

NBER WORKING PAPER SERIES

WHY DO WORLD WAR II VETERANS EARN MORE THAN NONVETERANS?

Joshua Angrist

Alan B. Krueger

Working Paper No. 2991

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 1989

We thank Orley Ashenfelter for raising a question which inspired this research and for helpful advice. We also thank David Card and Whitney Newey for helpful advice, and Angela Chang for outstanding research assistance. Naturally, we are responsible for any remaining errors. The data and computer programs used in the preparation of this paper are available upon request from the authors. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors not those of the National Bureau of Economic Research.

NBER Working Paper #2991
May 1989

WHY DO WORLD WAR II VETERANS EARN MORE THAN NONVETERANS?

ABSTRACT

Veterans of World War II are widely believed to earn more than nonveterans of the same age. Theoretical justifications for the World War II veteran premium include the subsidization of education and training, and preference for veterans in hiring. In this paper, we propose and test an alternative view: that the observed World War II veteran premium reflects the fact that men with higher earnings potential were more likely to have been selected into the Armed Forces. An empirical strategy is developed that allows estimation of the effects of veteran status while controlling for correlation with unobserved earnings potential. The estimation is based on the fact that from 1942 to 1947 priority for conscription was determined in chronological order of birth. Information on individuals' dates of birth may therefore be used to construct instruments for veteran status. Empirical results from the 1960, 1970, and 1980 Censuses, along with two other micro data sets, support a conclusion that World War II veterans earn no more than comparable nonveterans, and may well earn less. These results suggest that OLS estimates of the World War II veteran premium are severely biased by nonrandom selection into military service, and that the civilian labor market experiences of veterans of World War II were not very different from the experiences of Vietnam-era veterans.

Joshua D. Angrist
Industrial Relations Section
Firestone Library
Princeton University
Princeton, NJ 08544

Alan B. Krueger
Industrial Relations Section
Firestone Library
Princeton University
Princeton, NJ 08544

Cross-sectional comparisons show that, on average, veterans of World War II earn more and have lower unemployment rates than nonveterans of the same age. In contrast, Vietnam-era veterans earn less and have experienced more unemployment than comparable nonveterans in their cohort.¹ Differences in the consequences of veteran status over time raise several questions about the operation of the labor market and the effect of military service. In this paper, we empirically explore alternative explanations for the apparently successful civilian labor market experiences of World War II veterans.

A quantitative understanding of the effect of military service on subsequent civilian earnings is of interest because of issues relating to military recruitment, income distribution, and public policy towards veterans. The conflicting estimates of the effect of military service on civilian earnings for veterans of different periods, however, clouds our understanding of this issue.

One hypothesis that has been advanced to explain the observed pattern of veteran wage premiums is that because the Vietnam war was an unpopular cause, men who served during that era were discriminated against when they returned to the civilian labor market. On the other hand, since World War II was a widely-supported war, it has been argued that veterans from that period were given preferential treatment in civilian employment. Positive feelings about World War II may also have translated into political support for relatively generous subsidies for education and preferential treatment

¹For example, Rosen and Taubman (1982) estimate a 10 percent loss of earnings to Vietnam-era veterans and a 17 percent gain to World War II veterans. Martindale and Poston (1979), De Tray (1982) and Berger and Hirsch (1983) find the same pattern of veteran premiums. Cohany (1987) finds higher relative unemployment rates for veterans of the Vietnam period than World War II.

in hiring.²

The explanations mentioned above presume that the observed correlation between veteran status and earnings is truly caused by military service. A plausible alternative explanation, however, is that observed differences in labor market consequences of veteran status are instead generated by differences in the selection process for service. There is evidence that many college-educated young men from wealthy families avoided military service during the Vietnam period, while low-income young men were unable to avoid service or chose not to avoid service (see Cooper 1977). In contrast, nearly 75% of men from eligible cohorts served in the Armed Forces during World War II, and those who failed to serve were often physically or otherwise unfit. These selection forces are likely to induce a positive bias in estimates of the veteran premium for World War II veterans, and a negative bias for Vietnam-era veterans.

Angrist (1989a,b) uses the Vietnam-era draft lottery to attempt to overcome selection problems in estimates of the effect of Vietnam-era military service on earnings. His main conclusion is that even after controlling for selection forces, white Vietnam-era veterans suffered substantial earnings losses when they returned to the civilian labor market.³ Additionally, he finds that the earnings differential associated with serving in the military during the Vietnam era appears to be accounted for by the fact that veterans have less civilian-job experience. This

²Public perceptions of the treatment accorded returning Vietnam era veterans are described in Veterans Administration (1980, p. 36). See Educational Testing Service (1973) for information on GI education benefits.

³Angrist's sample is too small to draw precise inferences about blacks, but it appears that compared to nonveterans, black veterans of the Vietnam war period fared better than white veterans.

finding naturally leads one to question why World War II veterans appear to have fared comparatively better than Vietnam-era veterans.

This paper re-examines the World War II veteran premium using econometric techniques to control for nonrandom selection into the military. The key to consistent estimation is the use of instrumental variables that are correlated with veteran status but uncorrelated with other determinants of earnings. Instruments are available because after 1942 men were drafted for military service in chronological order of birth. Variations in military demand during the course of the war generate a correlation between birthdays and the probability of service, with men born in the beginning of the calendar year having a higher probability of conscription than men born later in the year. Birthdays are probably uncorrelated with personal characteristics other than age, and can therefore be used to identify the effect of veterans status on earnings.

Our econometric strategy is to use dummy variables indicating quarter of birth as instruments for veteran status in an earnings equation. Because a large sample of World War II era men is necessary to precisely estimate the veteran's premium with this approach, most of the analysis is based on the 5/100 sample of the 1980 Census. In marked contrast to the Ordinary Least Squares (OLS) estimates, the Two Stage Least Squares (TSLS) estimates indicate that veterans of World War II actually earn 6 to 12 percent less than comparable nonveterans. Furthermore, we find generally corroborative results in four other micro-data sets -- the 1960 Census, the 1970 Census, the Survey of Income and Program Participation, and the 1973 Current Population Survey - Social Security Administration Exact Match. The robustness of these results is established using a number of formal

specification tests and by varying the basic specification. Our findings support the conclusion that positive estimates of the World War II veteran premium are due entirely to nonrandom selection into the Armed Forces.

The paper is organized as follows. Section I describes the institutional process used to register and conscript men into the military between 1942 and 1947. Section II describes our primary data set, presents a descriptive analysis, and documents the correlation between birthday and veteran status. In Section III, basic results from the 1980 Census are discussed and contrasted with results from ordinary least squares regressions. This section also presents results of Generalized Method of Moments (GMM) specification tests, and presents estimates from four additional data sets. Section IV explores the potential bias of using an instrument that is correlated with age. Section V extends the analysis by performing specification tests of models which use regional variations in inductions to identify the veteran premium. Finally, Section VI discusses some implications of the empirical results.

I. Background on Selective Service

In 1940 Congress approved the first peacetime conscription measure in the history of the United States.⁴ This bill provided the basic legal authority for conscription throughout the World War II era, which officially ended March 31, 1947. Of the 16,354,000 men who served in the Armed Forces during the World War II era, ten million were drafted.⁵

In the first step of the conscription process the Armed Forces made a request of the Selective Service for a certain number of men to be inducted by a specific date. The Selective Service National Headquarters would then assign a quota to each state to meet the military's demand. These quotas were divided into smaller units known as "calls," and allocated to each local draft board by the state Director of Selective Service. Finally, the local draft boards were responsible for notifying registered men to report for induction in the order assigned to them on the basis of officially determined "order numbers."

There were a total of seven national draft registrations between 1940 and 1947. Virtually all male U.S. citizens or resident aliens were required to register. In the First through Fourth Registrations, held from 1940-42, order numbers were assigned by lottery. As a consequence of these lotteries, wherein numbers were randomly drawn from glass fishbowls, order numbers were randomly assigned to draft registration card serial numbers. In the Fifth and Sixth Registrations, held from 1942 to 1947, order numbers

⁴Unless otherwise noted, the material in this section is drawn from Selective Service System (1946, 1947, 1948a, 1950, 1952, 1955 and 1986).

⁵See Statistical Abstract of the United States 1987, Table 550.

were assigned on the basis of birthdays.⁶

A chronology of key events associated with the Fifth and Sixth Registrations appears in Table 1. In the Fifth registration, men born on January 1, 1922 were assigned the order number immediately following the highest order number assigned in the Third Registration. (The Fourth Registration involved older men who were not liable for service). Fifth Registration order numbers advanced by one for each date of birth through June 30, 1924. In the Sixth Registration, Order Numbers were assigned consecutively to men born on July 1, 1924 through March 31, 1929. The order number sequence in the Sixth Registration began after the highest number assigned in the Fifth Registration. Note that voluntary enlistment in the military was prohibited in the second part of the Sixth Registration.

Local draft boards proceeded down their lists of registrants in chronological birth order to fill their calls. Once a draft board had successfully called enough men to satisfy its quota, remaining men on the registrant list were not at risk for conscription unless another call was issued to the draft board. This process generated a correlation between date of birth and the probability of military service. Moreover, there is some regional variation in the relationship between date of birth and military service because quotas were assigned on the state level. Our empirical analysis focuses on registrants whose order numbers were assigned in the second part of the Sixth Registration (born after January 1, 1925) because, as shown below, the correlation between date of birth and veteran

⁶The Seventh Registration was relatively small, involving only male citizens residing abroad.

Table 1

Registration and Induction in 1942-1947^a

Date	Event	Birth Cohorts Registered	Number Registered
June 30, 1942	Fifth Registration	Born Jan. 1, 1922 to June 30, 1924 (18-20 year olds)	2,942,796
Dec. 11, 1942 to Dec. 31, 1942	Sixth Registration (Part I)	Born July 1, 1924 to Dec. 31, 1924 (18 year olds)	521,738
Jan. 1, 1943	Voluntary enlistments prohibited for 18-38 year olds		
Jan. 1, 1943 to March 31, 1947	Sixth Registration (Part II)	Born Jan. 1, 1925 to March 31, 1929 (Registered on 18th birthday)	3,987,723
Sept. 7, 1945	Voluntary enlistments reestablished for men age 18-37		

Notes:

a. Dates are from Selective Service System (1946, p. 71-74), Selective Service System (1952, p. 26) and the chronology in Selective Service System (1986). Number registered in the fifth registration is from Selective Service System (1955, p. 28). Number registered in the Sixth registration is from Selective Service System (1946, p. 73) for Part I and Selective Service System (1955, p. 28) for total number in the Sixth registration. Number registered in Part II of the Sixth registration is calculated as the total minus Part I registrations.

status is strongest for these cohorts.

Induction and Deferment

Following registration and assignment of order numbers, registrants were given a physical examination and classified by their local boards (Selective Service System 1955, Figure 1). As of August 1943, registrants could be granted a deferment under one of the following categories: dependency hardship, morally unfit, necessary civil activity, war production, agricultural worker, government official, ministry of religion or divinity student, neutral alien, or 45 years of age or older.⁷ At the examination stage, registrants were classified by the local board as awaiting induction (I-A), conscientious objector (I-A-O), or physically unfit (IV-F). After February 1, 1943 registrants were allowed to express a preference for serving in a particular branch of service.

At least one historian has observed that local draft boards "applied widely divergent standards in deciding whom to draft and whom to defer."⁸ For this reason, and because of regional variations in demographic characteristics, there was considerable variation across states in the fraction of registrants who served in the military. For example, 78 percent of registrants age 18-25 in New Jersey and Rhode Island were either in the service or discharged as of February 1, 1945, while only 54 percent of registrants from South Carolina and South Dakota had served in the

⁷ The list of deferments for occupational and other reasons frequently changed during the war. For example, the deferment of all married men was eliminated on December 5, 1943, and reinstated just for married men with three or more children on November 15, 1945 (Selective Service System, 1948b, pp. 317-318).

⁸ See Klingaman, 1988, pp. 73. Reportedly, one of the New York draft boards drafted men whose marriages were on the rocks but deferred men who were happily married.

military by this time (see Selective Service System 1950, p. 158-159). Regional variation in the execution of the draft provides another potential source of information for identifying the effect of veteran status.

More than six million registrants were found unsuitable for service and therefore rejected by the Selective Service System. The majority of registrants who were exempted or rejected for service were determined physically or otherwise unfit by their local draft board. For example, 64% of the 18-29 year old registrants who were not inducted or available for service on February 1, 1945 were deemed unfit for service. Another 34% of deferments were for occupational reasons (e.g., war production), and the remaining 2% were deferred for other reasons (Selective Service System 1950, p. 322).

Terms and Conditions of Service

The minimum term of service in the Armed Forces for draftees was 18 months, with an additional requirement to serve in the reserves after the tour of duty expired.⁹ The requirement to serve in the reserves, however, was waived for draftees who agreed to stay in the regular army for an additional 24 months. The GI Bill of Rights, which was first passed in 1944, entitled veterans to subsidized education and training benefits.¹⁰

The majority of soldiers who served in World War II were not involved in actual combat. Stoufer, et al. (1949, p. 312) estimate that "out of

⁹From September 1940 to August 1941 the minimum term of service was 12 months (see Selective Service System, vol. 3B).

¹⁰Assuming attendance at a private institution, the World War II era GI Bill was more generous than the Vietnam-era bill, whereas assuming attendance at a public institution, the Vietnam-era GI Bill was slightly more generous than the World War II era bill. See Educational Testing Service (1973) for a comparative analysis of the generosity of educational benefits available to veterans of World War II and the Vietnam era.

every four men in the Army one had a combat job and of the remaining three, one had a clerical job, one had a skilled job, and one had a semiskilled or unskilled job." Since three-fourths of the jobs in the military were noncombat jobs, Stoufer, et al. maintain that "the Army was in a position to give many soldiers training which would be useful in civilian life after the war."

In spite of the relatively low fraction of soldiers involved in combat, the battle-related death rate for U.S. forces during World War II was 1.8 per 100 troops, while 4.1 soldiers per 100 suffered a nonfatal wound.¹¹ This high rate of injury suggests that service-related disabilities may be an important factor in the labor market consequences of World War II veteran status.

II. Data and Descriptive Analysis

Our main empirical analysis is based on the 5/100 public use sample of the 1980 Census (the A sample).¹² This data set contains information for a self-weighted sample of 11 million individuals in the U.S. as of April 1, 1980. We limit our initial sample to male citizens in 1980 who were born between 1919 and 1929 and resided in the 48 Continental States or District of Columbia. The earnings information in the survey pertains to total wage and salary in the preceding calendar year (i.e., 1979). Men who served in the military between September 1940 and July 1947 are classified as

¹¹These statistics are derived from Statistical Abstract of the United States 1987, Table 550.

¹²The 1980 Census is described in the technical documentation for Census of Population and Housing, 1980: Public-Use Microdata Samples (Washington, D.C.: U.S. Department of Commerce, 1983).

veterans of the World War II period.¹³ Any individual whose response was allocated to the sex, age, quarter-of-birth, veteran status, veteran period, salary, or weeks worked questions was eliminated from the sample. We further restrict the sample to individuals whose weekly wage in 1979 was between \$25 and \$5,000.¹⁴

After these restrictions, the sample contains 335,989 men. Descriptive statistics for this sample are provided in Table 2. On average, World War II veterans earn higher wages and have more education than nonveterans. There is only a trivial difference, however, in the probability that veterans and nonveterans report that they have a disability that restricts the kind or amount of work that they can perform.

Limiting the sample to men covered by Part II of the Sixth Registration (i.e., born between 1925 and 1928) provides a sample of 131,354 observations. These men were between age 51 and 55 at the time of the survey. We know of no other publicly available data set that provides as large a sample of men in the relevant birth cohorts.

OLS Estimates

Table 3 presents OLS estimates of the World War II veteran premium along with the proportion of men who served in World War II by year of birth, tabulated from the 1980 Census. In column 3 the veteran premium is the difference in mean log earnings between veterans and nonveterans, and in column 4 each veteran premium is estimated from a log earnings equation that partials out education, disability status, race, marital status, and 8

¹³The survey does not allow a distinction between enlistees and draftees, or among the various branches of service.

¹⁴None of our results is sensitive to this outlier restriction.

Table 2
Means and Standard Deviations^a

Variable	World War II Veteran	World War II Non-Veteran
Annual Salary	21,386 (12,915)	19,251 (12,714)
Weekly Wage	442.59 (281.68)	398.45 (272.17)
Years of Education	12.36 (3.21)	11.38 (3.97)
Attended College	.35 (.48)	.28 (.45)
White	.95 (.23)	.87 (.34)
Work Disability	.10 (.30)	.09 (.29)
Currently Married	.90 (.31)	.85 (.35)
Age	55.75 (2.75)	53.58 (3.28)
Weeks Worked in 1979	48.69 (8.75)	48.41 (8.89)
Labor Force Participation Rate ^b	.86 (.35)	.83 (.38)
Sample Size	215,272	120,717

Notes:

a. Tabulated from the 1980 Census. Sample includes men born between 1919 and 1929, inclusive. Except for the labor force participation rate, all variables are based on the sample of men with work experience in 1979. Standard deviations are shown in parentheses.

b. Sample size for labor force participation rate of veterans is 268,374, and 160,798 for nonveterans.

region dummy variables. The finding of a positive World War II veteran premium for most cohorts is consistent with the previous literature (e.g. De Tray, 1982).

It is immediately apparent that the veteran premium is not constant across birth cohorts. The veteran premium tends to be lower for cohorts which have a relatively low fraction of World War II veterans. If veteran status were randomly assigned, we would not expect the veteran premium to vary with the fraction in each cohort who served. In addition, the results indicate that the veteran premium is more than halved when covariates are held constant. Both of these findings suggest that at least part of the veteran premium is due to the correlation of veteran status with other characteristics that are related to earnings.¹⁵ In terms of measurable characteristics, veterans tend to possess relatively more attributes that are rewarded in the workplace (see Table 2). If the unobserved determinants of earnings are not completely captured by the measured variables and have a similar correlation with veteran status, the OLS estimate of the veteran premium will be biased.

Quarter of Birth and Veteran Status

The information on date of birth available in the 1980 Census is limited to respondents' quarter of birth. Table 4 shows how the probability of veteran status varies by quarter and year of birth for men who were at risk of being drafted in the Fifth and Sixth Registrations. For example, 71% of men born in the first quarter of 1927 are World War II veterans while only 58% of men born in the fourth quarter of 1927 are World

¹⁵ See DeTray (1982) for an alternative explanation of why the veteran premium varies with the fraction who served based on the screening hypothesis.

Table 3

The World War II Veteran Premium by Cohort in 1980^a

Birth Year (1)	Proportion of Male Workers Who Served in World War II (2)	Proportional Annual Earnings Differential		Sample Size (5)
		Without Covariates ^b (3)	With Covariates ^c (4)	
1919	.732	.175 (.011)	.066 (.011)	23,469
1920	.755	.191 (.011)	.088 (.010)	27,494
1921	.779	.198 (.010)	.084 (.010)	29,042
1922	.780	.238 (.010)	.097 (.010)	29,314
1923	.766	.267 (.010)	.108 (.009)	30,196
1924	.759	.287 (.009)	.124 (.009)	32,043
1925	.750	.276 (.009)	.104 (.009)	32,233
1926	.739	.266 (.009)	.109 (.008)	32,271
1927	.656	.243 (.008)	.094 (.008)	33,759
1928	.308	.087 (.008)	.032 (.008)	33,091
1929	.122	-.034 (.012)	-.011 (.011)	33,077

Notes:

a. Separate regressions were estimated for each year. Standard errors are shown in parentheses. The sample is drawn from the 5% sample of the 1980 Census, and excludes men living outside the Continental U.S.

b. Calculated as the difference in mean log earnings between World War II veterans and nonveterans.

c. Each differential is the coefficient on a World War II veteran status dummy variable from a log wage regression including the following controls: years of completed education, disability status dummy, white dummy, currently married dummy, 8 region dummies, and an intercept.

War II veterans. These data are depicted graphically in Figure 1.

The probability of serving in the military declines monotonically with time for men born after the second quarter of 1925. This decline reflects the fact that induction calls peaked in 1943 and decreased thereafter. For each birth cohort after 1925, those born in the first quarter of the year have a higher probability of serving in the military than those born in the fourth quarter of the year.

Table 4 also shows the marginal significance levels for F-tests of the joint significance of three quarter-of-birth dummy variables in linear probability models of veteran status.¹⁶ Two sets of significance levels are reported: the first is for an F-test with three degrees of freedom in a regression of World War II veterans status on an intercept and three quarter-of-birth dummies; the second is for an F-test with 24 degrees of freedom in a regression of World War II veteran status on an intercept, eight region dummies and 24 region-quarter-of-birth interactions. That is, the second significance level is a test of the marginal significance of quarter-of-birth dummies in an equation that allows the effect of quarter of birth to vary by region.

The probability values indicate a statistically significant relationship between quarter of birth and World War II veteran status for men born from 1925-1929. Men born during this period were at risk of being drafted in Part II of the Sixth Registration. Moreover, the second part of the Sixth Registration coincides with the period in which voluntary enlistments were prohibited. Consequently, the wave of voluntary

¹⁶We note that the error variance in the linear probability model is necessarily heteroscedastic. When we correct for heteroscedasticity, however, there is only a trivial difference in the marginal significance levels.

Table 4

Proportion of World War II Veterans by Quarter of Birth

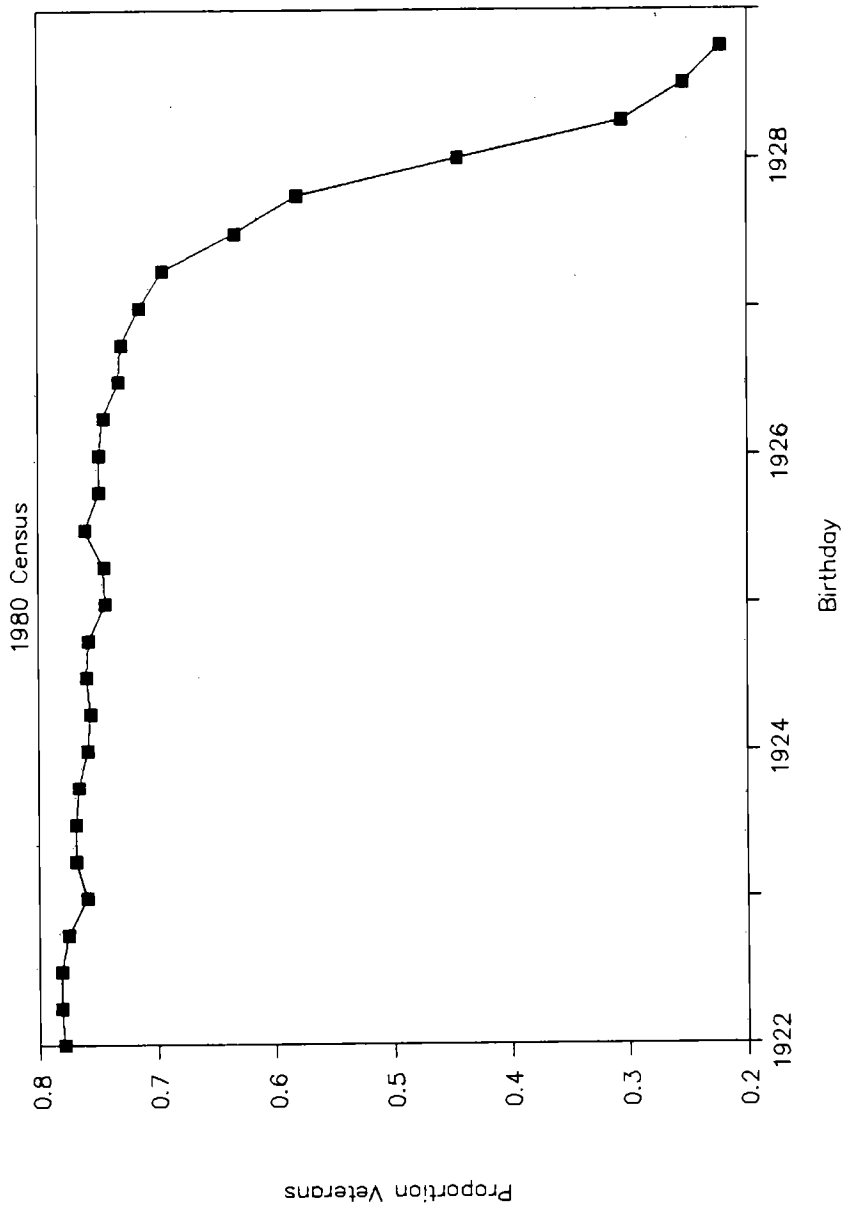
Year of Birth	Quarter of Birth				Marginal Significance Level of Quarter-of-Birth Dummies	
	I	II	III	IV	Quarters Alone ^a	Quarters*Region ^b
1922	.779	.782	.782	.776	.8033	.2149
1923	.760	.769	.769	.767	.5332	.4374
1924	.759	.757	.760	.758	.9827	.4223
1925	.744	.745	.761	.749	.0433	.0125
1926	.749	.745	.732	.730	.0131	.2208
1927	.714	.695	.633	.581	.0000	.0000
1928	.445	.306	.254	.223	.0000	.0000
1929	.170	.126	.109	.085	.0002	.0001

Notes:

- a. Significance level for F-test of joint significance of 3 quarter of birth dummies.
- b. Significance level for F-test of joint significance of 3 quarter of birth dummies and 3 quarter-of-birth dummies interacted with 8 region dummies, conditional on region main effects.

Figure 1

Proportion WWII Veterans by Qtr. of Birth



enlistments prior to January 1, 1943 may obscure any relationship between quarter of birth and veteran status for cohorts born before 1925.¹⁷

In the empirical analysis below, we limit our sample to men born between 1925 and 1928 because these cohorts were entirely encompassed by Part II of the Sixth Registration.¹⁸ For these cohorts, quarter of birth has been shown to satisfy one requirement for a valid instrumental variable: it is correlated with the endogenous variable of interest. The second requirement is that, conditional on age, quarter of birth may be excluded from a wage regression. The latter issue is taken up below.

III. Estimation and Results

We estimate human capital earnings functions of the form:

$$(1) \ln W = V\beta + X\xi + \sum_j Y_j r_j + \varepsilon \quad j = (1926, 1927, 1928)$$

where W is earnings, X is a row vector of explanatory variables, V is a veteran status dummy variable that equals one for veterans of World War II, β is (subject to the log approximation) the veteran premium, Y_j is a dummy variable indicating year of birth, ξ is a vector of parameters, and ε is an error term. OLS estimates of equation (1) will be biased if V and ε are correlated. In particular, it is likely that OLS estimates of β are biased

¹⁷Enlistments accounted for approximately 40% of World War II accessions from 1940-42 (Selective Service System 1948, Table 113).

¹⁸Individuals born in the last three quarters of 1929 were not covered by the Sixth Registration. In principle, the inclusion of men born in the first quarter of 1929 could be justified. In estimation of the earnings equations, however, a dummy variable is allowed for each year of birth. This considerably reduces the utility of information on a cohort of men from a single quarter.

upward because the Selective Service tended to reject registrants with a low earnings capacity (e.g., the disabled and low achievers on the Armed Forces Qualification Test).

The institutional description in the previous section suggests the following "first-stage" equation for the determination of veteran status:

$$(2) \quad V = Z\theta + \sum_j Y_j \alpha_j + X\pi + \mu \quad j = (1926, 1927, 1928)$$

where V is a veteran status indicator variable, $Z = (Q_1, Q_2, Q_3, Q_1 * Y_{1926}, \dots, Q_3 * Y_{1928})$ is a row vector of quarter-of-birth dummies and quarter-of-birth dummies interacted with year-of-birth dummies, Y_j is a year-of-birth dummy variable, X is the row vector of explanatory variables in equation (1), μ is an error term, and θ , α , and π are conformable vectors of parameters.¹⁹

Since earnings capacity is likely to be independent of the date an individual is born, quarter of birth is an appealing instrument for veteran status. For this approach to give a consistent estimate of β , however, Z must not appear in the earnings equation after conditioning on year of birth. This assumption is particularly likely to hold for middle-aged workers, who tend to have a flat age-earnings profile. Moreover, the exclusion restriction may be subjected to a variety of statistical tests.

Finally, we note that the effect of veteran status, β has been

¹⁹The fact that the endogenous regressor (V) is a dummy variable suggests that a probability model is in order for the first-stage equation. However, consistent estimation of the reduced form is not necessary for consistent estimation of the structural parameters (see Heckman, 1978 and Kelejian, 1971).

assumed constant in this model (e.g., veteran status has the same effect on everyone). An alternative specification would allow the effect to vary across individuals, denoted by the subscript i . For example, the random-coefficients model sets $\beta_i = \beta_0 + \bar{\beta}_i$. In this case, our TSLS procedure would identify $E(\beta_i \mid V_i = 1)$. That is, we estimate the veteran effect for those who served, free of selection bias (see Heckman and Robb, 1985).

Basic Results

Table 5 presents OLS and TSLS estimates of earnings equations with varying sets of covariates. In Columns 1 and 4 results are shown for equations that only include regressors for year of birth and veteran status. These results show a dramatic difference between the OLS and TSLS estimates -- the veteran premium is 22% with a t-ratio exceeding 50 in the OLS equation, and -13% with a t-ratio of -3.4 in TSLS estimation using quarter of birth as an instrument for veteran status.²⁰ The contrast between the OLS and TSLS estimates is also evident in Columns 2 and 5, which include additional regressors for race, marital status, 48 state dummy variables, and whether the individual works in an SMSA.

Columns 3 and 6 show the results of adding education and disability status to the list of regressors. These equations may be viewed as giving the "partial" effect of veteran status because veteran status may influence earnings through its effect on education (e.g., the GI Bill) and health. Comparing columns 2 and 3 shows that the addition of these variables nearly halves the OLS estimate of the veteran premium. The results in Column 6

²⁰ An approximate Hausman (1978) test of the difference between OLS and TSLS estimates can readily be calculated because the standard error of the OLS estimate is so small that the standard error of the Hausman test is essentially the standard error of the TSLS estimate. This test shows all three TSLS estimates to be significantly different from the OLS estimates.

Table 5
 Basic Results on the Return to World War II Service
 1980 Census, Men Born 1925-1928^a

Independent Variable	OLS			2SLS ^b		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	9.586 (.005)	9.008 (.014)	8.053 (.015)	9.846 (.030)	9.136 (.022)	8.086 (.034)
WW II Veteran (1 = yes)	.215 (.004)	.154 (.004)	.078 (.004)	-.132 (.039)	-.149 (.038)	-.064 (.034)
Born 1926	.016 (.006)	.018 (.005)	.011 (.005)	.012 (.006)	.015 (.005)	.009 (.005)
Born 1927	.039 (.005)	.036 (.005)	.018 (.005)	.007 (.007)	.008 (.006)	.005 (.006)
Born 1928	.124 (.006)	.101 (.006)	.050 (.005)	-.029 (.018)	-.032 (.017)	-.012 (.016)
White	--	.344 (.007)	.230 (.006)	--	.418 (.012)	.260 (.009)
Currently Married	--	.342 (.006)	.317 (.005)	--	.364 (.006)	.327 (.006)
48 State Dummies	No	Yes	Yes	No	Yes	Yes
SMSA	--	.145 (.007)	.112 (.007)	--	.159 (.007)	.116 (.007)
Years of Completed Education	--	--	.062 (.0005)	--	--	.065 (.0008)
Disability	--	--	-.420 (.006)	--	--	-.420 (.006)
R ²	.019	.094	.218	.0003	.083	.214
χ ² (DF = 11) [Prob-Value]	NA	NA	NA	15.4 [.164]	18.3 [.074]	22.3 [.022]

Notes:

a. Sample size is 131,354. Dependent variable is log annual earnings in 1979.

b. World War II veteran status is identified by the exclusion of quarter-of-birth dummies interacted with year-of-birth dummies.

indicate that the TSLS estimate of the veteran effect is still negative once education and health are held constant, although the effect is reduced in magnitude. Aside from the effect of veteran status on education, this attenuation in the TSLS estimates may also reflect the fact that education levels were increasing for succeeding generations of men in this era. This correlation between education and date of birth would lead individuals born in the beginning of the year -- who are more likely to be veterans -- to also have less education and therefore earn less.

The excluded instruments consist of three quarter-of-birth dummies interacted with four year dummies, so the TSLS models are over-identified by 11 degrees of freedom. The last row of Table 5 reports the GMM over-identification test statistic for tests of instrument-error orthogonality. All of the equations fail to reject the hypothesis of orthogonality at the 1% level, and only column 6 rejects at the 5% level.²¹ Given the large sample size, the failure to reject this null hypothesis in a classical hypothesis test is notable. These findings suggest that quarter of birth is a valid instrument for these equations.

Qualitatively similar estimates of the World War II veteran premium are found when we use log weekly earnings as the dependent variable instead of log annual earnings.²² These results are reported in Appendix Table A1.

Finally, we note that when we analyze separate samples of whites and nonwhites, the veteran premium (as estimated by TSLS) is more negative for

²¹This test is calculated as the sample size times the R^2 of a regression of the TSLS residuals from the second-stage equation on the included and excluded instruments.

²²Qualitatively similar results are also found when we narrow the sample to men who work in cities. These results are available on request.

nonwhites than whites. The difference in the veteran premium for blacks and whites, however, is well within the range of sampling error.

Other Data Sets and Time Periods

Table 6 presents additional OLS and TSLS estimates of the World War II veteran premium based on the 1960 Census (1% sample), the 1970 Census (3% sample), the first wave (1983) of the Survey of Income and Program Participation (SIPP), and the 1973 Current Population Survey - Social Security Exact Match (CPS-SSA). Our selection of alternative data sets was mainly dictated by the availability of information on quarter of birth for the cohorts at risk of conscription in World War II. To the extent possible, we have tried to construct samples and variables from these data sets that are comparable to our extract of the 1980 Census.²³ For convenience, we also report the results of the 1980 Census in Table 6.

Our motivation in examining additional data sets was to study how the World War II veteran premium evolves over the lifecycle, and to explore the robustness of the basic results in other samples. Unfortunately, since these data sets provide smaller samples than the 1980 Census, the veteran premium is less precisely estimated. Nonetheless, the results reveal some interesting patterns.

First, consider the estimates based on the three Censuses, which follow the same cohort of men over time. Interestingly, the OLS estimates show that the positive veteran premium declines as the cohort grows older. For example, when we only control for birth year, veterans earn 31% more than nonveterans in 1960, 26% more in 1970, and 22% more in 1980.

²³Details of the sample selection and variable definitions for each data set are given in Appendix A.

Table 6

Estimates of the World War II Veteran Premium in Other Data Sets and Time Periods^a

Data Set and Year	Wage Variable	Covariates	OLS	TSLS	χ^2 Over-Identification Statistic (DF=11) ^b
Panel A					
1960 Census 1% Sample (N=37,469)	Log Annual Earnings in 1959	3 Birth Year Dummies	.309 (.008)	-.001 (.080)	9.21 [.603]
		Above plus White; Married; 48 State Dummies; SMSA Dummy	.201 (.007)	-.078 (.075)	6.59 [.831]
		Above plus education	.115 (.007)	-.023 (.069)	4.75 [.943]
1970 Census 3% Samples (N=108,306)	Log Annual Earnings in 1969	3 Birth Year Dummies	.258 (.005)	-.024 (.045)	19.20 [.058]
		Above plus White; Married; 8 Region Dummies; SMSA Dummy	.184 (.005)	-.043 (.042)	15.78 [.149]
		Above plus education	.090 (.004)	.023 (.038)	15.83 [.148]
1980 Census 5% Sample (N=131,354)	Log Annual Earnings in 1979	3 Birth Year Dummies	.215 (.004)	-.132 (.039)	15.41 [.164]
		Above plus White; Married; 48 State Dummies; SMSA Dummy	.154 (.004)	-.149 (.038)	18.33 [.074]
		Above plus education and disability dummy	.078 (.004)	-.064 (.034)	22.31 [.022]

-- Continued --

Table 6 -- Continued

Data Set and Year	Wage Variable	Covariates	OLS	TSLS	χ^2 Over-Identification Statistic (DF=11) ^b
Panel B					
1973 CPS-SSA Exact Match (N=2,292)	Log Annual Earnings in 1972	3 Birth Year Dummies	.171 (.034)	-.412 (.346)	11.34 [.415]
		Above plus White; Married; 8 Region Dummies; SMSA Dummy	.143 (.033)	-.232 (.301)	10.75 [.464]
		Above plus education	.088 (.031)	-.092 (.270)	8.69 [.650]
1983 SIPP (N=512)	Log Monthly Earnings in 1983	3 Birth Year Dummies	.182 (.065)	-.094 (.340)	10.83 [.458]
		Above plus White; Married; 8 Region Dummies	.103 (.064)	-.260 (.331)	9.45 [.580]
		Above plus education and good health dummy	-.010 (.060)	-.355 (.310)	7.22 [.781]

Notes: a. Each sample contains men born in 1925-1928. The data sets are described in Appendix B.

b. Prob-value for GMM over-identification test given in brackets.

The table also shows that although the World War II veteran premium is positive and statistically significant when estimated by OLS, it typically becomes negative and statistically insignificant when estimated by TSLS, using quarter-of-birth dummies as instruments for veteran status. Given the large sampling variation in the TSLS estimates based on the CPS-SSA Exact Match and SIPP, however, it is difficult to draw meaningful conclusions about the magnitude of the World War II veteran premium in these data sets.²⁴

It is also interesting to compare the TSLS estimates for the cohorts over time in the Censuses. The TSLS estimates of the veteran premium using the 1980 Census are the most negative. One explanation for the more negative veteran premium in 1980 is based on age-related correlation between quarter of birth and earnings. Between 1960 and 1970, the cohorts we examine were between age 31 and 45, which would place them on the upward sloping portion of the cross-sectional age-earnings profile. In contrast, by 1980 this cohort of men reached their early fifties and would have been at the peak or downward sloping portion of the cross-sectional age-earning profile.²⁵ As a result, in 1960 and 1970 men born in the beginning of the year would be expected to earn more than men born at the end of the year, while in 1980 the exact opposite correlation would be expected. Since quarter of birth is always negatively correlated with veteran status, this

²⁴ The CPS Exact Match and SIPP identify individuals' month of birth. In auxiliary TSLS estimates, we compared 11 month-of-birth dummies to three quarter-of-birth dummies for these data sets and found that using month-of-birth dummies as instruments only slightly increased the precision of the veteran premium, and did not qualitatively change the estimated coefficients.

²⁵ For example, Mincer's (1974) estimate of the experience-earnings profile implies that earnings peak at age 52 for high school graduates, and age 56 for college graduates (see Table 5.1).

implies that the TSLS estimates will be biased upwards in 1960 and 1970, and biased downwards in 1980.²⁶ This consideration led us to explore the potential downward bias in the 1980 estimates. The results of this exploration are summarized in Section IV, below.

Finally, results from the three censuses have implications for the hypothesis that the effect of military service on civilian earnings is due to a loss of civilian labor market experience. Under the standard assumption of a concave experience-earnings relationship, the loss-of-experience hypothesis predicts that the veteran premium would decline with labor market experience. Angrist (1989b) and Crane and Wise (1987) tentatively conclude that the gap in earning between Vietnam-era veterans and nonveterans is consistent with this hypothesis. In contrast to the prediction of the loss-of-experience model, however, the TSLS point estimates based on the censuses show that the World War II veteran premium may have become more negative over time. The robustness of this result is explored below.

IV. Specification Analysis

If quarter of birth cannot be legitimately excluded from equation 1, then the TSLS estimates will be inconsistent. We have taken three alternative approaches to gauge the extent of omitted variable bias in our basic TSLS estimates. In the first and second approach, we assume that the relationship between earnings and quarters takes a known parametric form

²⁶For example, men born in the first quarter are older and more likely to be veterans than those born later in the year. If men born in the first quarter earn less even after controlling for veteran status, then use of a first-quarter dummy as an instrumental variable will bias estimates of the veteran premium downward.

and construct two TSLS estimators that are robust to misspecification. In a third approach, we use estimates of the effect of quarter of birth on earnings for a sample of men for whom veteran status is orthogonal to quarter of birth to control for bias due to an imperfect instrument.

Formally, let the correctly specified wage equation be:

$$(3) \ln W = V\beta + Z\gamma + \varepsilon$$

where the variables are defined as before and we have dropped the covariates for expositional purposes. The inclusion of the term $Z\gamma$ signifies that some function of the instruments, Z , is assumed to belong directly in the estimating equation. In addition, we continue to assume that $\text{cov}(V, \varepsilon) \neq 0$, which renders OLS estimates of (3) biased and inconsistent. It can easily be verified that if $\ln W = Vb + \nu$ is estimated by TSLS using the Z variables as instruments, the probability limit of the veteran premium b is given by:

$$(4) \text{plim } b = \beta + \text{plim } (V'P_Z V)^{-1} V' Z \gamma$$

where $P_Z = Z(Z'Z)^{-1}Z'$.²⁷ Note that depending on the magnitude of γ and the extent of correlation between ε and V , TSLS estimation with endogenous instruments may well yield an estimate of β with a lower mean squared error than OLS.

In principle, omitted variables misspecification of the type discussed above should be detected by the GMM over-identification test. The test

²⁷ If Z were a scalar variable, the asymptotic bias of the TSLS of β simplifies to γ/ρ , where ρ is the coefficient from an OLS regression of V on Z .

results presented so far, however, have shown no evidence of such misspecification. On the other hand, the GMM test is for an unrestricted alternative hypothesis where there is arbitrary correlation between the instruments and the regression error. Below, we attempt to develop more powerful tests and more robust estimates of β by assuming three specific functional forms for the omitted variable, $Z\gamma$.

A. Constructing an Instrument that is Orthogonal to Age

One might suspect that quarter of birth has a small effect on earnings because it is correlated with age. Since veteran status and quarter of birth are related in a nonlinear way, however, it is possible to construct instruments from quarter of birth that are uncorrelated with age, but correlated with veteran status. Specifically, we rely on the fact that a dummy variable indicating whether an individual is born in the middle two quarters of the year is orthogonal to linear age if births are uniformly distributed over quarters of the year.²⁸ Thus, models for the earnings of a single cohort are just identified using this indicator as an instrument; estimates for more than one cohort are over identified when tabulated using middle-quarter indicators for each cohort.

The first three rows of Table 7 present estimates of the World War II veteran premium using an indicator for the middle two quarters and interactions with the three year-of-birth dummies as instruments. Although the new instrument leads to an imprecise estimate of the veteran premium, these estimates tend to be zero or slightly negative. As would be expected if the sample under study were on the downward sloping segment of the age-

²⁸This approach was suggested to us by Orley Ashenfelter.

earnings profile, the orthogonal instrument yields estimates that are less negative than the previous TSLS results.

It is interesting to consider the relationship between this approach and the GMM test results reported above. One interpretation of the GMM test is as a Hausman (1978) test for equality of the different instrumental variables estimates that can be computed from an over-identifying set of instruments (Newey 1985). The power of the GMM test depends critically on the requirement that among the over-identifying set of instruments, at least one subset of instruments consistently estimates the structural coefficients. Since an indicator for the middle two quarters is a linear combination of instruments for individual quarters, this indicator is implicitly included in the instrument list that underlies our basic TSLS results. Thus, the GMM test has some power against models that are misspecified solely due to exclusion of a linear age term.

B. Controlling for Quadratic Age in Quarters

An alternative to the use of an instrument that is orthogonal to linear age is to assume that $Z\gamma$ has a more general functional form. For example, age-earnings profiles are often estimated with a quadratic specification. Assuming that the relationship between earnings and quarter of birth is quadratic leads to the following specification:

$$(4) \quad \ln W = V\beta + A\gamma_1 + A^2\gamma_2 + \sum Y_j\tau_j + \nu \quad j = (1926, 1927, 1928)$$

where A is age measured to the quarter of a year. The coefficient β can be consistently estimated by using quarter-of-birth dummies as instruments and including a quadratic age term along with the other regressors.

Table 7

Exploration of Robustness of Veteran Premium^a

Estimator	Covariates	Veteran _b Premium
1. Middle-of-the-year dummies interacted with year-of-birth dummies. Orthogonal with age by construction.	3 cohort dummies	-.048 (.134)
	Above plus white, married, 48 state dummies, and SMSA dummy.	-.017 (.126)
	Above plus education and disability status.	.012 (.114)
2. TSLS controlling for quadratic age effect measured in quarters.	3 cohort dummies, age and age squared.	-.125 (.113)
	Above plus white, married, 48 state dummies, and SMSA dummy.	-.099 (.105)
	Above plus education and disability status.	-.075 (.095)
3. Control Function Estimation	3 cohort dummies	-.072 (.047)
	Above plus white, married, 48 state dummies, and SMSA dummy.	-.083 (.046)
	Above plus education and disability status.	-.025 (.044)

Notes:

a. Data set is 1980 census. The dependent variable is log annual earnings in 1979. Estimators 1 and 2 are based on a sample of 131,354 men born between 1925 and 1928. Estimator 3 is based on a sample of 193,593 men born between 1923 and 1928.

b. Standard errors are shown in parentheses.

The results of TSLS estimation of equation (4) are shown in rows 4-6 of Table 7. The estimated veteran premium from this exercise is also imprecise, but the magnitude of the premium is remarkably similar to those estimated in our basic specification. It is also worth noting that the coefficients on the linear and squared age terms never exceed their standard errors.

C. Control Function Estimation

Our third set of specification tests is based on the fact that the correlation between quarter of birth and veteran status was found to be insignificant for men born between 1922 and 1924. We can therefore use these cohorts to estimate the direct effect of quarter of birth on earnings (γ), and use this additional information to correct the TSLS estimates.

The estimation technique is described in detail in Appendix B. Briefly put, to control for a possible relationship between earnings and quarters conditional on veteran status we have used the sample of men born in 1923 and 1924 to construct an estimate of γ , denoted $\hat{\gamma}$. We then form $\hat{Z}\hat{\gamma}$, and include this variable as a regressor in TSLS estimation of wage equations for the sample of men born 1925-1928.²⁹ In a different context, Heckman and Robb (1985) have referred to this approach as the use of a "control function".

Focusing on the sample of 62,329 men in the 1980 Census born between 1923-1924, we find that individuals born in the first quarter of the year earn a statistically significant 3% less than those born in the fourth

²⁹ To implement this procedure we actually estimate the equations for the two samples jointly, as described in Appendix B. The two-step procedure described in the text is asymptotically equivalent, but less efficient than the joint estimation.

quarter, controlling for birth cohort. However, we find no evidence of a correlation between quarter of birth and earnings for men born 1923-1924 in the 1960 or 1970 Censuses (i.e., at age 36-37 and age 46-47). These results suggest that within birth year, quarters of birth are correlated with earnings in 1980 because members of the sample are on the downward sloping portion of the age-earnings profile. On the other hand, quarter-of-birth dummies are likely to be valid instruments in 1960 and 1970 because in these years the age-earnings profile for the cohorts under study is relatively flat.

Results of control function estimation for the sample drawn from the 1980 Census are presented in the last three rows of Table 7. These estimates continue to show a negative, albeit attenuated, World War II veteran earnings premium.

Overall, the estimates in Table 7 suggest that using quarter of birth as an instrument for World War II veteran status in 1980 leads to somewhat exaggerated estimates of the earnings loss experienced by World War II veterans. The control function estimates of -3% to -8% may be closer to the true earnings loss, and this range is similar to the TSLS estimates based on the 1960 and 1970 Censuses. These estimates, however, are still dramatically different from the large, positive OLS estimates.

V. Alternative Exclusion Restrictions

The basic specification presented in Section II seems to provide fairly robust estimates of the effects of World War II veteran status. Having presented what we believe to be an accurate estimate, it seems natural to consider whether instruments more commonly available than

quarter of birth lead to similar inferences. Our inquiry is motivated by the work of Lalonde (1986), who has presented evidence that econometric techniques for program evaluation do a poor job of identifying the effects of manpower training programs. Following Heckman, Hotz and Dabos (1989), however, we consider over-identification test statistics along with the coefficient estimates derived from conventional instruments. This is in the hope that when conventional estimates are far off the mark, the over-identification test will reject the specification.

The conventional instruments we use to estimate the effect of World War II veteran status are a set of eight dummy variables indicating the region of the U.S. in which individuals were born, and year-of-birth dummies.³⁰ The results of this investigation are summarized in Table 8. Each row corresponds to a different model. All models include dummies for race, marital status, and SMSA status in both the wage equation and the instrument list. Additional covariates and instruments are listed for each model. If a variable is listed as a covariate but not as an instrument, it is excluded from the wage equation.

The first model is similar to our basic specification in Column 4 of Table 5. Model 2 is similar to model 1, but also includes current region dummies in both the equation and instrument list. As before, results from these models find a negative veteran premium with an acceptable over-identification test statistic.

In model 3, the instrument list is augmented by region-of-birth dummies and a full set of region-of-birth/year-of-birth/quarter-of-birth

³⁰We note that we additionally restrict the sample to men who reported their place of birth and were born in the U.S.

Table 8
 Estimates of the Veteran Premium Under Alternative Exclusion Restrictions^a

Model	Instruments and Covariates ^b								Veteran Premium	X ² Over-ID Stat (DF) [Prob-Value]	
	Quarter	X Year	Quarter of Birth	Year of Birth	Region of Birth	Current Region	Year X				
							Quarter	Region			
1. Covariates Instruments	I	I		C					-.127	19.1	(11)
				I					(.037)	[.0590]	
2. Covariates Instruments	I	I		C		C			-.135	20.9	(11)
				I		I			(.037)	[.0340]	
3. Covariates Instruments	I	I		C					.414	1454.5	(139)
				I		I			(.030)	[.0000]	
4. Covariates Instruments	I	I		C		C			.082	420.4	(139)
				I		I			(.031)	[.0000]	
5. Covariates Instruments	I	I		C		C			-.074	152.7	(139)
				I		I			(.034)	[.0940]	
6. Covariates Instruments	I	I		C		C			.980	93.7	(7)
				I		I			(.041)	[.0000]	
7. Covariates Instruments	I	I		C		C			-.059	9.6	(2)
				I		I			(.012)	[.0008]	
8. Covariates Instruments	I	I		C		C			-.042	287.5	(10)
				I		I			(.010)	[.0000]	

Notes:

a. Sample is 125,472. Standard errors for the estimated veteran premiums are shown in parentheses.

b. All equations and instrument lists include dummies for race, marital status, and SMSA residence. An I signifies that the variables were included with the instruments and a C signifies that the variables were included as covariates in the earnings equation.

interactions. Current region appears neither in the equation nor in the instrument list. The results are striking. The veteran status coefficient is .41 with a small standard error. The over-identification test statistic, however, exceeds 1,450. Compared to a chi-square distribution with 139 degrees-of-freedom, this provides strong evidence that the exclusion restrictions used to estimate model 3 are not valid.³¹ In this case, then, the specification test seems to have performed as one would hope -- a precise but apparently misleading coefficient estimate is associated with a strong rejection of the model specification.

A more plausible formulation than that in model 3 would also include current region in the wage equation. This idea is implemented in model 4. Although less dramatic, results from this model are qualitatively similar to those from model 3. A positive and statistically significant veteran premium of .082 is associated with an over-identification test statistic of 420. Again, we would confidently reject the exclusion restrictions underlying the model.

The model in row 5 is the same as model 4 except the wage equation includes region-of-birth dummies. Excluded instruments are region-of-birth/year-of-birth/quarter-of-birth interactions. In this case, it appears that inclusion of region-of-birth dummies directly in the second-stage equation removes the source of instrument-error nonorthogonality in model 4, and leads to a veteran status coefficient similar to those reported in Table 6.

³¹Degrees of freedom for this test are calculated as follows: there are 3 main quarter effects, 8 main region effects, 9 quarter-year interactions, 24 quarter-region interactions, 24 year-region interactions, and 72 quarter-year-region interactions, giving a total of 140 exclusions restrictions to identify one parameter.

The specification tests so far have been for models in which we are confident that at least a subset of the instruments yield consistent estimates. In the typical empirical investigation, however, the researcher does not know on a priori grounds that one set of instruments is valid. Consequently, it would be interesting to know if the over-identification test can provide meaningful evidence about specification errors when it is not clear that any of the instruments are valid. In models 6-8 we use three alternative exclusion restrictions to attempt to mimic the typical situation. Model 6 is identified by excluding region-of-birth dummies, model 7 is identified by excluding year-of-birth dummies, and model 8 is identified by excluding both region-of-birth and year-of-birth dummies from the wage equation. Importantly, none of these models uses information on quarter of birth to identify the effects of veteran status.

Models 7 and 8 yield plausible estimates of the veteran premium, while the estimate in row 6 (which is identified by only excluding region-of-birth dummies) is implausibly large. The table shows, however, that all three models are rejected by a specification test at the one-percent level.

The results of this brief exploration are mixed. Superficially plausible models -- using region-of-birth as instruments for veteran status in a wage equation that includes current region -- leads to incorrect inferences (rows 4 and 6). The over-identification test statistic, however, provides strong evidence of misspecification for models giving estimates that we believe are far off the mark. On the other hand, when the TSLS estimates are near what appears to be the true effect of veteran status, the over-identification test is not rejected when the interaction between region-of-birth dummies and quarter-of-birth dummies

are relied on for identification, but is rejected when year-of-birth dummies are the source of identification.

Finally, we note that when the same models were estimated on the sample drawn from the 1970 Census, qualitatively similar conclusions were reached.³²

³²A table containing these auxiliary results is available on request.

VI. Summary and Conclusion

This paper presents and estimates a variety of econometric models to measure the impact of serving in the Armed Forces during World War II on the subsequent civilian earnings of veterans. Given the nonrandom selection process used to draft men into the Armed Forces during World War II, it is unlikely that OLS estimates will give an unbiased estimate of the veteran premium. Therefore, we control for selectivity by exploiting the draft system that was used to call men to service between 1943 and 1947. Specifically, we construct instrumental-variables estimates of the veteran premium using one's quarter of birth as an excluded instrument.

Empirical results from the 1960, 1970, and 1980 Censuses support a conclusion that World War II veterans earn no more than comparable nonveterans, and may well earn less. Moreover, this finding seems to hold in two smaller data sets -- the CPS-SSA Exact Match and the SIPP -- and is found to be robust in a variety of specification tests. Estimates derived from the 1980 Census show the largest and most precisely estimated loss of earnings to veterans, on the order of -6% to -12%. Our TSLS estimate of the veteran premium is likely to be negatively biased in 1980, however, because by that time the cohort of men under study had reached the downward sloping portion of the age-earnings profile. Nonetheless, the TSLS estimate of the veteran premium is negative or close to zero even for age groups whose position on the age-earnings profile would induce a positive bias in the veteran premium. We interpret this as particularly strong evidence that World War II military service does not have a positive effect on civilian earnings.

Overall, a conservative estimate of the World War II veteran premium

consistent with the evidence from all three censuses would be roughly -5%. This estimate is in marked contrast to the OLS estimates, which indicate a positive World War II veteran's premium in the range of 10% to 20%. The wildly different OLS estimates indicate the potential hazards of estimating wage gaps when individuals are allocated to groups by a nonrandom selection process.

The answer to the question: "Why do World War II veterans earn more than nonveterans?", appears to be that World War II veterans would have earned more than nonveterans even had they not served in the military. Military service, in fact, may have reduced World War II veterans' earnings from what they otherwise would have been. This finding helps reconcile the apparently contradictory estimates of the effects of military service for veterans of World War II and veterans of the Vietnam era.

Table A1
 Dependent Variable: Log Weekly Wage
 1980 Census, Men Born 1925-1928^a

Independent Variable	OLS			2SLS ^b		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	5.747 (.004)	5.227 (.012)	4.549 (.013)	5.975 (.026)	5.34 (.019)	4.578 (.015)
WW II Veteran (1 = yes)	.193 (.004)	.140 (.004)	.074 (.003)	-.111 (.034)	-.127 (.032)	-.050 (.029)
Born 1926	.012 (.005)	.014 (.005)	.010 (.004)	.009 (.005)	.011 (.005)	.009 (.004)
Born 1927	.034 (.005)	.031 (.005)	.019 (.004)	.005 (.006)	.006 (.006)	.007 (.005)
Born 1928	.109 (.005)	.089 (.005)	.048 (.005)	-.025 (.016)	-.029 (.015)	-.007 (.013)
White	--	.302 (.006)	.201 (.006)	--	.367 (.010)	.226 (.008)
Currently Married	--	.250 (.005)	.236 (.005)	--	.269 (.006)	.244 (.005)
48 State Dummies	No	Yes	Yes	No	Yes	Yes
SMSA	--	.176 (.006)	.144 (.006)	--	.188 (.006)	.148 (.006)
Years of Completed Education	--	--	.056 (.0004)	--	--	.058 (.0007)
Disability	--	--	-.178 (.005)	--	--	-.178 (.005)
R ²	.021	.097	.205	.0003	.085	.201
X ² (DF = 11) [Prob-Value]	NA	NA	NA	24.4 [.011]	26.1 [.006]	30.3 [.001]

Notes:

a. Sample size is 131,354. Dependent variable is log weekly earnings in 1979.

b. World War II veteran status is identified by the exclusion of quarter-of-birth dummies interacted with year-of-birth dummies.

APPENDIX A

Data and Sample Selection1. 1970 Census of Population

The 1970 Census is described in Public Use Samples of Basic Records From the 1970 Census (Washington, DC: U.S. Department of Commerce, 1972). Our sample combines individuals from three separate public-use files -- the State, County Group, and Neighborhood files. Each file contains a self-weighted, mutually exclusive sample of one percent of the population, yielding a total sample of three percent of the population. The public-use data sets we used are based on the questionnaire that was administered to 15% of the population. These samples were selected because they are the only public-use samples of the 1970 Census that contain information on veteran status.

To the extent possible, the sample and variables were constructed to be comparable to the sample from the 1980 Census. We limited the sample to men born 1925-1928. Birth year was derived from reported age and quarter of birth. Individuals who had allocated responses to the age, sex, earnings, veteran status, or veteran period questions were excluded from the sample. Finally, individuals who were in the military as of the survey date were excluded from the sample.

The dependent variable, the log of annual earnings in 1969, is reported in intervals of \$100. This variable was converted to a continuous variable by taking the arithmetic average of the interval endpoints. We found that our estimates were not very different when the geometric average was used instead of the arithmetic average.

Unfortunately, geographic information is limited to 9 major Census

regions since the state of residence is not available in two of the three public-use files. The Neighborhood file directly identifies region of residence instead of state of residence. For the County Group file, regions were identified by the location of county-groups since this file does not indicate state or major region. In some instances, county groups straddled two states. In these cases, counties were allocated to the region that contained the greatest land-mass of the county group. For the States file, regions were derived from state of residence.

A dummy variable was created that equals one for men who served in the military during any part of the World War II period (defined as September 1940 - July 1947), and zero otherwise. The marital status dummy variable equals one if the individual is currently married, and zero otherwise. Years of education is the number of years of schooling completed. A dummy variable is included that equals one for men who indicate their race as white and zero otherwise. Three dummy variables were constructed indicating year of birth on the basis of individuals' current age and quarter of birth.

Unlike the 1980 Census, information is not available on disability status in the 1970 or 1960 Censuses. As a result, we do not control for disability status in the analyses of the 1960 and 1970 Censuses.

2. 1960 Census of Population

A description of the 1960 Census is provided in A Public Use Sample of Basic Records From the 1960 Census. (Washington, DC: U.S. Department of Commerce, 1975. Our sample is drawn from the one percent State sample.

To the extent possible, the sample and variables were constructed to be comparable to the sample from the 1980 Census. We limited the sample to

men born 1925-1928. Birth year was derived from reported age and quarter of birth. The 1960 Census has a smaller set of allocation flags than the later Censuses. Individuals who had allocated responses to the date of birth or veteran status questions were deleted from the sample. In addition, individuals residing in Hawaii and Alaska were deleted from the sample. Finally, individuals who were in the military as of the survey date were excluded from the sample.

The dependent variable, the log of annual earnings in 1959, is reported in intervals of \$100 for those earning less than \$10,000 and in intervals of \$1,000 for those earning more than \$10,000. This variable was converted to a continuous variable by taking the arithmetic average of the interval endpoints. We found that our estimates were not very different when we used the geometric average instead of the arithmetic average.

A dummy variable was created that equals one for men who served in the military during any part of the World War II period (defined as September 1940 - July 1947), and zero otherwise. The marital status dummy variable equals one if the individual is currently married, and zero otherwise. Years of education is the number of years of schooling completed. A dummy variable is included that equals one for men who indicate their race as white and zero otherwise. Three year of birth dummies were constructed on the basis of current age and quarter of birth. Eight region of birth dummies were derived from the state of birth question.

3. 1973 CPS-SSA Exact Match

This data set is described in Studies from Interagency Data Linkages, Report No. 9, US Department of Health, Education and Welfare, 1979.

The CPS-SSA Exact Match data set consists of matched records from the March 1973 Current Population Survey and administrative Social Security records. The two data sources were matched on the basis of respondents' Social Security numbers. Some variables in our analysis are obtained from the SSA administrative data, and others are taken from the CPS.

We first describe the variables derived from the CPS. The dependent variable, the log of annual earnings in 1972, was derived from the CPS question. A World War II veteran status dummy variable was created that equals one for men who served in the military during any part of the World War II period (defined as September 1940 - July 1947), and zero otherwise. The marital status dummy variable equals one if the individual is currently married, and zero otherwise. Years of education is the number of years of schooling completed. Finally, eight region dummies and a dummy variable for residence in an SMSA were created from the CPS fields.

The variables taken from the SSA portion of the data set were month of birth (which was used to create quarter-of-birth dummies), birth year (which was used to create cohort dummies), and race.

The sample was narrowed to individuals born 1925-1928, and who were categorized as males in both the CPS and SSA portions of the data set.

4. The Survey of Income and Program Participation (SIPP)

This data set is described in Survey of Income and Program Participation User's Guide, 2nd Edition - Draft, U.S. Department of Commerce, Bureau of the Census, Washington D.C., 1984. The extract we use is drawn from the SIPP 1984 Panel, Wave I Rectangular File. The data refer to the interview month and the four preceding months. Reference month 4 is

the month preceding the interview month and is one of chronological months September 1983 - December 1983. The SIPP variable mnemonics used in the User's Guide are given in parentheses in the description below.

The sample was selected using the following rules:

- i.) Birth year between 1925-1928 (U_BRTHYR)
- ii.) Wage and salary earnings in reference month 4 not imputed (WS1CAL01)
- iii.) Not a resident of Hawaii in reference month 4 (H4_STATE \neq 15)
- iv.) Had wage and salary earnings in reference month 4 (WS1_2032 > 0)
- v.) Weekly wage between \$25 and \$5,000, where the weekly wage is the monthly wage and salary earnings divided by weeks worked in reference month 4 (WS1_WKS4).

The dependent variable is the log of wage and salary earnings in reference month 4 (WS1_2032). Instruments consist of quarter-of-birth dummies derived from month of birth (U_BRTHMN). The tape location for month of birth does not appear in the SIPP codebook, but month of birth is on the Wave I rectangular file public-use tape in columns 5343-5344.

The following regressors were used in the analysis:

- i.) Health affects work dummy (SC1386 = 1)
- ii.) World War II veteran dummy, defined as a veteran (VETSTAT = 1) who served in the World War II era (U_SRDTE = 3)
- iii.) White dummy (RACE = 1)
- iv.) Married in interview month dummy, defined as married, spouse absent or present (MS_5 = 1 or 2)
- v.) Education, defined as highest grade (HIGRADE) completed
- vi.) 8 region dummies based on state of residence in reference month 4 (H4_STATE).

APPENDIX B

A Control Function Estimator for Instrumental Variables

The key to identification in instrumental variables estimation is knowledge of the relationship between the excluded instruments and the error term in the equation to be estimated. In TSLS it is assumed that the instrument has a zero coefficient in the second-stage equation. In other cases, however, it may be that a known function of the instruments belongs in the second-stage equation, but that conditional on inclusion of this function consistent instrumental variables estimation is feasible. Heckman and Robb (1985) have referred to variants of this approach as "control function estimators."

In the current context, control function estimation is motivated by the assumption that quarter of birth appears in the earnings equation with a functional form that can be estimated. In particular, we assume that for cohorts which evidenced no correlation between quarters of birth and earnings, the regression of earnings on quarter of birth can be consistently estimated. This idea is made more precise by writing the basic regression model for two samples, indexed by 1 (men born 1923-1924) and 2 (men born 1925-1928):

$$(A1) \quad \begin{aligned} \ln W_1 &= Z_1 \gamma + (V_1 \beta + \epsilon_1) \\ \ln W_2 &= V_2 \beta + Z_2 \gamma + \epsilon_2 \end{aligned}$$

In the first sample, OLS regression of $\ln W_1$ on Z_1 is assumed to provide a consistent estimate of γ because Z_1 and V_1 are orthogonal. In

addition, V_1 and Z_1 are assumed to be independent so that the variance of $\ln W_1$ conditional on Z_1 is homoscedastic.

Our strategy is to estimate the above system by stacking observations for men from the two subsamples.³³ To ensure that system estimation will result in consistent estimates of parameter standard errors, allowance should be made for the fact that the residual variance in the two subsamples will differ even if the variance of ϵ_1 and ϵ_2 are assumed to be equal. Estimates of the residual variances of the two error terms are obtained by first estimating equations for the two samples separately. That is, the following equation for sample 2 was estimated by TSLS using the estimated $\hat{\gamma}$ from the first sample:

$$(A2) \quad y_2 = V_2\beta + [Z_2\hat{\gamma}] + (Z_2(\gamma - \hat{\gamma}) + \epsilon_2),$$

The sample variance of the residual from the first equation gives a consistent estimate of the variance of ϵ_1 . The sample variance of the residual from the second equation gives a consistent estimate of the variance of ϵ_2 because $\hat{\gamma}$ is a consistent estimate of γ .³⁴

Applying the two step procedure in the 1980 census, the standard deviation of ϵ_1 was found to be 2% higher than the standard deviation of ϵ_2 in the model without covariates, 6% higher in the model with states,

³³When estimating models with covariates other than year-of-birth dummies, we omitted the additional covariates from the first equation. This is appropriate assuming that the additional covariates are also orthogonal to quarter-of-birth dummies.

³⁴Note that coefficient estimates from this two-step procedure are also consistent. The two-step estimator is similar to Heckman's (1979) two-step estimator for sample selection models.

married, white and SMSA, and 15% higher in the model with these covariates plus education and disability. The results reported in the text were obtained by using these estimates of the residual variance ratios to weight the first equation in estimation of the two equation system.

References

- Angrist, Joshua. "Using the Draft Lottery to Measure the Effects of Military Service on Civilian Earnings," forthcoming in Research in Labor Economics Volume X, edited by Ronald Ehrenberg. Greenwich: JAI Press, Inc., 1989a.
- Angrist, Joshua. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," Industrial Relations Section, Princeton University; Mimeo, January 1989b.
- Berger M.C. and B.T. Hirsch (1983). "The Civilian Earnings Experience of Vietnam-Era Veterans." The Journal of Human Resources 18 (4) (Fall 1983): 455-479
- Browning, Harley, Sally Lopreato and Dudley Preston, Jr. "Income and Veteran Status: Variations Among Mexican Americans, Blacks and Anglos." American Sociological Review 38 (February 1973): 74-85.
- Cohany, S.R. "Labor Force Status of Vietnam Era Veterans." Monthly Labor Review 110 (2), (1987): pp. 11-17.
- Cooper, V.L. Military Manpower and the All-Volunteer Force. RAND Corporation, Report R-1450-ARPA, Santa Monica, 1977.
- Crane, J.R. and David Wise. "Military Service and the Civilian Earnings of Youths." In Public Sector Payrolls, edited by D.A. Wise. Chicago: University of Chicago Press, 1987, pp. 119-146.
- De Tray D.N. "Veteran Status as a Screening Device." American Economic Review 72 (March 1982): 133-42.
- Educational Testing Service. "Educational Assistance to Veterans: A Comparative Study of Three GI Bills," James L. Bowman, Project Director, Princeton: Educational Testing Service, 1973.
- Hausman, Jerry A. "Specification Tests in Econometrics." Econometrica 46 (1978): 1251-1271.
- Heckman, James J. "Dummy Endogenous Variables in a Simultaneous Equation System." Econometrica 46 (July 1978): 931-959.
- Heckman, James J. and Richard Robb. "Alternative Methods for Evaluating the Impact of Interventions." In Longitudinal Analysis of Labor Market Data, edited by J. Heckman and B. Singer. New York: Cambridge University Press, 1985.
- Heckman, James J., Joseph Hotz, and M. Dabos. "Do We Need Experimental Data to Evaluate the Impact of Manpower Training on Earnings?" Forthcoming in Evaluation Review, 1989.
- Kelejian, H.H. "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables." Journal of the American Statistical Association 66 (June 1971): 373-374.

Klingaman, William. 1941. New York: Harper & Row, Publishers, 1988.

Lalonde, Robert. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." American Economic Review 76 (September 1987): 604-620.

Martindale, Melanie, and Dudley Poston, Jr. "Variations in Veteran/Nonveteran Earnings Patterns Among World War II, Korea, and Vietnam War Cohorts." Armed Forces and Society 5, No. 2 (February, 1979): 219-243.

Mincer, Jacob. Schooling, Experience and Earnings. New York: National Bureau of Economic Research, 1974.

Newey, Whitney. "Generalized Method of Moments Specification Testing." Journal of Econometrics 29 (1985): 229-256.

Rosen S. and P. Taubman. "Changes in Life Cycle Earnings: What Do Social Security Data Show?" Journal of Human Resources 17 (3) (Summer 1982): 321-38.

Selective Service System. Registration and Selective Service. Special Monograph Number 4. Washington DC: U.S. Government Printing Office, 1946.

Selective Service System. Selective Service Registrations July 31, 1945 to March 31, 1947, Second Edition. Washington DC: U.S. Government Printing Office, 1947.

Selective Service System. Quotas, Calls and Inductions. Special Monograph Number 12, Volume I. Washington DC: U.S. Government Printing Office, 1948a.

Selective Service System. Selective Service and Victory. Washington DC: U.S. Government Printing Office, 1948b.

Selective Service System. The Classification Process. Special Monograph Number 5, Volume III. Washington DC: U.S. Government Printing Office, 1950.

Selective Service System. Problems of Selective Service. Special Monograph Number 16, Volume I. Washington DC: U.S. Government Printing Office, 1952.

Selective Service System. The Operation of Selective Service. Special Monograph Number 17, Volume I. Washington DC: U.S. Government Printing Office, 1955.

Selective Service System. A Short History of the Selective Service System. Office of Public Affairs, Selective Service System National Headquarters. Washington, DC: U.S. Government Printing Office, 1986.

Stoufer, Samuel, et al. Studies in Social Psychology in World War II. vol. 1. Princeton, NJ: Princeton University Press, 1949.

Taussig, M.K. "The Rationale for Veterans Benefits." In Those Who Served: Report of the 20th Century Fund Task Force on Policies Toward Veterans. New York: The 20th Century Fund, 1974.

U.S. Department of Commerce. A Public Use Sample of Basic Records From the 1960 Census. Washington, DC: U.S. Department of Commerce, 1975.

U.S. Department of Commerce. Census of Population and Housing, 1980: Public-Use Microdata Samples. Washington, DC: U.S. Department of Commerce, 1983.

U.S. Department of Commerce. Public-Use Samples of Basic Records From the 1970 Census. Washington, DC: U.S. Department of Commerce, 1972.

U.S. Department of Commerce. Survey of Income and Program Participation: Wave I Rectangular File. Washington, DC: Bureau of the Census, 1985.

US Department of Health, Education and Welfare. Studies from Interagency Data Linkages, Report No. 9. Washington, DC: Social Security Administration, 1979.

Veterans Administration. Myths and Realities: A Study of Attitudes Toward Vietnam Era Veterans. Submitted by the Veteran's Administration to the Senate Committee on Veterans' Affairs, July 1980.