Erik Johnston and Tori Farris

Professor Joseph Kuehn

ECON 440: Advanced Econometrics

6 December 2024

Exploring Effect of Military Service on Earnings Using Propensity Score Matching

Part 1: Introduction

Many young adults in the United States pursue military service as a viable option for employment upon entry to the workforce. For some, enlistment occurs upon completion of some secondary schooling, like high school, while for others enlistment occurs after completion of a four year degree. In some cases, active duty military members attend school while serving as they are encouraged to do so through military programs that offer tuition assistance in exchange for a commitment to serve for a certain number of years.

The American military provides educational and financial incentives in addition to job security. For this reason, we believe there likely exists a relationship between military engagement, or veteran status, and income. This paper investigates the effect of military service on earnings. To answer this research question, we applied the method of propensity score matching to data sourced from the IPUMS dataset. Through this analysis, we found that veteran status has a positive effect on earnings, with varying magnitudes depending on the propensity matching method.

Part 2: Data Summary

In order to perform this analysis, we extracted a total of seven variables from the IPUMS USA data pool. In selecting these data we searched for variables that would exogenously explain not only veteran status, but our response variable, personal income. Further, the explanatory variables we selected had to uniquely explain variation in personal income and propensity for veteran status, meaning they could not be affected by other explanatory variables or the error term in the model. This criteria ensures accurate propensity scores, and ultimately conclusions.

We included multiple years (2016, 2021, 2022, and 2023). Years 2017-2020 contained data that were approximations due to issues IPUMS suffered with data collection. We selected sex (Male = 1, Female = 0) as an explanatory variable for the reason that men compose a significantly larger portion of the veteran pool in the United States. According to the U.S. Department of Labor, women make up only 10% of the veteran population. This suggests that men are more likely to enlist in the military. We also selected age (years) for the reason that physical fitness is a component of eligibility. Further, age may implicitly arise as a factor that the U.S. considers in its recruitment strategies for the reason that younger people are more likely to have a longer career in the military. To illustrate, the average age of recruits in the 2024 military fiscal year was 22 years and 4 months. We seek to capture this age effect by including it in the model. We included marital status (Married = 1, Single = 0) for the reason that decisions about how to provide for a family are likely to play a role in one's decision to enlist. A married person may be attracted by the prospect of job security, alternatively they may be dissuaded by the possibility of being relocated around the country or globe. Further, we included race (1 = White or Asian-American, 0 = All other recognized racial categories) for the reason that the U.S. military has a history of implementing targeted recruitment strategies on the basis of race.

Asian-Americans were collapsed into the same value as Whites under the recoded race variable for the reason that Asian-Americans on average make more than all other racial groups IPUMS collects data on. Coding these groups together reflects socioeconomic benefits that are shared by both groups. Finally, we included family total income (in dollars) to represent one's socioeconomic status in the model. Socioeconomic status has historically impacted military enlistment due to the military's perceived advantages, such as job stability and educational benefits.

Our response variable, total personal income is measured in dollars. In cleaning this data, we restricted the range of possible values between \$0 and \$9,999,997, in order to account for how IPUMS represents missing and unknown entries. The propensity variable veteran status is also represented as a binary variable (1 = veteran/active duty, 0 = non-veteran/civilian). We would like to note that throughout this paper we will refer to veteran status. According to how we recoded the variable, veteran status effectively includes active-duty service members in addition to veterans.

In Figure 2.1, one can see that while the sample represents non-veterans approximately equally across the age variable, there is a distinct skewness to the age of veterans. In the plot one can see veteran status is left-skewed, with some bunching of age in ages 65-75. This is likely due to the United States' involvement in the Korean War (1950-1953) and Vietnam War (1955-1975). This may impact our data in the way that individuals in the 65-75 age range are approaching retirement, or already are retired. This could potentially reduce the amount of income reported, impacting our results.

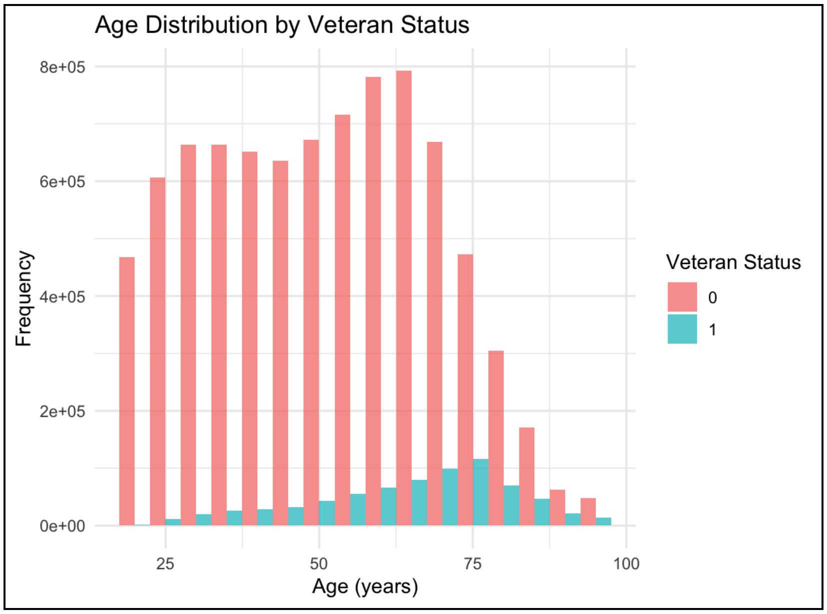


Figure 2.1

In Figure 2.2 one can observe the proportion of veteran status between racial categories as they have been specified in our data. We previously mentioned that the United States in the past has made targeted efforts to recruit on the basis of race. Specifically, the U.S. Armed Forces have made efforts to recruit members of racial minority groups. We see here that the proportions of veteran status between racial categories are relatively similar in value (8.79% in 1, 5.86% in 0). This indicates that in the our sample, the aforementioned recruitment tactics may not be having as strong of an effect as one could anticipate.

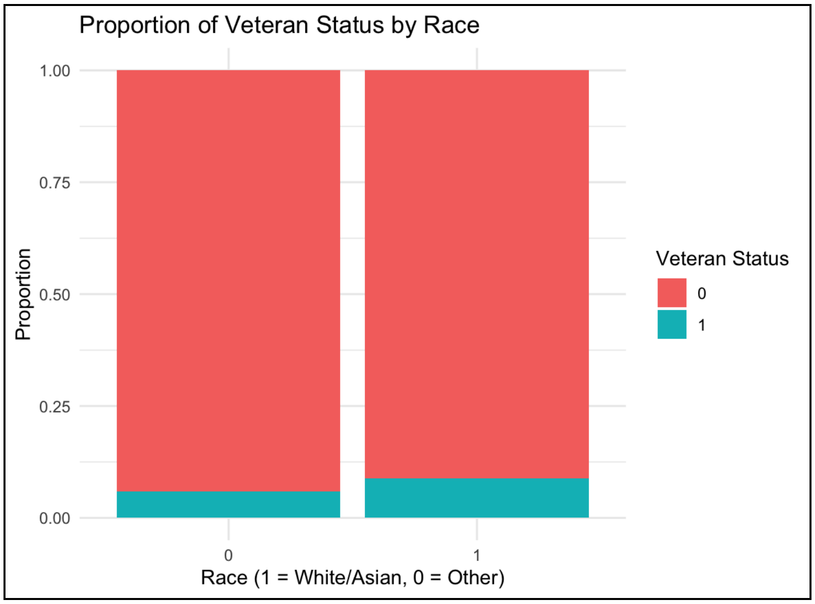


Figure 2.2

Finally, before going through formal modelling steps to determine the impact of veteran status on income, we look at a simple comparison of mean income between veterans and non-veterans.

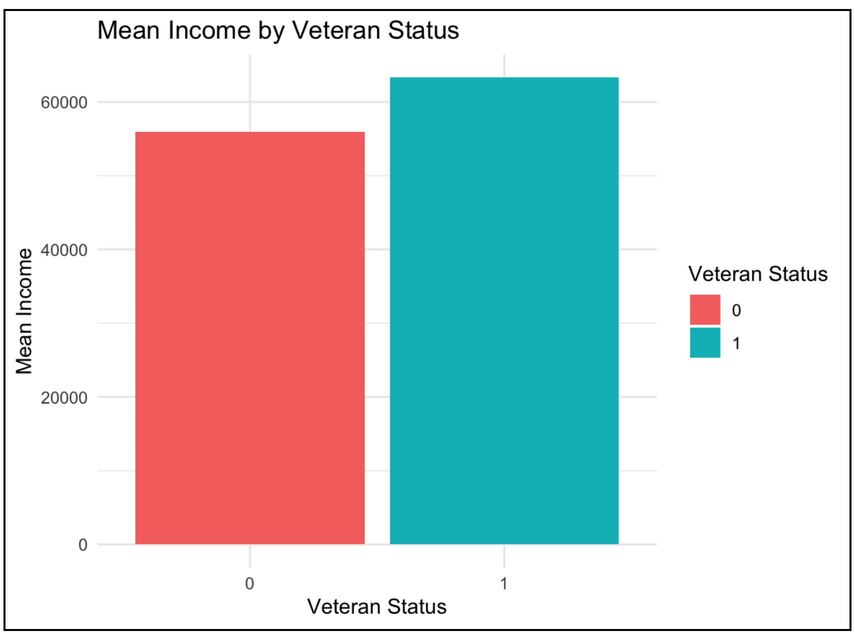


Figure 2.3

Without controlling for possible confounding variables, we observe that veterans on average make more money ($63,292) compared to non-veterans ($55,920). This result lends itself to our belief that there exists a relationship between veterans status and earnings, however we cannot draw conclusions at this point.

Part 3: Model

One big issue with a basic linear regression in this case is selection bias, because the choice to join military service is not random. Their decision could be influenced by factors such as socioeconomic background, education level, and family military history. These same factors could also correlate with long term earnings, which complicates finding causality based on military service.

Propensity Score Matching addresses this problem by comparing individuals who served in the military with those who didn't, but who are otherwise similar based on other observable factors. By matching individuals with similar "propensity scores" (the likelihood of joining the military given these characteristics), the method creates comparable groups, allowing for a fair analysis of how military service affects long-term earnings.

Using a logistic regression, we predicted the likelihood of an individual joining the military (the treatment group) based on observable characteristics such as marriage status, family income, race, age, and sex. These predicted probabilities are the propensity scores.

We then plotted histograms of the propensity scores for each group (veterans and non-veterans).

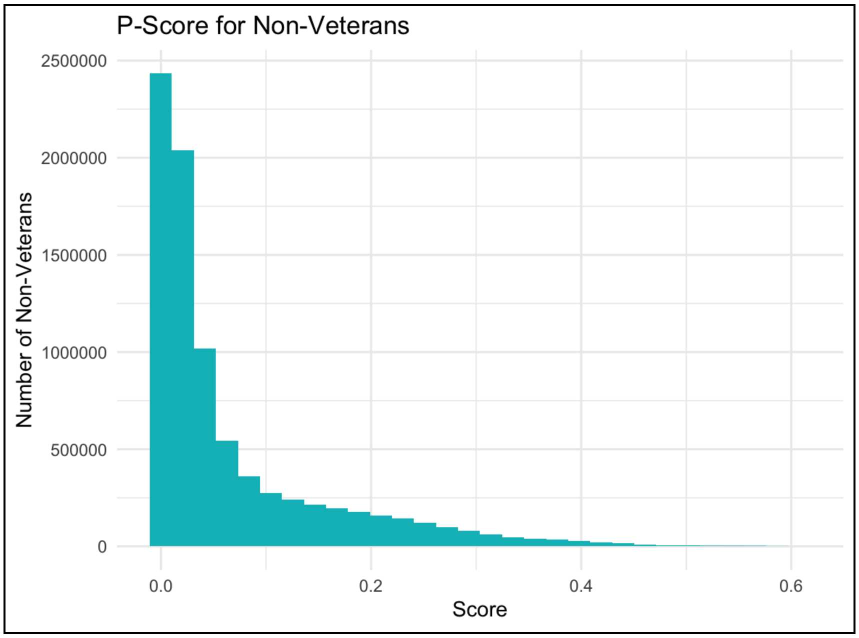


Figure 3.1

The non-veterans propensity scores skew to the right, showing that a large portion of the non-veterans are predicted to have low probabilities to be veterans. This makes sense, as these individuals chose not to go into military service.

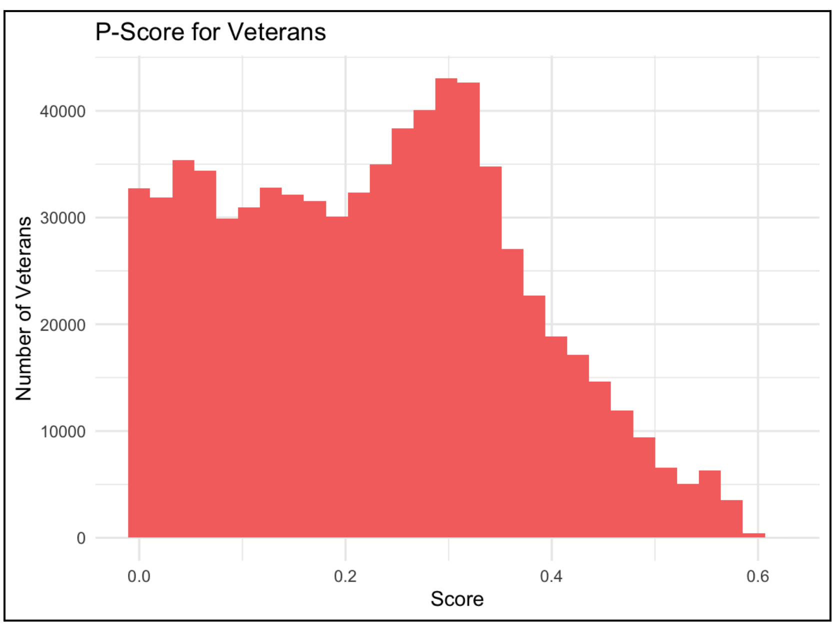


Figure 3. 2

For those who are veterans, their propensity scores (probability that they will be a veteran, based on our model) are more or less uniform from 0 to 0.2, increase to a peak around 0.3, then taper off. Veterans have a broader range of predicted probabilities of choosing to join the military. This wider spread suggests more diversity in the covariates for veterans compared to non-veterans.

We then checked for overlap between veterans and non-veterans graphically to ensure that for every veteran in the treatment group, there are non-veterans in the control group with similar propensity scores. This is the common support condition, and is necessary to make meaningful comparisons between the groups.

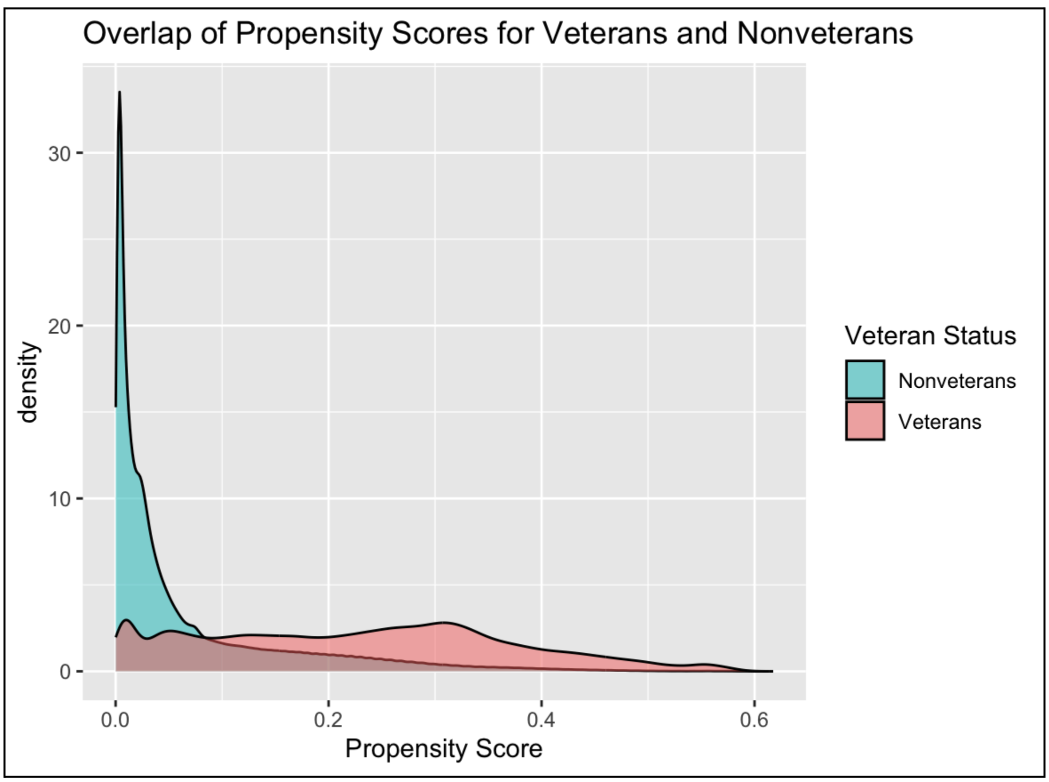


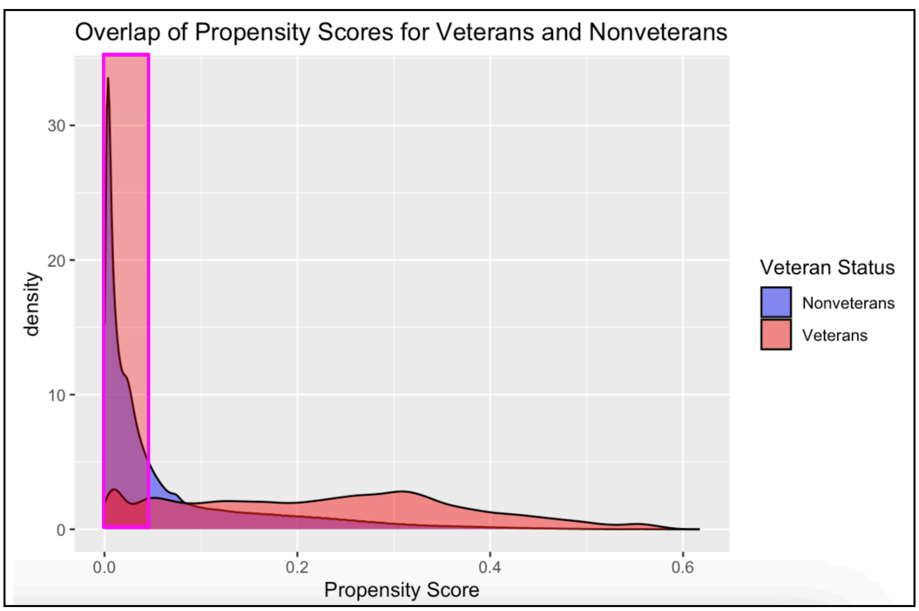
Figure 3.3

For comparison purposes, we calculated three average treatment effects:

1. ATE from OLS regression
2. ATE using propensity score matching and inverse propensity score weighting
3. ATE using propensity score matching and trimming

The OLS regression gives us a baseline in which to compare our other estimates. The weighting method helps account for differences in the distribution of propensity scores between veterans (treatment group) and non-veterans (control group). Because the control group is overrepresented on the low end of the propensity scores, weighting adjusts for this imbalance, ensuring that each observation contributes appropriately to the analysis.

Because there is a high volume of extremely low propensity scores for non-veterans, we trimmed the data to a range of 0.05 to 0.95 in order to focus analysis on the overlap between the two groups.



We trimmed the data between the thresholds of 0.05 and 0.95 (not illustrated)

Figure 3. 4

However, each model has its own assumptions and limitations. OLS Regression assumes all relevant variables are included, but it cannot account for selection bias, leading to potentially biased results when comparing veterans to non-veterans. Propensity Score Matching (PSM) helps address selection bias by comparing individuals with similar characteristics, but it assumes all confounding factors are observable and included. In this case, skewed propensity scores can still introduce bias. A Weighted Model improves reliability by balancing the data, but they rely on the same assumption of no unmeasured confounding and can amplify noise if the propensity score model is misspecified. A Trimmed Model further reduces bias by removing extreme cases, but this limits the generalizability of results to the broader population and depends on subjective trimming thresholds. Across all models, unmeasured confounders and potential differences in the treatment effect across subgroups remain key challenges.

Part 4: Results

Summary of Model Results: Effect of Military Service on Income

Model	Average Treatment Effect	Standard Error	Significance Level	Observations
OLS	-417.2	66.33	*** (p-value: 3.19e-10)	9,110,012
Propensity Score Matching	7371.478	N/A*	N/A*	9,110,012
Propensity Score - Weighted	4675.732	N/A*	N/A*	9,110,012
Propensity Score - Trimmed	1788.898	N/A*	N/A*	9,110,012

**For propensity score models, observing statistical significance and calculating standard errors are accomplished through bootstrapping, however due to the large number of observations in our data (over 9 million), our computers were unable to perform this technique.*

Figure 4. 1

When initially examining the effect of Veteran Status on income using an Ordinary Least Squares (OLS) regression model, the coefficient for Veteran Status suggests that military service is associated with a decrease in income of $417.20 on average, compared to non-veterans. This result is statistically significant at the 0.001 level, meaning there is strong evidence to suggest that veteran status has a negative impact on income, according to the observed data. However, this conclusion may be misleading because the OLS model does not account for selection bias — that is, individuals who choose to join the military may differ in important ways from those who do not, which could affect their income. Without adjusting for this bias, the OLS estimate may not accurately represent the true effect of military service.

To address this issue, we turn to propensity score matching (PSM), which aims to adjust for selection bias by matching veterans and non-veterans with similar characteristics (i.e., propensity scores) but different treatment statuses. After applying propensity score matching, the Simple Difference in Means shows that being a veteran actually has a positive effect on income, with veterans earning on average $7,371.49 more than non-veterans.

However, this result is likely overestimated due to a significant skew in the propensity scores of non-veterans. Because of this skew, the matching process gives more weight to certain non-veterans who are less similar to veterans, leading to a biased upward estimate of the treatment effect. Essentially, the matching process does not fully correct for the imbalances in the data, which causes the positive effect to appear larger than it likely is in reality.

The weighted model ATE of $4,675.73 provides an even more accurate estimate of the effect of military service on income. The weighting process ensures that non-veterans who are more similar to veterans in terms of their characteristics (as reflected in their propensity scores) receive more weight, while those who are less similar receive less. This adjustment results in a more balanced and reliable comparison, with the estimate still positive, but less extreme than the simple difference in means.

To refine the estimate further and address potential outliers or extreme propensity scores, we trimmed the model, removing the most extreme cases that could disproportionately influence the results. This trimmed ATE of $1,788.90 indicates that after addressing both selection bias and extreme observations, military service is still associated with an increase in income, but the effect is more modest than the untrimmed, weighted estimate. To answer our original research question of how military service impacts earnings, we contest that the trimmed model best represents the real effect, concluding that military service increases income by $1,788.90, while acknowledging the limitations of our analysis due to the inability to bootstrap for statistical significance.

Part 5: Conclusion

Our propensity score analysis demonstrates a positive effect of military service on earnings with a magnitude of $1,788.90. When applying the technique of propensity score matching, our findings suggest that military service in fact has a positive effect on income. The OLS model spuriously reports a negative impact of military service likely due to unaccounted selection biases, while the propensity score matching models, particularly the trimmed model, provide a more nuanced understanding of the true relationship. We can conclude that by choosing to join the military, assuming all confounding factors are properly accounted for, individuals may experience a modest increase in income compared to their non-veteran counterparts. However, this conclusion is contingent on the accuracy of the matching process and the assumption that all relevant variables influencing both military service and income have been included in the analysis.

Future research could further refine these estimates by exploring additional variables that might influence both military service and income, such as educational attainment, post-service occupation, and geographic location. Additionally, future studies could utilize longitudinal data to better capture the long-term effects of military service on earnings.

References

Segal, David R., et al. "The All-Volunteer Force in the 1970s." *Social Science Quarterly*, vol. 79, no. 2, 1998, pp. 390–411. *JSTOR*, http://www.jstor.org/stable/42863796. Accessed 5 Dec. 2024.

"2021 Employment Situation of Women Veterans." *Veterans' Employment and Training Service Demographics*, U.S. Department of Labor, 29 Apr. 2021, www.dol.gov/agencies/vets/womenveterans/womenveterans-demographics#:~:text=%E2%80%9CThere%20are%20nearly%202%20million,only%201.5%25%20are%20veterans.%E2%80%9D.

"Fiscal Year 2024-2025 Recruiting Media Roundtable with Service Leaders." *U.S. Department of Defense*, www.defense.gov/News/Transcripts/Transcript/Article/3952493/fiscal-year-2024-2025-recruiting-media-roundtable-with-service-leaders/. Accessed 5 Dec. 2024.