 - .283) - (.675 - .672)] = .056.$$

This is larger but less precise (s.e. = .021) than the estimates reported below.

[15] Aspects of the DMDC record-keeping system are described in Berryman, Bell, and Lisowski (1983) and Orvis and Gahart (1990).

system. Social Security numbers (SSNs) were used to link information on applicants with information on entrants. Details of how this link was accomplished are given in my earlier paper on military applicants (Angrist (1993a)).

After matching applicant records to information on entrants, a random sample from the resulting data set was matched to SSA earnings histories. The sample matched to earnings was limited to men aged 17–22 who applied during calendar years 1976–82, had valid sex and race codes, data on Armed Forces Qualification Test (AFQT) scores collected on certain ASVAB test forms,[16] and at least a 9th grade education but no more than a 4-year college degree. The target population contains 2.2 million white men and 900,000 nonwhite men. The sample drawn from this population contains roughly 750,00 applicants, half of whom are nonwhite. Conditional on race, over 90 percent of this sample is self-weighting, but some groups were over-sampled to satisfy SSA confidentiality requirements. The overall sample size was determined by the SSA. Additional information about the sample design is provided in the data appendix.

The SSA keeps track of the earnings of all workers covered by Social Security in a data base called the Summary Earnings Record (SER). SSA programmers were able to locate earnings records for 697,944 applicants in the full sample of 753,095. Data on Social Security taxable earnings were then added to the applicant sample, generating a micro data set containing information collected at the time of application, veteran status, and Social Security earnings for each year from 1974 through 1991.

SSA earnings data have a number of limitations. First, the SER records a zero for any individual without earnings from covered employment in a given year. However, almost all nongovernmental employees are covered by the main Social Security programs, OASDI (old-age, survivors, and disability insurance), and HI (Medicare).[17] Zeros in the SER may also be attributable to the fact that earnings information is recorded by the SSA with a lag.[18] Another limitation of SSA data is that earnings records for each year are censored at the maximum amount subject to FICA taxation (the "taxable maximum") for that year. Since 1980, however, at least 90 percent of all employee earnings have been reported to the Social Security Administration. Over 85 percent of employee plus self-employment earnings have been reported, and over 85 percent of covered male workers have earnings that fall below the taxable maximum (US Department of Health and Human Services (1993, Tables 3.B2 and 4.B2)).

Another limitation is that the SSA does not release individual earnings data to researchers. The micro data set linking 697,644 military applicants to their

[16] The applicant population was restricted to those tested on ASVAB Form 5/6/7 or later so that AFQT scores would be roughly comparable across years, and because this form was the first used extensively in the ASVAB testing program (Maier and Truss (1983)).

[17] About 95 percent of all jobs in the United States were covered as of 1991 (US Department of Health and Human Service (1993, page 9)). Members of the armed services have been covered since 1956 and most federal employees have been covered since 1983. Coverage of earnings from self-employment remains incomplete.

[18] The problem of late reporting of Social Security earnings is discussed by Card and Sullivan (1988) and in the Appendix to Angrist (1990).

earnings was therefore used to produce a data set containing average earnings for each of 8,760 cells defined by race, year of application, AFQT score group, veteran status, schooling level at the time of application, year of birth, and a variable for applicants in 1977–78 that was not used in this project. The released data set contains cell-identifiers for the 5,654 cells with 25 or more observations in the population, along with the average FICA earnings in these cells for 1974–91.[19] Additional earnings variables for each cell include standard deviations, the fraction with zero earnings, and the fractions with earnings at or above the taxable maximum.[20] The data released by the SSA were also subject to a confidentiality-edit that masks some or all of the earnings variables in small cells in the sample. This confidentiality edit is described in the data appendix, which also provides additional information on FICA coverage. In what follows, I refer to the released data set as the DMDC-SER matched sample.

### 3.1 *Descriptive Statistics*

The fraction of white applicants who ended up enlisting ranges from a high of 55 percent in 1979 to a low of 49 percent in 1981. The corresponding figures for nonwhite applicants range from a high of 50 percent in 1978 to a low of 36 percent in 1981. These figures are given in Panel A of Table I, which provides descriptive statistics for the population of applicants to the military from which the sample for DMDC-SER match data was drawn. The large increase in the number of applicants in 1980 is the result of increases in veterans benefits, military pay, and expenditures devoted to Army recruiting efforts in the early 1980's (Gilroy, Phillips, and Blair (1990)). Declines in the fraction of applicants who enlisted in 1980 and 1981 are partly the result of increases in entry standards generated by corrections to the ASVAB score scale.

The proportion veteran is slightly higher in the DMDC-SER matched sample than in the population. This can be seen in Panel B of Table I, which shows the size and proportion veteran by year of application for applicants in the DMDC-SER match. The excess representation of veterans is probably because military earnings are FICA-covered, so that all veterans have had at least one job covered by FICA. The control group is also limited to nonveterans who have had at least one FICA-covered job.

The typical applicant in the DMDC-SER matched sample was aged 18–20 at the time he applied, had an 11th or 12th grade education, and scored in the lower to middle range of the AFQT score scale.[21] Other statistics not shown in Table I indicate that roughly 30 percent of applicants in the sample were aged

[19] The lower bound on population cell size is a confidentiality restriction. Note that while many cells have fewer than 25 observerations, over 98 percent of the population is in larger cells.

[20] Persons with multiple employers can have earnings above the FICA taxable maximum on the SER.

[21] DMDC data on AFQT scores include up to 3 score scales, depending on the year of application and whether the applicant's record has been updated. All of the AFQT scores used in this project are compatible and based on what is known as the corrected 1944 score scale. I converted other scales to this scale using unpublished tables provided by the DMDC.

TABLE I

APPLICANT POPULATION AND SAMPLE

| | Application Year | | | | | | |
|---|---|---|---|---|---|---|---|
| Race | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | 1982 |
| A. *Population*[a] | | | | | | | |
| White | 339.5 | 286.9 | 235.9 | 253.1 | 348.6 | 387.3 | 309.8 |
| Percent veteran[b] | 53 | 52 | 54 | 55 | 53 | 49 | 52 |
| Nonwhite | 128.6 | 114.8 | 103.6 | 119.5 | 134.3 | 149.3 | 112.5 |
| Percent veteran | 44 | 46 | 50 | 46 | 41 | 36 | 43 |
| B. *Sample*[c] | | | | | | | |
| White | 49.2 | 46.5 | 40.0 | 39.4 | 52.9 | 57.9 | 47.3 |
| Percent veteran | 56 | 53 | 55 | 57 | 54 | 50 | 53 |
| Nonwhite | 50.9 | 48.1 | 44.6 | 51.9 | 57.0 | 63.7 | 48.7 |
| Percent veteran | 49 | 49 | 52 | 49 | 44 | 38 | 45 |

[a] The population is as in Angrist (1993a, Table 4), excluding those with less than a 9th grade education at the time of application. Numbers reported are thousands.
[b] Veterans are applicants identified as entrants to the military within two years following application.
[c] Approximately 90 percent of the sample is self-weighting, conditional on race.

18 when they applied to the military, 25 percent were aged 19, and 16 percent were aged 20. A total of 40 percent of applicants in the sample were high school graduates, 4 percent were GED certified, and 34 percent had completed 11th grade only. Out of nearly 700,000 applicants in the sample, only 739 were college graduates. In 1979, 67 percent of white applicants and 78 percent of nonwhite applicants had AFQT scores in categories III and IV, corresponding to the 10th through 64th percentiles of the AFQT reference population.

### 3.2. *Earnings Profiles*

Average earnings profiles for cohorts defined by application-year are shown in Figure 1. The profiles exhibit a high rate of initial earnings growth, with declining and even negative growth rates later.[22] This concavity is familiar from cross-section studies in labor economics. In this case, however, at least part of the decline in earnings growth at the end of sample period is attributable to the lags with which data on earnings are recorded in the SER. Another feature of the profiles is the dip in earnings growth around 1980–82. This is a business cycle effect that appears to have hit earlier applicants the hardest. The profiles also show an increase in the rate of earnings growth in the year of application.

[22] Earnings data were deflated to 1991 dollars using the Consumer Price Index (U.S. Bureau of the Census (1992, Table 738)).

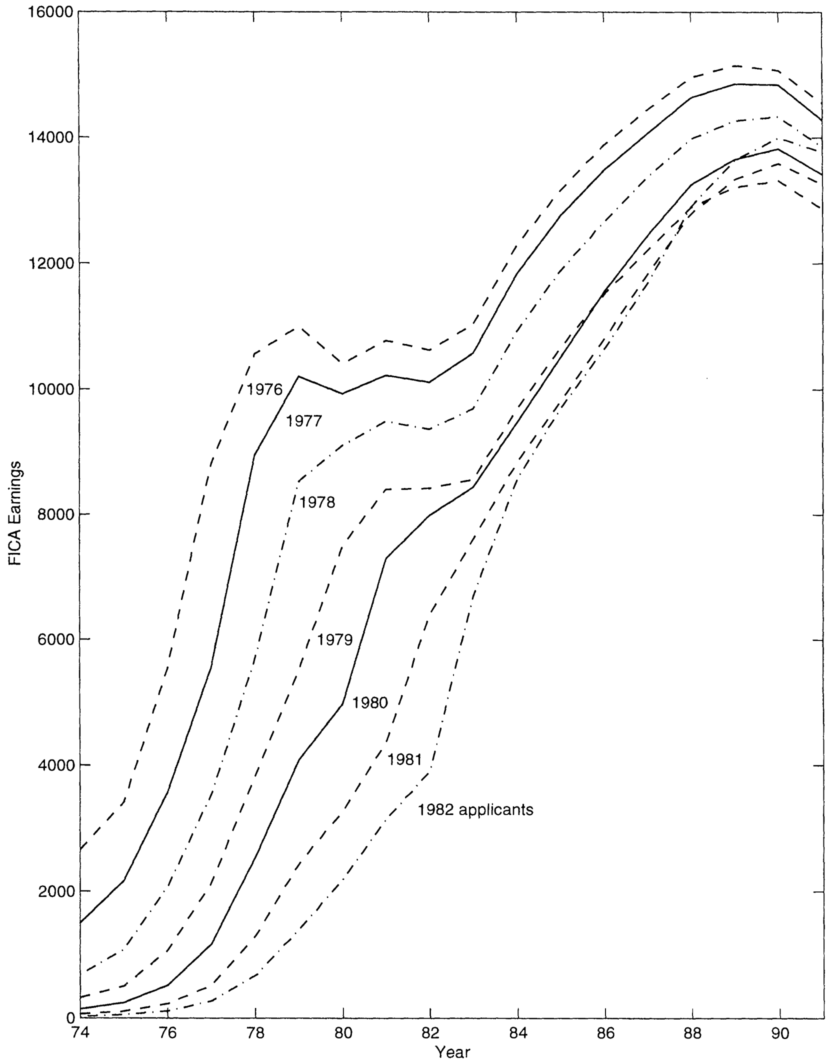


FIGURE 1.—Social Security earnings profiles of men who applied to the military in 1976–82, by year of application. Earnings are in 1991 dollars.

This is partly due to a reduction in the number of men with zero earnings in that year.

Although Figure 1 shows profiles for each of the 1976–82 applicant cohorts, the remaining analysis is limited to the sample of applicants who applied from 1979–82, with AFQT scores in groups III and IV. There are 128,968 whites and 175,262 nonwhites in this restricted sample, which also excludes a handful of college graduates. There are a number of reasons for these restrictions. First, identification using instrumental variables requires that treatment effects be independent of application year, an assumption more likely to be satisfied when the earlier applicant years are excluded because the package of veterans benefits offered in 1979 was usually threadbare.[23] Beginning in 1979, applicants were offered increasingly generous packages of veterans benefits that remain in place today. Moreover, many of the veterans who applied from 1976–78 probably had an unusually rough start as new entrants in the civilian labor market during the 1980–82 recession. The sample is restricted to applicants with AFQT scores in the middle range because those most affected by recent increases in military entrance standards consist of men with AFQT scores in groups III and IV.[24] And, as noted above, the majority of applicants of both races have scores that fall in this range.

In the sample of 1979–82 applicants with AFQT scores in categories III and IV, veterans earned more than nonveterans in every year in which they applied to the military. This can be seen in Figure 2, which plots earning profiles by veteran status and application year (constructed using population cell counts to aggregate individual cell means). Differences in earnings by veteran status are reported with standard errors in columns 2 and 6 of Table II, separately by race. To save space, differences in the table are for earnings averaged across application cohorts. Because the sample is so large, all of the post-1978 differences are very precisely measured and significantly different from zero. Some of the earlier small differences are significant as well. The veteran earnings gap reached a peak of 1500 dollars for whites and 2900 dollars for nonwhites in 1982–83, and remained substantial through the end of the sample period.[25] The fact that pre-application-year differences are small tends to support the interpretation of the veteran/nonveteran contrast as an unbiased estimate of $E[Y_1 - Y_0 \mid D = 1]$. In Section 4, however, I show that these simple contrasts are misleading.

### 3.3 *Employment Rates*

The Social Security earnings data analyzed here include observations of zero earnings. An alternate use of Social Security data, originally suggested by Card

[23] The main benefit for schooling in 1976–82 was the Veterans Educational Assistance Program (VEAP), which offered benefit levels considerably lower than those offered under the Post-Korean GI Bill. Beginning in 1979, however, the individual services (especially the Army) began to offer a range of additional benefits which, in combination with VEAP, approached the benefit levels in previous and later programs (Angrist (1993b)).

[24] The number of enlistments by men with AFQT scores in group IV fell from 17,700 in 1989 to 340 in 1992. Enlistments by applicants scoring in AFQT group IIIb fell by almost 40 percent over the same period (Angrist (1993a, Table 2)).

[25] For veterans, 1981–83 earnings data mostly reflect military earnings, while data from later years increasingly reflect civilian earnings. My tabulations from the 1990 Census show that the median length of service for AVF veterans aged 24–33 in 1990 (this includes men aged 17–22 in any year from 1979–1982) was between 3 and 4 years. Only 25 percent served more than 5 years and less than 15 percent were still in the military in 1990.

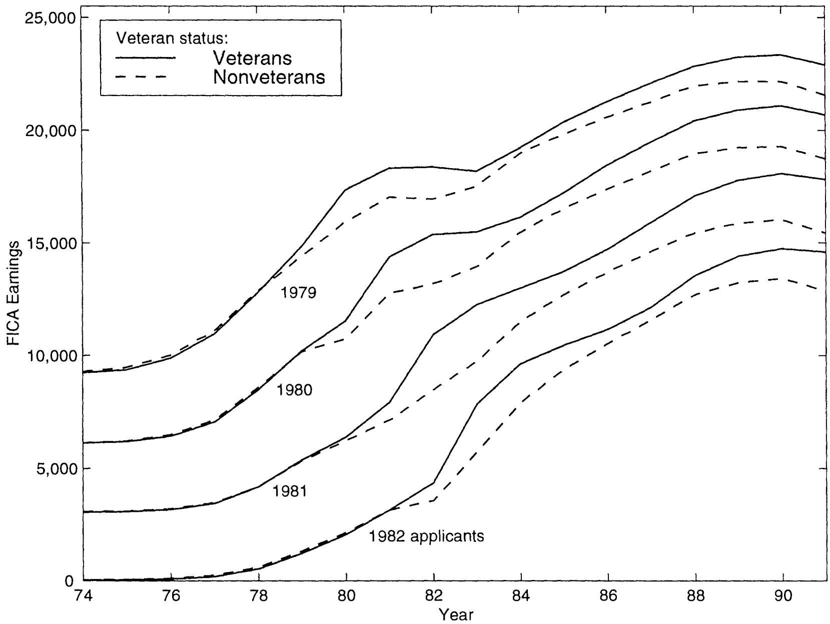


FIGURE 2.—Earnings profiles by veteran status and application year for men who applied 1979–82, with AFQT scores in categories III and IV. The plot shows the actual earnings of men who applied in 1982, earnings + \$3,000 for men who applied in 1981, earnings + \$6,000 for men who applied in 1980, and earnings + \$9,000 for men who applied in 1979.

and Sullivan (1988), is to treat an indicator of nonzero earnings itself as an outcome variable. Averages of this indicator can be interpreted as employment rates. An analysis of employment rates is possible in this case because the DMDC-SER match includes information on the fraction of each cell with zero earnings.

There are important problems associated with the use of Social Security zeros to measure employment rates. Potentially most serious is the fact that a zero could represent employment in the uncovered sector or filing delay on the part of employers or the SSA. This possibility is of particular concern here because veterans are more likely to be employed in the public sector, where workers are less likely to have FICA-covered earnings.[26] However, FICA coverage was extended to most sectors of the US economy in the 1980's, including many previously uncovered jobs in the public sector. Bias from filing delay is not reduced by these extensions, of course, but filing delay seems unlikely to affect

[26] Chay (1995) analyzes a match of March CPS data to Social Security data from the 1960's and 1970's and finds that most of the men with zero Social Security earnings were employed in the public sector.

TABLE II

ALTERNATIVE ESTIMATES OF THE EFFECTS OF MILITARY SERVICE

| | Whites | | | | Nonwhites | | | |
|---|---|---|---|---|---|---|---|---|
| Year | Mean (1) | Difference in Means[c] (2) | Controlled Contrast (3) | Regression Estimates (4) | Mean (5) | Difference in Means (6) | Controlled Contrast (7) | Regression Estimates (8) |
| A. *Earnings*[a] | | | | | | | | |
| 74 | 182.7 | −26.1 | −14.0 | −13.0 | 157.2 | −4.9 | −2.0 | −3.9 |
| | | (7.0) | (9.2) | (9.4) | | (4.4) | (6.0) | (5.8) |
| 75 | 237.9 | −41.4 | −14.2 | −12.0 | 216.9 | −.6 | −17.1 | −15.2 |
| | | (6.3) | (7.6) | (7.8) | | (4.5) | (6.0) | (5.5) |
| 76 | 473.4 | −47.9 | −14.8 | −12.7 | 413.6 | −14.5 | −33.3 | −30.2 |
| | | (8.1) | (9.0) | (9.3) | | (6.4) | (8.0) | (7.4) |
| 77 | 1012.9 | −7.1 | −8.6 | −9.4 | 820.9 | −13.0 | −56.0 | −51.3 |
| | | (11.3) | (12.3) | (12.2) | | (9.1) | (11.1) | (10.0) |
| 78 | 2147.1 | 40.3 | −23.5 | −22.4 | 1677.9 | 58.1 | −53.6 | −42.5 |
| | | (16.7) | (18.1) | (17.2) | | (13.4) | (16.1) | (14.1) |
| 79 | 3560.7 | 188.0 | −8.4 | −11.2 | 2797.0 | 340.3 | 119.1 | 122.3 |
| | | (21.0) | (23.2) | (21.6) | | (16.2) | (20.1) | (17.2) |
| 80 | 4709.0 | 572.9 | 178.0 | 175.9 | 3932.2 | 1154.3 | 741.6 | 738.5 |
| | | (23.4) | (27.2) | (24.6) | | (18.0) | (23.4) | (19.5) |
| 81 | 6226.0 | 855.5 | 249.5 | 249.9 | 5218.8 | 1920.0 | 1299.9 | 1318.5 |
| | | (27.2) | (32.4) | (29.1) | | (20.7) | (28.2) | (23.1) |
| 82 | 7200.6 | 1508.5 | 783.3 | 782.4 | 6150.2 | 2917.1 | 2186.0 | 2210.1 |
| | | (30.3) | (36.4) | (32.5) | | (23.4) | (32.0) | (26.0) |
| 83 | 8398.1 | 1390.5 | 588.8 | 601.5 | 7221.1 | 2889.9 | 2103.8 | 2142.3 |
| | | (34.4) | (41.1) | (36.6) | | (27.0) | (36.7) | (29.8) |
| 84 | 9874.2 | 652.8 | −235.7 | −198.5 | 8377.2 | 2202.9 | 1333.0 | 1428.9 |
| | | (39.5) | (46.9) | (41.7) | | (30.5) | (41.4) | (33.4) |
| 85 | 10972.7 | 469.8 | −521.3 | −459.6 | 9306.8 | 1955.5 | 932.3 | 1059.2 |
| | | (44.6) | (52.6) | (46.8) | | (34.4) | (46.2) | (37.3) |
| 86 | 12004.5 | 543.7 | −557.3 | −491.7 | 10106.2 | 1881.3 | 720.9 | 872.3 |
| | | (50.4) | (59.0) | (52.5) | | (38.7) | (51.2) | (41.6) |
| 87 | 13045.7 | 663.9 | −548.0 | −464.3 | 10833.0 | 2050.1 | 751.0 | 925.0 |
| | | (54.6) | (63.9) | (56.8) | | (41.8) | (55.2) | (44.8) |
| 88 | 14136.1 | 904.3 | −415.5 | −311.7 | 11480.1 | 2175.0 | 708.2 | 923.7 |
| | | (58.3) | (68.2) | (60.6) | | (44.9) | (59.5) | (48.1) |
| 89 | 14716.1 | 1169.1 | −248.6 | −136.3 | 11751.4 | 2379.1 | 799.7 | 1031.9 |
| | | (61.0) | (71.2) | (63.2) | | (47.6) | (62.7) | (50.9) |
| 90 | 14886.1 | 1300.8 | −154.5 | −53.2 | 11904.3 | 2483.6 | 824.9 | 1064.0 |
| | | (63.0) | (73.6) | (65.2) | | (49.4) | (65.4) | (52.7) |
| 91 | 14407.9 | 1559.6 | 29.8 | 146.2 | 11518.7 | 2758.8 | 1026.1 | 1277.9 |
| | | (64.6) | (75.6) | (66.9) | | (50.8) | (67.2) | (54.3) |
| B. *Employment Rates*[b] | | | | | | | | |
| 74 | 11.8 | −1.5 | −.43 | −.5 | 10.8 | −.58 | −.78 | −.7 |
| | | (.29) | (.38) | (.4) | | (.18) | (.22) | (.3) |
| 75 | 13.6 | −2.4 | −.8 | −.7 | 13.6 | .1 | −.62 | −.5 |
| | | (.22) | (.28) | (.3) | | (.17) | (.22) | (.2) |
| 76 | 20.8 | −2.0 | −.76 | −.7 | 20.6 | −.52 | −.65 | −.6 |
| | | (.22) | (.25) | (.2) | | (.18) | (.23) | (.2) |
| 77 | 34.7 | .000 | −.46 | −.5 | 32.9 | −.60 | −1.22 | −1.0 |
| | | (.23) | (.26) | (.2) | | (.20) | (.25) | (.2) |

TABLE II—*Continued*

| | Whites | | | | Nonwhites | | | |
|---|---|---|---|---|---|---|---|---|
| Year | Mean (1) | Difference in Means[c] (2) | Controlled Contrast (3) | Regression Estimates (4) | Mean (5) | Difference in Means (6) | Controlled Contrast (7) | Regression Estimates (8) |
| 78 | 54.4 | .7 | −.44 | −.5 | 50.3 | .8 | −.53 | −.4 |
| | | (.23) | (.26) | (.2) | | (.21) | (.25) | (.2) |
| 79 | 71.4 | 2.5 | .6 | .5 | 66.6 | 3.2 | 1.6 | 1.6 |
| | | (.22) | (.24) | (.2) | | (.20) | (.24) | (.2) |
| 80 | 81.2 | 4.9 | 3.0 | 3.0 | 75.8 | 7.9 | 6.1 | 6.2 |
| | | (.20) | (.22) | (.2) | | (.18) | (.22) | (.2) |
| 81 | 88.5 | 7.0 | 5.4 | 5.4 | 82.2 | 11.4 | 9.6 | 9.6 |
| | | (.17) | (.19) | (.2) | | (.17) | (.21) | (.2) |
| 82 | 89.9 | 10.8 | 8.6 | 8.8 | 83.6 | 17.9 | 15.6 | 15.7 |
| | | (.17) | (.18) | (.2) | | (.17) | (.21) | (.2) |
| 83 | 91.0 | 9.6 | 7.7 | 8.0 | 85.4 | 15.9 | 13.7 | 14.2 |
| | | (.16) | (.18) | (.2) | | (.16) | (.20) | (.2) |
| 84 | 92.2 | 6.6 | 5.4 | 5.5 | 87.9 | 11.1 | 9.6 | 10.0 |
| | | (.15) | (.17) | (.2) | | (.15) | (.18) | (.2) |
| 85 | 91.9 | 5.3 | 4.2 | 4.3 | 88.4 | 9.3 | 7.9 | 8.3 |
| | | (.15) | (.17) | (.2) | | (.15) | (.18) | (.2) |
| 86 | 91.1 | 4.9 | 3.8 | 4.0 | 87.6 | 8.6 | 7.0 | 7.4 |
| | | (.16) | (.18) | (.2) | | (.15) | (.18) | (.2) |
| 87 | 90.8 | 4.5 | 3.5 | 3.6 | 87.3 | 8.2 | 6.7 | 7.2 |
| | | (.16) | (.18) | (.2) | | (.15) | (.19) | (.2) |
| 88 | 90.7 | 3.9 | 3.0 | 3.1 | 86.9 | 7.8 | 6.3 | 6.7 |
| | | (.16) | (.18) | (.2) | | (.16) | (.19) | (.2) |
| 89 | 89.9 | 4.0 | 3.1 | 3.2 | 85.6 | 7.8 | 6.3 | 6.7 |
| | | (.17) | (.19) | (.2) | | (.16) | (.20) | (.2) |
| 90 | 88.8 | 4.2 | 3.2 | 3.3 | 83.8 | 8.0 | 6.3 | 6.7 |
| | | (.18) | (.20) | (.2) | | (.17) | (.21) | (.2) |
| 91 | 86.4 | 4.7 | 3.2 | 3.5 | 80.5 | 8.9 | 6.6 | 7.1 |
| | | (.19) | (.21) | (.2) | | (.19) | (.22) | (.2) |

[a] Panel A shows average earnings, differences in earnings by veteran status, controlled contrasts (matching estimates), and regression estimates for 1979–82 applicants with AFQT scores in groups III and IV. This sample includes 128,968 white applicants and 175,262 nonwhite applicants, about half of whom served in the military.
[b] Panel B presents similar statistics for employment rates (the fraction with positive earnings).
[c] Standard errors for each contrast are given in parentheses.

veterans differently from nonveterans. Moreover, the earnings records used here were last updated in 1992 and should be largely complete through 1989.

To provide an empirical assessment of filing delay and coverage problems, I drew an extract from the 1990 Census 5 Percent Public Use Microdata Sample (Bureau of the Census (1993)) that is roughly comparable with the sample used here. The Census extract includes over 900,000 men aged 24–33 in 1990, who were US citizens and had at least a ninth grade education. Of the whites in this sample, 13.5 percent are veterans. The corresponding figure for nonwhites is 17.6 percent. The Census was chosen for comparison because, unlike most US surveys, it includes data on active-duty military personnel. A drawback of the Census is that there is no way to limit the sample to those who applied for military service.

For whites, the Census fraction with positive wage and salary earnings in 1989 is 90.8 percent, while the DMDC-SER data for 1989 generate an employment rate of 89.9 percent. The Census employment rate for nonwhites is 82.3 percent, close to the corresponding DMDC-SER rate of 85.6 percent. Thus, the comparisons of employment levels suggest that information in the DMDC-SER matched data set is virtually complete. Also relevant is the comparison across data sets of gaps in 1989 employment rates by veteran status. For whites, this is 3 percent in the Census data and about 4 percent in the DMDC-SER data. The corresponding gap for nonwhites is 8.7 percent in the Census data and 7.8 percent in the DMDC-SER data. The fact that veteran employment gaps are similar in the two data sets suggests that veteran/nonveteran comparisons in the DMDC-SER data are not biased by a problem of under-coverage of veterans in public sector jobs.

The employment rates in the DMDC-SER matched sample range from a low of 10–12 percent in 1974 to a high of 92 percent for whites and 88 percent for nonwhites in 1984 and 1985. These statistics are from columns 1 and 5 in Panel B of Table II, which gives the average employment rate in each year, calculated as the fraction with positive wage and salary earnings. The low rates at the beginning of the sample period are to be expected for a sample of very young men. The declining rates at the end of the sample period are at least partly due to filing delay; a similar decline is apparent in the Card and Sullivan (1988) and Angrist (1990) Social Security data.

Comparisons of DMDC-SER employment rates by veteran status for each year suggest that a significant part of the veteran earnings advantage is attributable to higher employment rates for veterans. Differences in employment rates by veteran status are reported in columns 2 and 6 of Panel B in Table II. The employment contrasts by veteran status show that in years after 1977, veterans are more likely to be working than nonveterans. The gap in favor of veterans reaches a high of 11 percent for whites and 18 percent for nonwhites in 1982. The gap falls in later years to around 4–5 percent for whites and 8–9 percent for nonwhites.

## 4. MATCHING AND REGRESSION ESTIMATES

### 4.1 *Matching Estimates*

Because the DMDC-SER sample is confined to applicants to the AVF, comparisons of earnings by veteran status such as those in Table II control for veterans' decisions to apply to the military. On the other hand, veterans are carefully selected by the military on the basis of personal characteristics, like schooling and test scores, that are clearly related to future earnings. This fact motivates the matching estimator discussed in Section 2. It is worth mentioning again, however, that the modest pre-application differences observed in Figure 2 and Table II provide little evidence of selection bias. Of course, part of the problem with the use of such early comparisons as a specification check is that earnings or labor force participation as a teenager may not be related to earnings potential as an adult.

Matching estimates of veteran effects suggest that the simple comparisons of earnings by veteran status overestimate the effect of military service on earnings and employment. Estimates of $\hat{\alpha}_c$ for the earnings of whites, reported in column 3 of Table II and averaged over application cohorts, range from a high of only 783 dollars in 1982 to a low of −557 dollars in 1986. Standard errors for these estimates are less than 60 dollars.[27] Estimates of $\hat{\alpha}_c$ for the earnings of nonwhites are much larger than those for whites, although they are also substantially smaller than the corresponding simple comparisons. The largest estimate for nonwhites is 2,186 dollars in 1982 and the smallest is 708 dollars in 1988. The 1991 estimate of 1,026 dollars for nonwhites is less than 9 percent of nonwhite's average 1991 FICA earnings. The 1991 estimate for whites is about 30 dollars and is not statistically different from zero. Moreover, $\hat{\alpha}_c$ is negative for whites in every year after 1983 except 1991.

As with the uncontrolled comparisons by veteran status, estimates of $\hat{\alpha}_c$ for employment rates suggest that a substantial component of the effect of veteran status on earnings is an increase in employment rates. In years when most of the veterans would have been in the military (1981–83), employment rates for white veterans are 5.4 to 8.6 percent higher than those of nonveterans while employment rates for nonwhite veterans are 9.6 to 15.6 percent higher than those of nonveterans. Matching estimates of the veteran employment advantage fall to around 3 percent in later years for white veterans. Estimates of the veteran employment advantage for nonwhite veterans remain above 6 percent through 1991. The difference between $\hat{\alpha}_c$ and the simple comparisons for both earnings and employment is largely explained by the fact that, among applicants, veterans have higher test scores and are more likely than nonveterans to be high school graduates.

### 4.2 *Additional Matching Estimates*

A simple but important check on the matching estimates is to see whether the pattern of estimates by application year matches the expected timing of entry and discharge from military service for applicants who applied in different years. We have already seen that pre-application year differences by veteran status are small even in uncontrolled comparisons, so that the pre-application contrast is obviously not a foolproof specification test. But the time series of treatment effects by application year contains additional information of interest, such as changing effects as veterans are discharged from the military and reenter the civilian labor market.

[27] The variance of $\hat{\alpha}_c$ was computed treating $\delta_k$ and $N_{1k}$ as fixed and assuming that veteran and nonveteran earnings are statistically independent.

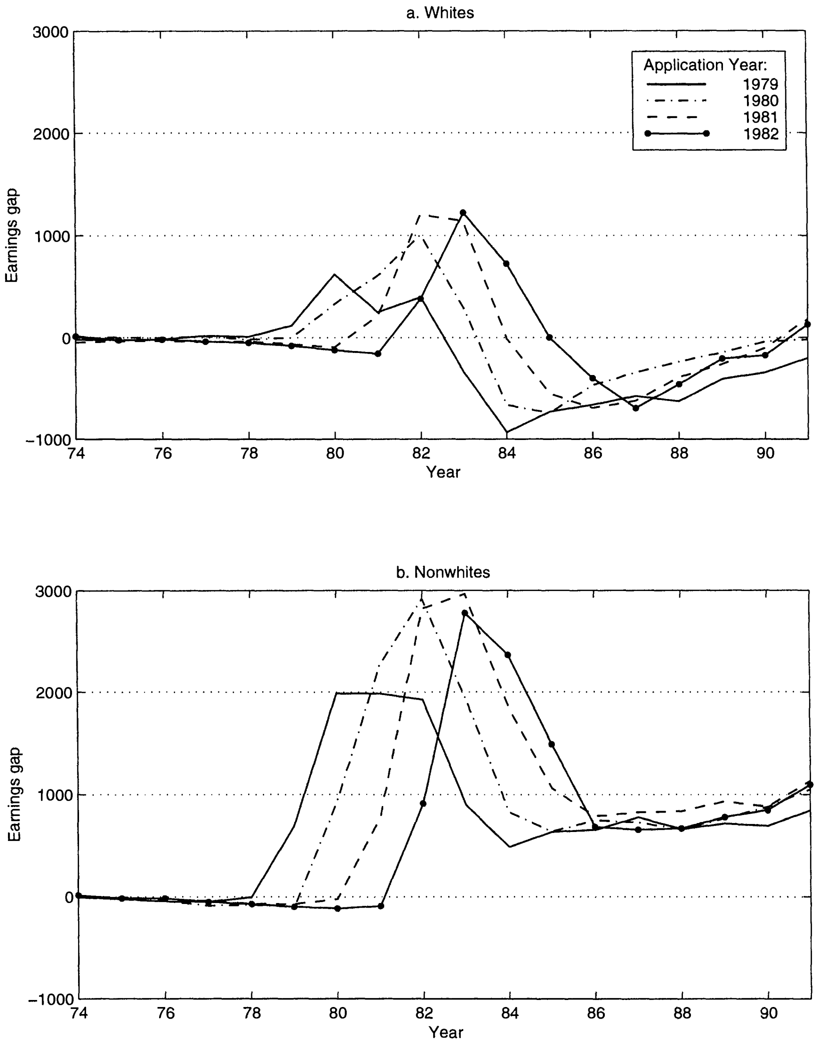


FIGURE 3.—Controlled contrasts by application year and calendar year for whites (a) and nonwhites (b).

Estimates of $\hat{\alpha}_c$ by application year, plotted in Figure 3 for both race groups, show that $\hat{\alpha}_c$ is close to zero until the year of application, at which time there is a sharp break. This break is especially well-defined for nonwhite applicants. After the break, veterans earn considerably more than nonveterans for a few years, while subsequent declines in $\hat{\alpha}_c$ are spaced at one-year intervals, beginning in 1983 for 1979 applicants. This suggests that the estimates of $\hat{\alpha}_c$ reflect a staggered return to civilian life by most of the veterans in each cohort, since the median length of service for men in this sample is between 3 and 4 years. The negative veteran effects for whites in the years when they were leaving the military may be due in part to college attendance. Note, however, that these negative effects persist through 1990.

Another feature of interest in Figure 3 is the modest dip in earnings for some cohorts just before the year of application. Similar pre-program declines in earnings have been observed in studies of subsidized training programs (Ashenfelter (1978), Ashenfelter and Card (1985)) and are a potential source of bias in nonexperimental evaluations. Because earnings are serially correlated, transient negative shocks could have a lasting impact on veteran-nonveteran contrasts. But the small pre-application declines in earnings observed in Figure 3 seem unlikely to have imparted any substantial bias in the estimates for later years.

Finally, note that Figure 3 suggests that by 1986, after the majority of veterans in this sample would have left the military, the effects of military service are similar across application years. This is important because, as noted in Section 2.3, the causal interpretation of the IV estimates turns partly on the assumption that treatment effects are independent of the instruments (application year dummies) conditional on AFQT score group and other covariates.

### 4.3 *Regression Estimates*

A natural question raised by the matching results is whether a regression strategy generates similar estimates. To describe the regression estimator used here, recall that the random variable $X$ takes on values to indicate each possible combination of year of birth, schooling, application year, and AFQT score group. Conditional on race, these are the cell identifiers other than veteran status. In the sample of men who applied in 1979–82, with AFQT scores in groups III or IV, and earnings data in 1991 for both veterans and nonveterans, there are 466 possible values of $X$ for whites and 429 possible values of $X$ for nonwhites.[28] As before, index these possible values by $k = 1, \dots, K$. Observations on the dependent variable can then be written $\bar{y}_{Dk}$, denoting average earnings for men with veteran status $D$ and covariate-combination $k$.

Regression estimates of the effect of military service are based on the following model, estimated separately, for each calendar year and race group:

$$\bar{y}_{Dk} = \beta_k + \alpha_r D + \bar{\epsilon}_{Dk}, \tag{17}$$

where $\beta_k$ is an effect for covariate-combination $k$, $\alpha_r$ is a veteran effect, and $\bar{\epsilon}_{Dk}$ is an error term that is orthogonal to $D$ and $X$ by definition of $\beta_k$ and $\alpha_r$. Estimates of $\alpha_r$, denoted $\hat{\alpha}_r$, were computed by weighted least squares using population cell counts ($N_{Dk}$) as weights. This weighting scheme, when applied to grouped data, produces the same estimates as would be generated using micro data weighted by inverse sampling rates.

[28] The number of cells can vary from year to year because of the confidentiality edit.

In Section 2, $\alpha_r$ was shown to be a conditional-variance-weighted average of covariate-specific treatment effects, whereas $\alpha_c$ weights covariate-specific effects by the proportion of veterans at each value of $X$. In practice, the regression and matching estimates are almost identical through 1984. This can be seen in Table II, which reports $\hat{\alpha}_r$ as well as $\hat{\alpha}_c$ for each year. In contrast with the 1974–84 results, however, regression estimates for each year after 1984 are larger than the corresponding matching estimates. The largest difference is roughly 250 dollars, for the earnings of nonwhites in 1991. This is equal to about 25 percent of the corresponding matching estimate. Because the regression and matching estimates are highly correlated, the difference in the two estimates is precisely measured, with a standard error of about 20 dollars for the 1991 earnings of nonwhites.[29] Differences between $\hat{\alpha}_r$ and $\hat{\alpha}_c$ for earlier years, and for the earnings of whites, are estimated with equal or better precision.

The divergence between regression and matching estimates after 1984 is probably explained by differences in the long term impact of military service on men with covariate values that place them in low-probability-of-service and high-probability-of-service groups. This can be seen in Figure 4, which plots estimated treatment effects for average 1988–91 earnings conditional on probability of service (grouped into 7.5 point intervals). There is a strong negative relationship between treatment effects and the probability of service for both race groups. The matching estimator gives the small covariate-specific estimates for men with high probabilities of service the most weight, while the larger covariate-specific estimates for men with low probability of service are given less weight. The regression estimator, in contrast, gives more weight to covariate-specific estimates where the probability of military service conditional on covariates is close to one-half. This leads to a higher overall treatment effect.

## 5. ESTIMATION USING THE ASVAB MISNORMING

The estimates in Section 3 control for the major characteristics used by the military to screen applicants to the Armed Forces. It should be noted, however, that a large fraction of the applicants who do not enlist nevertheless appear to qualify for enlistment (Berryman, Bell, and Lisowski (1983)). This raises the possibility that even after controlling for observed covariates, potential earnings might not be independent of veteran status. The possibility of selection bias in the matching estimates motivates an IV strategy that relies on different identifying assumptions. The source of identifying information underlying these estimates are the changes in the probability of enlistment for applicants with low

[29] The covariance between the regression and matching estimates was calculated as follows. Both estimators can be written in the form $\Sigma_k\ \omega_{jk}[\bar{y}_{1k}-\bar{y}_{0k}]$, where $[\bar{y}_{1k}-\bar{y}_{0k}]$ is the $k$th covariate-specific treatment contrast and $\omega_{jk}$ (for $j=c$ or $j=r$), is the relevant weighting function. Assuming independence across cells and treating the weights as nonstochastic, the covariance is therefore $\Sigma_k\ \omega_{ck}\,\omega_{rk}\,V[\bar{y}_{1k}-\bar{y}_{0k}]$. Standard errors for the regression estimates were calculated in the same way as the standard error of the matching estimates, i.e., by treating the weights as fixed and assuming that the earnings of veterans and nonveterans are independent.

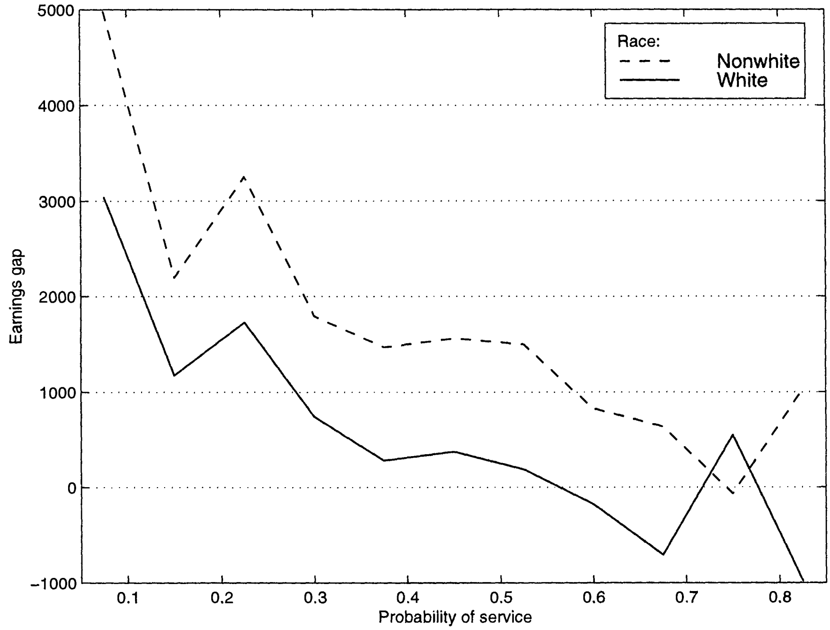


FIGURE 4.—Controlled contrasts by race and probability of service. These estimates are for pooled 1988–91 earnings.

AFQT scores documented in Table III, which shows that AFQT-IV applicants were as much as 29 percentage points less likely to enlist if they applied in 1981–82 than in 1979–80. In contrast, enlistment probabilities for AFQT-III applicants changed little.

IV estimates of the effect of veteran status were constructed by replacing $E[Y \mid W = w_k, S, A]$ with the corresponding sample mean in equations (11) and (12) using a weighted least squares procedure described in the Appendix. There is no need to use sample estimates for $E[D \mid W = w_k, S, A]$ because this proportion is known from population cell counts. In view of the requirement that treatment effects be independent of application years (Assumption 3), earnings data were dropped for each application-year cohort in that cohort's year of application, and data on the earnings of cohorts observed before the year of application were not pooled with data on the earnings of cohorts observed after the year of application.

The IV estimates of $\alpha^*$ in equation (11) are generally consistent with the matching and regression estimates discussed in Section 4. This can be seen in Figures 5 and 6, which plot the estimates for earnings and employment along with 95 percent confidence bands. Pre-treatment estimates are close to zero, though in some cases, significantly different from zero. Estimates for 1980–82, when most veterans in the sample were entering the military, are large and positive for both racial groups. In some cases, these positive estimates are even

TABLE III

PROBABILITIES OF ENLISTMENT BY AFQT SCORE AND APPLICATION YEAR

| Race | Application Year[a] | AFQT Score Group IVc (10–15)[b] | IVb (16–20) | IVa (21–30) | IIIb (31–49) | IIIa (50–64) |
|---|---|---|---|---|---|---|
| Whites | 1979 | 26.4 | 51.7 | 59.1 | 62.3 | 63.1 |
| | 1980 | 14.5 | 42.8 | 51.5 | 59.4 | 62.3 |
| | 1981 | 3.3 | 22.6 | 35.9 | 59.0 | 62.5 |
| | 1982 | 5.4 | 23.2 | 36.0 | 57.1 | 63.1 |
| | 1979–1982 Difference | 21.0 | 28.5 | 23.1 | 5.2 | 0 |
| Nonwhites | 1979 | 29.6 | 58.0 | 65.9 | 68.1 | 68.5 |
| | 1980 | 19.8 | 50.6 | 60.2 | 66.5 | 67.7 |
| | 1981 | 4.4 | 30.1 | 47.0 | 66.8 | 68.4 |
| | 1982 | 6.3 | 28.7 | 48.0 | 66.5 | 68.6 |
| | 1979–1982 Difference | 23.3 | 29.3 | 17.9 | 1.6 | –.1 |

[a] The table shows probabilities for the population of 1979–82 noncollege graduate applicants with AFQT scores in groups III and IV. This population includes 849,983 white applicants and 403,686 nonwhite applicants.
[b] AFQT percentile scores in parentheses.

larger than the matching and regression estimates. Also, like the matching and regression estimates, the IV estimates for the earnings of whites after 1982 are almost all negative, though not significantly different from zero. Note, however, that the IV estimates suggest that long-run employment effects may be zero or negative while the matching and regression estimators generate small positive effects.

In contrast with the negative long-run effects for whites, the IV estimates of equation (11) for nonwhite earnings and employment generally show positive effects. For example, estimates for the 1988–91 earnings of nonwhites are all positive and significantly different from zero. The pooled estimate for the 1988–91 earnings of nonwhites is 849 dollars with a standard error of 374. The pooled estimate for employment is 2.6 percent with a standard error of 1.2 percent.[30]

Equation (11) is identified by assuming that any earnings consequences of application year which occur other than as a result of veteran status can be captured by a linear function. Chi-square goodness of fit statistics for the estimates of equation (11), a measure of how well this model fits the cell means, suggest that this assumption may be violated for nonwhites. For example, the

[30] Pooled estimates were computed by using average earnings across years as the dependent variable. The variance of pooled earnings and employment rates was computed by assuming the cross-year correlation structure is constant across cells. See the appendix to Angrist (1990) for details of a similar calculation. Pooled estimates for employment include some cells excluded from the yearly estimates because there are fewer pooled cells with employment rates equal to zero or one (in which case the cell variance is estimated to be zero).

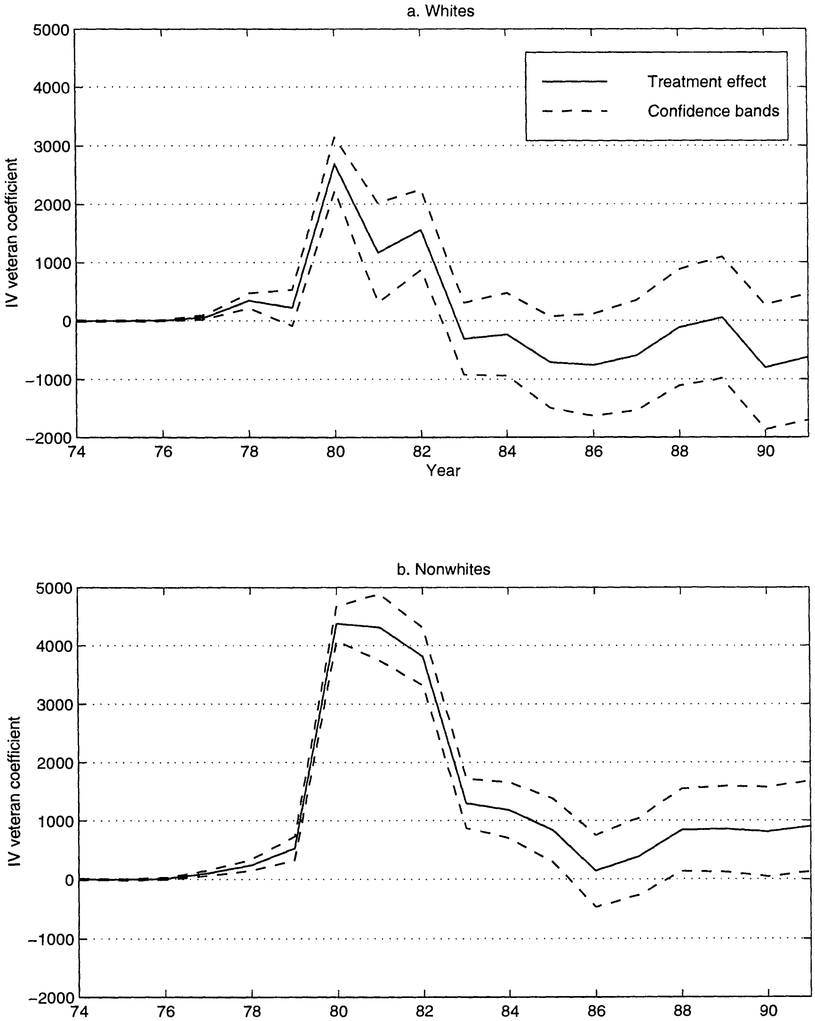


FIGURE 5.—Instrumental variables estimates of earnings effects from a model with linear controls for time-since-application for whites (a) and nonwhites (b).

test statistic for pooled 1988–91 earnings has a marginal significance level close to one percent. The strong restrictions embodied in equation (11) can be partially relaxed, however. In particular, the term $(t-A)\gamma$ can be replaced by $\delta_{kA}$, an additive application-year effect for each value of $W$, $w_k$. This leads to equation (12).

Except for years when most of the veterans would have been in the military, estimates of equation (12) show little evidence of an effect of veteran status on the earnings of whites, although there are positive employment effects through

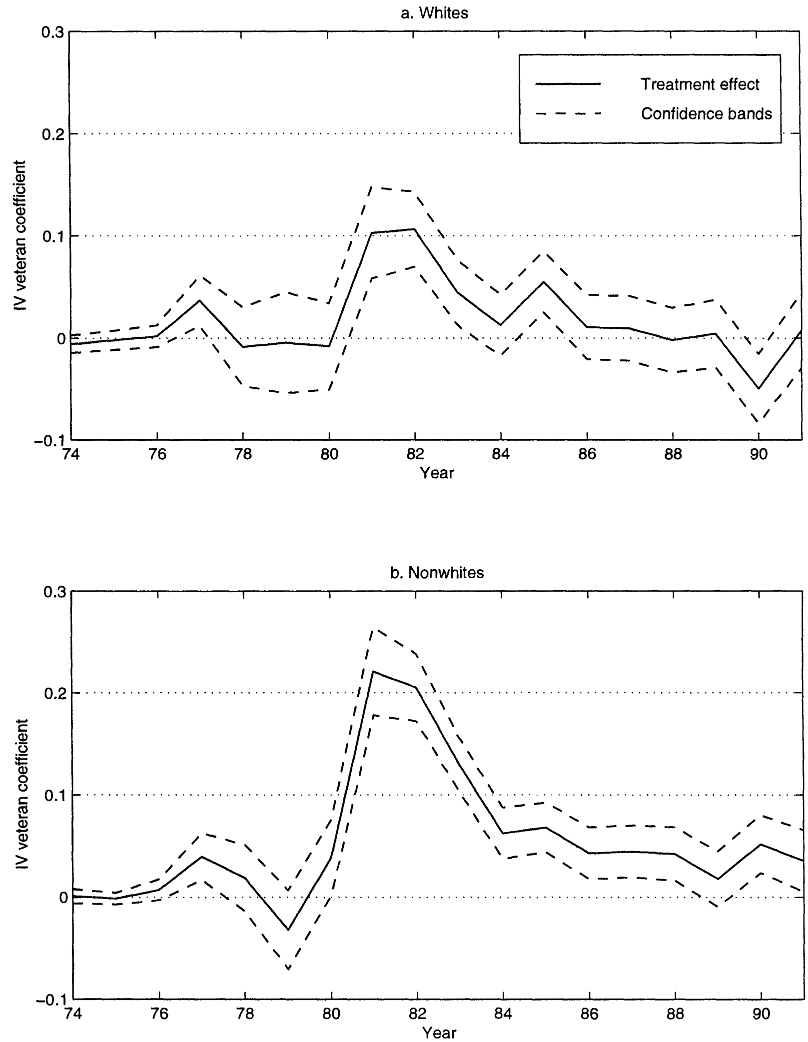


FIGURE 6.—Instrumental variables estimates of employment effects from a model with linear controls for time-since-application for whites (a) and nonwhites (b).

1988. These estimates are reported in Table IV for both whites and nonwhites. The IV estimates for nonwhites show large positive effects on earnings through 1984. Unlike the corresponding matching estimates, however, the IV estimates for nonwhites are negative and imprecise after 1985. On the other hand, IV estimates of employment effects for nonwhites are positive and significantly different from zero through 1988 and also in 1990. The pooled 1988–91 employment estimate for nonwhites is 3.3 percent with a standard error equal to 1.5 percent. A comparison of goodness-of-fit statistics for equations (12) and (11)

TABLE IV
IV ESTIMATES

| | Whites | | | | Nonwhites | | | |
|---|---|---|---|---|---|---|---|---|
| Year | Earnings Estimate[b] (1) | $\chi^2$ (2) | Employment Estimate (3) | $\chi^2$ (4) | Earnings Estimate (5) | $\chi^2$ (6) | Employment Estimate (7) | $\chi^2$ (8) |
| 74 | 2.5 | 93 | .43 | 79 | 1.1 | 102 | .11 | 117 |
| | (10.3) | (93) | (.79) | (93) | (6.5) | (113) | (.40) | (113) |
| 75 | −1.2 | 137 | −.4 | 163 | .6 | 187 | −.3 | 164 |
| | (6.0) | (156) | (.57) | (156) | (5.8) | (177) | (.32) | (177) |
| 76 | −5.8 | 189 | −.7 | 204 | .9 | 204 | .33 | 210 |
| | (8.3) | (211) | (.67) | (211) | (9.0) | (222) | (.55) | (222) |
| 77 | 16.2 | 243 | 2.95 | 295 | −22.0 | 259 | −.5 | 300 |
| | (28.3) | (252) | (1.70) | (247) | (30.5) | (243) | (1.56) | (241) |
| 78 | 204.9 | 289 | 7.63 | 297 | 42.2 | 273 | 2.34 | 303 |
| | (80.7) | (266) | (2.64) | (239) | (62.2) | (256) | (2.03) | (253) |
| 79 | 5.3 | 246 | 9.61 | 213 | 28.1 | 139 | 1.38 | 176 |
| | (199.5) | (182) | (3.23) | (160) | (142.9) | (173) | (2.34) | (170) |
| 80 | 1799.0 | 242 | 5.44 | 210 | 4487.6 | 188 | 17.57 | 154 |
| | (526.5) | (159) | (4.26) | (141) | (302.3) | (152) | (3.13) | (142) |
| 81 | 2449.3 | 200 | 12.66 | 159 | 4028.8 | 174 | 16.02 | 161 |
| | (592.9) | (155) | (3.70) | (139) | (356.5) | (147) | (3.28) | (136) |
| 82 | 2973.4 | 187 | 16.14 | 188 | 3625.0 | 160 | 21.72 | 173 |
| | (430.8) | (176) | (2.39) | (169) | (279.8) | (169) | (1.92) | (154) |
| 83 | 1652.1 | 314 | 9.33 | 338 | 2217.3 | 239 | 14.43 | 278 |
| | (408.3) | (268) | (2.12) | (250) | (262.8) | (257) | (1.65) | (248) |
| 84 | 153.9 | 257 | 3.38 | 246 | 957.3 | 267 | 6.47 | 284 |
| | (467.0) | (268) | (2.00) | (245) | (298.3) | (257) | (1.58) | (235) |
| 85 | −788.7 | 279 | 7.31 | 261 | 2.0 | 286 | 7.01 | 279 |
| | (517.1) | (268) | (2.00) | (244) | (335.6) | (257) | (1.52) | (246) |
| 86 | −595.7 | 238 | 2.89 | 238 | −685.6 | 311 | 4.73 | 261 |
| | (575.3) | (268) | (2.10) | (247) | (376.3) | (257) | (1.58) | (236) |
| 87 | 12.4 | 266 | 3.53 | 264 | −454.7 | 297 | 4.60 | 281 |
| | (613.5) | (268) | (2.11) | (249) | (403.7) | (257) | (1.57) | (242) |
| 88 | −4.6 | 275 | 1.68 | 254 | −381.3 | 292 | 5.48 | 302 |
| | (650.2) | (268) | (2.13) | (251) | (432.0) | (257) | (1.64) | (241) |
| 89 | −368.2 | 248 | −.3 | 287 | −456.7 | 312 | −.1 | 309 |
| | (676.5) | (268) | (2.22) | (250) | (450.0) | (257) | (1.70) | (243) |
| 90 | −995.6 | 252 | −5.6 | 291 | −578.5 | 315 | 3.85 | 345 |
| | (694.6) | (268) | (2.29) | (257) | (465.7) | (257) | (1.74) | (245) |
| 91 | −776.6 | 268 | −1.0 | 242 | −612.3 | 308 | .69 | 286 |
| | (700.9) | (268) | (2.45) | (264) | (473.7) | (257) | (1.85) | (253) |

[a] The table reports estimates and goodness-of-fit statistics for equation (12) in the text.
[b] Standard errors and chi-square degrees of freedom in parentheses.

(not reported) indicates that in most years equation (12) typically provides a significantly better fit to the cell means. The test statistics for equation (12), reported in Table IV, imply that in many years this model also passes an omnibus goodness-of-fit test.

Overall, the IV estimates suggest that the long-run consequences of military service may be even smaller than indicated by the matching estimates. For example, the IV estimates for whites show small negative employment effects while the corresponding matching estimates show small, though statistically significant, positive effects. Similarly, for nonwhites, the IV estimates of employment effects, while still positive, are smaller than the matching estimates. Moreover, unlike the matching estimates, which show modest positive effects on the earnings of nonwhites, one of the IV specifications for nonwhites (equation (12)) leads to negative and insignificant earnings effects.

There are at least three possible—and not mutually exclusive—explanations for the differences between the IV and matching estimates. First, the matching estimates may still be biased upwards because Assumption 1 fails to hold. For example, in addition to the covariates used here, applicants may use other information about their civilian labor market prospects when making enlistment decisions. Second, the identifying assumptions upon which the IV estimates are based may fail to be satisfied. In either of these two cases, there is no reason for the IV and matching estimates to be related in any particular way. Third, while the matching estimates capture the effect of military service on the entire population of veterans, the theorem in Section 2.3 shows that the IV estimates characterize the impact of treatment on a group that is not necessarily representative of the population of veterans or even of the population of low-scoring veterans. In particular, the IV estimates capture the effect of treatment on men with low scores who get into the military *because of misnorming*.

Men who get into the military with low test scores typically do so in spite of these scores. In contrast, men who got into the military because of the ASVAB misnorming are men with low test scores whom the military would have excluded if their true scores had been known (these are men with $s = 1$ and $D_1 > D_0$). On one hand, in light of the covariate-specific estimates in Figure 4, the fact that this group has low scores suggests that they should benefit more from military service than the entire population of veterans. On the other hand, unlike other low-scoring veterans, conditional on having low AFQT scores they were not actually qualified for military service. Some studies suggest that such unqualified recruits were less likely to complete their initial tour of duty, get promoted, reenlist, or obtain an honorable discharge (see, e.g., Heisey, Means, and Laurence (1985), Cooke and Quester (1992)). These factors may have reduced any long-run labor market benefits from military service.

## 6. SUMMARY AND CONCLUSIONS

The main empirical findings are summarized in Table V, which reports differences-in-means, matching, regression, and IV estimates of the effects of veteran status on earning and employment for selected post-application years. The last application year in the sample is 1982. In 1983, when most of the veterans studied here were still in the military, all but one of the estimates (IV-1 for whites) suggest that veterans earned more than nonveterans, and all of the estimates show that veterans were more likely to be employed. In 1984–87, when many veterans were reentering the civilian labor market, white veterans

TABLE V
SUMMARY OF RESULTS

| | | | Estimator | | | | |
|---|---|---|---|---|---|---|---|
| Race | Period | Variable | Difference in Means | Matching | Regression | IV-1 [eq. (11)] | IV-2 [eq. (12)] |
| *Whites* | During Service (1983) | Earnings | 1390.5 (34.4) | 588.8 (41.1) | 601.5 (36.6) | −309.9 (312.8) | 1652.1 (408.3) |
| | | Employment | 9.6 (.2) | 7.7 (.2) | 8.0 (.2) | 4.5 (1.6) | 9.3 (2.1) |
| | Transition Years (1984–87)[a] | Earnings | 582.6 (45.5) | −465.6 (53.5) | −403.5 (47.6) | −615.9 (407.3) | −329.5 (524.7) |
| | | Employment | 5.3 (.1) | 4.2 (.2) | 4.4 (.1) | −1.0 (1.3) | 3.4 (1.7) |
| | | Employment[b] | | | | 2.1 (1.4) | 3.9 (1.9) |
| | After Service (1988–91) | Earnings | 1233.4 (60.3) | −197.2 (70.5) | −88.8 (62.5) | −407.7 (521.8) | −548.5 (666.0) |
| | | Employment | 4.2 (.2) | 3.1 (.2) | 3.3 (.2) | −2.7 (1.4) | −2.5 (1.9) |
| | | Employment[b] | | | | −.3 (1.6) | −1.2 (2.1) |
| *Nonwhites* | During Service (1983) | Earnings | 2889.9 (27.0) | 2103.8 (36.7) | 2142.3 (29.8) | 1296.4 (215.6) | 2217.3 (262.8) |
| | | Employment | 15.9 (.2) | 13.7 (.2) | 14.2 (.2) | 12.9 (1.4) | 14.4 (1.7) |
| | Transition Years (1984–87) | Earnings | 2022.4 (35.2) | 934.3 (47.0) | 1071.3 (38.1) | 605.7 (283.9) | −62.8 (343.7) |
| | | Employment | 9.3 (.1) | 7.8 (.2) | 8.2 (.1) | 5.1 (1.1) | 5.8 (1.3) |
| | | Employment[b] | | | | 6.4 (1.2) | 5.7 (1.5) |
| | After Service (1988–91) | Earnings | 2449.1 (47.4) | 839.7 (62.7) | 1074.4 (50.7) | 849.1 (374.3) | −500.5 (449.4) |
| | | Employment | 8.1 (.2) | 6.4 (.2) | 6.8 (.2) | 2.6 (1.2) | 3.3 (1.5) |
| | | Employment[b] | | | | 4.3 (1.4) | 2.2 (1.7) |

[a] Pooled estimates use earnings and employment averaged over four years as the dependent variable.
[b] For comparability with the yearly estimates, these pooled estimates exclude cells with employment rates equal to zero or one in any calendar year.

were actually earning less than comparable nonveterans. This finding is consistent across all estimation methods and contradicts simple comparisons by veteran status, which show white veterans earning more than nonveterans in these years. In contrast with the negative estimates for the 1984–87 earnings of whites, most of the estimates for the 1984–87 earnings of nonwhites are positive, although one IV estimate is negative and very close to zero.

By 1988–91, when most veterans were no longer in the military, the earnings of white veterans and nonveterans appear to have converged. But most of the estimates suggest that nonwhite veterans may have enjoyed an earnings advantage through the end of the sample period, and all of the estimates suggest at least a small employment advantage for nonwhites. Although the long-run earnings effects for nonwhites are estimated to be less than 10 percent of average earnings, the finding of some long-run labor market benefits for nonwhite veterans stands in contrast with earlier negative results for veterans of the misnorming era (Lawrence, Ramsberger, and Gribben (1989)). The possibility of modest earnings benefits for nonwhite veterans should be taken into account by policy-makers concerned about the social cost of military downsizing. On the other hand, claims that blacks who serve in the military have enjoyed unparalled labor market benefits (see, e.g., Moskos and Butler (1996)) are not supported by the findings reported here. Even among applicants, simple comparisons by veteran status clearly exaggerate the effect of military service. The true long-run earnings effect seems to be on the order of the benefit from between one and two years of extra schooling, an alternative career-development option that is likely to be available to most applicants for military service.

Results from each of the estimation methods summarized in Table V have been interpreted as average causal effects that are valid for various groups in the population of veterans. Under the assumption that veteran status is ignorable conditional on the covariates used by the military to select from the applicant pool, the matching and regression estimators produce weighted averages of covariate-specific average treatment effects. The weighting schemes differ and regression does not, in general, estimate the effect of treatment on the treated. IV estimates also have a weighted average interpretation, but this average applies to a narrower group than either the regression or matching estimates. In this case, the IV estimates are for a group of low-scoring applicants that would have been disqualified if the military had accurate test scores; men in this group may have benefited less from military service than other low-scoring applicants.

The discussion in this paper highlights the fact that in evaluation research there is generally no single treatment effect of interest. In the absence of a theory explaining the response to treatment, researchers should expect methods that exploit different sources of variation to produce different results. On the other hand, except for the simple comparison of means, the estimators used here generate broadly similar results. This suggests that simple economic mechanisms may explain the coarse pattern of treatment effects. In this case, positive effects on the earnings of AVF veterans in the early eighties seem to be explained by the fact that veterans were insulated from a major cyclical downturn while in the military. The proximate cause for any longer-term positive effects on earnings would seem to be an increase in employment rates, explained perhaps by continued military service and the hiring preferences enjoyed by veterans in the public sector.

*Dept. of Economics, Massachusetts Institute of Technology, 50 Memorial Dr., Cambridge MA 02139-4307, U.S.A., and NBER; angrist@mit.edu*

*Manuscript received August, 1994; final revision received June, 1997.*

## APPENDIX A: PROOF OF THEOREM AND ESTIMATION DETAILS

I. PROOF OF THEOREM: Assumption 2 implies that we can write

$$Y = Y_0 + [Y_1 - Y_0]D = \beta_0 + \beta_1 s + \delta a + [Y_1 - Y_0][D_0 + (D_1 - D_0)a] + \epsilon \tag{A.1}$$

where $\epsilon$ is mean-independent of $a$ and $s$. Taking expectations conditional on $a$ and $s = 1$, we have

$$E[Y \mid a = 1, s = 1] = \beta_0 + \beta_1 + \delta + E[(Y_1 - Y_0)D_1 \mid a = 1, s = 1], \tag{A.2}$$

$$E[Y \mid a = 0, s = 1] = \beta_0 + \beta_1 + E[(Y_1 - Y_0)D_0 \mid a = 0, s = 1]. \tag{A.3}$$

Differencing A.2 and A.3 and using Assumption 3,

$$E[Y \mid a = 1, s = 1] - E[Y \mid a = 0, s = 1] = \delta + E[(Y_1 - Y_0)(D_1 - D_0) \mid s = 1]. \tag{A.4}$$

Likewise,

$$E[Y \mid a = 1, s = 0] - E[Y \mid a = 0, s = 0] = \delta + E[(Y_1 - Y_0)(D_1 - D_0) \mid s = 0] = \delta, \tag{A.5}$$

where the second equality is because $s = 0$ implies $D_1 - D_0 = 0$ by Assumption 4. Also by Assumption 4,

$$E[(Y_1 - Y_0)(D_1 - D_0) \mid s = 1] = E[Y_1 - Y_0 \mid D_1 > D_0, s = 1]P[D_1 > D_0 \mid s = 1].$$

Subtracting (A.5) from (A.4) therefore gives

$$\begin{aligned}(E[Y \mid a = 1, s = 1] - E[Y \mid a = 0, s = 1]) - (E[Y \mid a = 1, s = 0] - E[Y \mid a = 0, s = 0])\\ = E[Y_1 - Y_0 \mid D_1 > D_0, s = 1]P[D_1 > D_0 \mid s = 1].\end{aligned} \tag{A.6}$$

Using Assumptions 2 and 3 and the fact that $D = D_0 + (D_1 - D_0)a$, we have

$$\begin{aligned}&E[D \mid a = 1, s = 1] - E[D \mid a = 0, s = 1] = P[D_1 > D_0 \mid s = 1], \quad \text{and}\\ &E[D \mid a = 1, s = 0] - E[D \mid a = 0, s = 0] = 0.\end{aligned} \tag{A.7}$$

Putting (A.6) and (A.7) together establishes that $\alpha_{IV} = E[Y_1 - Y_0 \mid s = 1, D_1 > D_0]$.

Note that Assumptions 2–4 are not the weakest assumptions that can be used to give IV estimates a causal interpretation. For example, Assumption 4 can be modified to allow men with $s = 0$ to have $D_1 \leq D_0$ and not just $D_1 = D_0$:

ASSUMPTION 5: $P[s(D_1 - D_0) \geq (1 - s)(D_1 - D_0)] = 1$.

An argument similar to the one above shows that if Assumptions 2, 3, and 5 hold,

$$\alpha_{IV} = \frac{\begin{array}{c}\{E[Y_1 - Y_0 \mid s = 1, D_1 > D_0]P[D_1 > D_0 \mid s = 1]\\ +E[Y_1 - Y_0 \mid s = 0, D_1 < D_0]P[D_1 < D_0 \mid s = 0]\}\end{array}}{P[D_1 > D_0 \mid s = 1] + P[D_1 < D_0 \mid s = 0]}.$$

This is a weighted average of treatment effects for two groups of men, low-scorers whose enlistment prospects were helped by misnorming, and high-scorers whose prospects were hurt.

II. Details of IV Estimation: The general estimating equation can be written as

$$\bar{y}_j = \beta_j + \alpha^* p_j + \bar{\epsilon}_j, \tag{A.8}$$

where $\bar{y}_j$ is the average earnings in the $j$th cell defined by $W$, $S$, and $A$; $p_j$ is the probability of military service in the $j$th cell; and $\beta_j$ is a restricted cell effect that may include a parametric function of $A$ but excludes interactions between $A$ and $S$. The data were aggregated to the level of cell $j$ using population frequency counts.

For the purposes of estimation, $\alpha^*$ was assumed to be constant, so that the efficient estimator of (A.8) is weighted least squares using the reciprocal of the variance of $\bar{\epsilon}_j$ as weights. In this case, the cell residual variance is also the variance of the dependent variable because the $p_j$ are known population proportions and $\beta_j$ is fixed within cells. The asymptotic covariance matrix of the resulting weighted least squares estimator can be derived directly or as the limiting distribution of a two-sample instrumental variables estimator in the special case where the moments from the first sample are $\bar{y}_j$ and the moments from the second sample ($p_j$) have zero sampling variance. See Angrist and Krueger (1992, Lemma 1) for details. A small Monte Carlo study resampling cell means and variances suggested that standard errors and confidence intervals based on the normal approximation provide reasonably accurate coverage of the relevant population parameter.

## APPENDIX B: Data

Sample Design: The 735,095 records submitted to SSA to be matched to earnings were selected as a stratified random sample. First, the population of 3,023,642 applicants meeting the criteria described in the text was grouped into 8,760 cells defined by race (white, nonwhite), year of application (1976–82), AFQT Group (I, II, IIIa, IIIb, IVa, IVb, IVc, V), veteran status, schooling completed at time of application (college graduate, some college, high school graduate, GED certified, grade 11, grade 9 or 10), year of birth (1954–1965) and another binary variable for 1977–78 applicants that was not used here. To satisfy SSA confidentiality requirements, population cells with less than 25 applicants were not sampled. There are 5,654 cells with 25 or more observations in this breakdown.

To ensure that the sampling scheme did not miss small cells and that no sampled cell had fewer than 15 observations, sampling was of three types. Random samples were drawn from large cells, but all observations in cells with between 25 and 34 observations in the population were selected. Medium-sized cells were over-sampled so as to generate an expected cell size of 35. The sampling ratio for large cells was .15 for whites, in cells with 200 or more observations, and .46 for nonwhites, in cells with 75 or more observations. This generated a sample of 352,035 white applicants and 401,060 nonwhite applicants.

The minimum sample cell size for white applicants is 17, and the minimum sample cell size for nonwhite applicants is 15. The median cell size for white applicants is 40 and the median sample cell size for nonwhite applicants is 57. The sample is self-weighting for those observations in population cells with 200 or more observations for whites and 75 or more observations for nonwhites. Roughly 93 percent of each racial group falls in cells in the self-weighting part of the sample.

SSA programmers succeeded in locating and validating earnings data for 697,944 out of the 753,095 DMDC applicants searched for on the SER. Matches were validated using SSA data on sex, race, and year or birth. A proposed match was assumed to be valid if the DMDC and SSA sex codes matched, and if the Social Security and DMDC years of birth differed by no more than 3 in absolute value. Note that the empirical sampling rate in each cell differs from the self-weighting rates of 15 and 46 percent because of random variation, and because of over-sampling of small cells and lost observations in the earnings match.

Inflating the matched sample observations by the inverse empirical sampling frequency shows that the 697,044 records that were matched and validated using SSA data represent 848,708 nonwhite applicants and 2,149,718 white applicants. Thus, (848.71/862.6 =) 98.4 percent of the nonwhite applicants in the population are represented in the earnings match, and (2149.7/2161.2 =) 99.5 percent of the white applicants are represented.

II. Social Security Earnings Coverage and Taxable Maximum:[31] OASDI and HI (Medicare) are contributory programs in which covered workers pay a tax and report earnings to the SSA via employers or the Internal Revenue Service. The SSA keeps track of taxable earnings on the SER, which is the source of the earnings data used here. Under the Federal Insurance Contributions Act (FICA) or the Self-employment Contributions Act (SECA), earnings are taxed for Social Security purposes up to the Social Security taxable maximum, and are therefore recorded on the SER only up to this maximum. The taxable maximum has been raised every year. By 1980, over 85 percent of covered male workers had earnings below the maximum. A separate and higher taxable maximum was instituted for HI in 1991. All taxable earnings, whether reported for the purposes of OASDI or HI, and whether reported as part of a mandatory or voluntary coverage provision, should appear on the SER.

About 95 percent of all jobs in the US were covered by the OASDI program as of 1991. Exceptions fall into five major categories: (i) Federal civilian employees hired before 1984, (ii) railroad workers, (iii) some employees of State and local governments already covered under a retirement system, (iv) household and farm workers with low earnings, and (v) persons with no wage and salary earnings and very low earnings from self-employment. Members of the uniformed services have been covered since 1956, and receive noncontributory wage credits to improve their insured status. Recent important changes include the 1983 coverage of most federal employees and employees of nonprofit organizations, HI coverage of many state and local employees in 1986, OASDI coverage of State and local employees without an employer retirement plan in 1990, and coverage of reserve soldiers in 1987.

III. SSA Confidentiality Edit: The confidentiality edit accepted as input the uncensored grouped data set showing cell-identifiers and earnings variables for the 5,654 cells defined above. Cells were masked to ensure that the file released satisfied SSA confidentiality requirements. Masking was implemented according to the following algorithm for each of the 18 years of earnings data in each of the 5,654 cells in the sample. Let

$EARN$ = average earnings,

$POSERN$ = average positive earnings,

$LOGERN$ = average log earnings,

$STD$ = standard deviation of earnings,

$STDPOS$ = standard deviation of positive earnings,

$STDLOG$ = standard deviation of log earnings,

$ZERO$ = number in cell with zero earnings,

$XMAX$ = number in cell with earnings exactly at the taxable maximum,

$GEMAX$ = number in cell with earnings at or above the taxable maximum,

$FREQ$ = sample cell count.

1. If $\{(FREQ - ZERO)*POSERN\} < 1001$ then mask:
   *EARN*, *POSERN*, *LOGERN*, *STD*, *STDPOS*, *STDLOG*, *ZERO*, *GEMAX*, *XMAX*.
2. If $(FREQ - ZERO) < 3$ then mask:
   *EARN*, *POSERN*, *LOGERN*, *STD*, *STDPOS*, *STDLOG*, *ZERO*.
3. If $ZERO < 3$ then mask:
   *ZERO*, *POSERN*, *STDPOS*.
4. If $\{(XMAX > 2)$ and $(GEMAX - XMAX) < 3)\}$ then mask:
   *GEMAX*.

[31] This section draws on US Department of Health and Human Services (1993).

5. If $GEMAX < 3$ then mask:
   *GEMAX*.
6. If $XMAX < 3$ then mask:
   *XMAX*.
7. If $STD = 0$ then mask:
   *EARN*, *STD*.
8. If $STDPOS = 0$ then mask:
   *POSERN*, *LOGERN*, *STDPOS*, *STDLOG*.

Rules 1 and 2 effectively eliminate all earnings information on cells with very low earners or few positive earners. Information on the number of individuals with zero earnings and the number with earnings at or above the taxable maximum is also masked in cells with few individuals in these categories (rules 3–6). Information on positive earnings is masked when there are few individuals with zero earnings (rule 3). In the unlikely event that everyone in a cell has the same earnings, earnings information in the cell is masked (rule 7–8). Finally, for the purposes of empirical work reported in this paper, the following modification was made to the editing process. When the *ZERO* variable in a cell is masked but the *EARN* variable for that cell is not masked (this situation occurs when there are substantial earnings in a cell but few zeros), I assume there are no individuals with zero earnings in the cell.

The impact of the confidentiality edit is primarily on earnings data for young men in the first few years of the sample period. For example, there are 1,099 cells with data on the earnings of men born 1962 or later. The confidentiality edit masks earnings data for 1974 in all but 51 of these cells. But only 91 cells in this group have 1978 earnings data masked, and only one cell is masked in 1980. After 1980, no cells are masked for this young group. Of the 1,459 cells for men born in 1957 or earlier, no cells have data on average earnings that were masked. Among the 3,096 cells with data on men born between 1958 and 1961, no earnings data are masked after 1976. Earnings for 1976 are masked in only 32 cells, and 1974 earnings are masked for 1,747 cells.

## REFERENCES

ANGRIST, J. D. (1990): "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence From Social Security Administrative Records," *American Economic Review*, 80, 313–335.

——— (1993a): "The Misnorming of the U.S. Military's Entrance Examination and its Effect on Minority Enlistments," University of Wisconsin, Institute for Research on Poverty, Discussion Paper No. 1017–93.

——— (1993b): "The Effect of Veterans Benefits on Education and Earnings," *Industrial and Labor Relations Review*, 46, 637–652.

——— (1995): "Using Social Security Data on Military Applicants to Estimate the Effect of Voluntary Military Service on Earnings," NBER Working Paper No. 5192.

ANGRIST, J. D., AND G. W. IMBENS (1995): "Two-Stage Least Squares Estimates of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 89, 431–442.

ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 90, 444–472.

ANGRIST, J. D., AND ALAN B. KRUEGER (1992): "The Effect of Age of School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples,"*Journal of the American Statistical Association*, 87, 328–336.

——— (1994): "Why Do World War Two Veterans Earn More Than Nonveterans?" *Journal of Labor Economics*, 12, 74–97.

ASHENFELTER, O. A. (1978): "Estimating the Effect of Training Programs on Earnings," *Review of Economics and Statistics*, 69, 47–57.

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

BARNOW, B., G. CAIN, AND A. GOLDBERGER (1981): "Selection on Observables," *Evaluation Studies Review Annual* 5. Beverly Hills: Sage, pp. 43–59.

BERGER, M. C., AND B. T. HIRSCH (1983): "The Civilian Earnings Experience of Vietnam-Era Veterans," *Journal of Human Resources*, 18, 455–479.

BERRYMAN, S. E., R. M. BELL, AND W. LISOWSKI (1983): *The Military Enlistment Process: What Happens and Can it be Improved?* Santa Monica: Rand Report R-2986-MRAL.

BEUSSE, W. E. (1974): "The Impact of Military Service on Low Aptitude Men," Air Force Human Resources Laboratory, Alexandria, VA.

BRYANT, R. B., V. A. SAMARANAYAKE, AND A. WILHITE (1993): "The Effect of Military Service on the Subsequent Civilian Wage of the Post-Vietnam Veteran," *Quarterly Review of Economics and Finance*, 74, 15–30.

BUREAU OF THE CENSUS (1993): *Census of Population and Housing, 1990: Public Use Microdata Sample A*. Washington, DC: Bureau of the Census (distributed by CIESIN, Ann Arbor).

BUSINESS WEEK (1992): "Where Troop Cuts Will Be Cruelest: For Blacks, the Services Have Been the Best Employer Around," June 8, 72–73.

CARD, D. E., AND DANIEL SULLIVAN (1988): "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment," *Econometrica*, 56, 497–530.

CHAY, K. Y. (1995): "Evaluating the Impact of the 1964 Civil Rights Act on the Economic Status of Black Men Using Censored Longitudinal Earnings Data," Mimeo, Industrial Relations Section, Princeton University.

COCHRAN, W. G. (1968): "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies," *Biometrics*, 24, 295–313.

COOKE, T. C., AND A. O. QUESTER (1992): "What Characterizes Successful Enlistees in the All-Volunteer Force: A Study of Male Recruits in the US Navy," *Social Science Quarterly*, 73, 238–252.

COOPER, R. V. L. (1977): *Military Manpower and the All-Volunteer Force*, Report R-1450-ARPA. Santa Monica: RAND Corporation.

CUTRIGHT, PHILLIPS (1974): "The Civilian Earnings of White and Black Draftees and Nonveterans," *American Sociological Review*, 39, 317–327.

DEFENSE DEPARTMENT (1988): *Population Representation in the Military Services*, Fiscal year 1987. Washington, D.C.: Office of the Assistant Secretary of Defense (Force Management and Personnel).

——— (1992): *Population Representation in the Military Services*, Fiscal year 1991. Washington, D.C.: Office of the Assistant Secretary of Defense (Force Management and Personnel).

DEHEJIA, R. H., AND S. WAHBA (1995): "Causal Effects in Non-experimental Studies: Re-evaluating the Evaluation of Training Programs," Mimeo, Harvard Economics Department.

DE TRAY, D. N. (1982): "Veteran Status as a Screening Device," *American Economic Review*, 72, 133–142.

EITELBERG, MARK J. (1988): "Test-Scoring Errors May Have Saved the All-Volunteer Force," *Navy Times*, 25, Sept. 12.

GILROY, C. L., R. L. PHILLIPS, AND J. D. BLAIR (1990): "The All-Volunteer Army: Fifteen Years Later," *Armed Forces and Society*, 16, 329–350.

GOLDBERGER, A. S. (1972): "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations," University of Wisconsin, Institute for Research on Poverty, Discussion Paper No. 123–72.

HECKMAN, J. J., AND R. ROBB (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis for Labor Market Data*, ed. by J. Heckman and B. Singer. New York: Cambridge University Press.

HEISEY, J. G., B. MEANS, AND J. H. LAURENCE (1985): "Military Performance of Low-Aptitude Recruits: Project 100,000 and the ASVAB Misnorming" (FR-PRD-85-2), Human Resources Research Organization, Alexandria, VA.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–476.

IMBENS, G. W., AND W. VAN DER KLAAUW (1995): "The Cost of Conscription in the Netherlands," *Journal of Business and Economic Statistics*, 13, 207–216.

IMBENS, G. W., J. B. LIEBMAN, AND N. EISSA (1997): "The Econometrics of Differences in Differences," Mimeo, Harvard Department of Economics.

LAURENCE, J. H. (1992): "Crew Cuts: Effects of the Defense Drawdown on Minorities," Chapter 2 in *Military Cutbacks and the Expanding Role of Education*. Washington D.C.: Office of Educational Research and Improvement, US Department of Education.

LAURENCE, J. H., P. F. RAMSBERGER, AND M. A. GRIBBEN (1989): "Effects of Military Experience on the Post-Service Lives of Low-Aptitude Recruits: Project 100,000 and the ASVAB Misnorming," FR-PRD-89-29, Human Resources Research Organization, Alexandria, VA.

MAGNUM, S. L., AND D. E. BALL (1989): "The Transferability of Military-Provided Occupational Training in the Post-Draft Era," *Industrial and Labor Relations Review*, 42, 230–245.

MAHONEY, R. (1991): "Is Military Job Training a Dead End?" *Sojourners*, 20, February–March, 230–245.

MAIER, M. H., AND W. H. SIMS (1986): *The ASVAB Score Scales: 1980 and World War II*. Alexandria, VA: Center for Naval Analysis, Report No. 116.

MAIER, M. H., AND A. R. TRUSS (1983): *Original Scaling of ASVAB Forms 5/6/7: What Went Wrong?* Alexandria, VA: Center for Naval Analysis, Research Contribution No. 457.

MOSKOS, CHARLES, AND J. S. BUTLER (1996): *All That We Can Be: Black Leadership and Racial Integration the Army Way*. New York: Basic Books.

ORVIS, B. R., AND M. T. GAHART (1990): *Enlistment Among Applicants for Military Service*, Report R-3359-FMP. Santa Monica: Rand Corporation.

PHILLIPS, R. L., P. J. ANDRISANI, T. N. DAYMOUNT, AND C. L. GILROY (1992): "The Economic Returns to Military Service: Race-Ethnic Differences," *Social Science Quarterly*, 73, 340–359.

ROSENBAUM, P. R., AND D. B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

——— (1984): "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association*, 79, 516–524.

RUBIN, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688–701.

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

——— (1978): "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics*, 6, 34–58.

SELTZER, C. C., AND S. JABLON (1974): "Effects of Selection on Mortality," *American Journal of Epidemiology*, 100, 367–372.

STAFFORD, F. P. (1991): "Partial Careers: Civilian Earnings and the Optimal Duration of an Army Career," in *Military Compensation Policies and Personnel Retention: Models and Evidence*, ed. by C. Gilroy, D. Horne, and D. A. Smith. Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences, Ch. 7.

USA TODAY (1990): "Does Military Experience Help?" 119, August, 10–11.

U.S. CONGRESS (1989): "Social Representation in the U.S. Military," Congressional Budget Office. Washington D.C.: U.S. Government Printing Office.

U.S. BUREAU OF THE CENSUS (1992): *Statistical Abstract of the United States 1992*. Washington D.C.: U.S. Govt. Printing Office.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES (1990): *Annual Statistical Supplement to the Social Security Bulletin*. Social Security Administration, Office of Research and Statistics, Washington D.C.