The Hidden Pre-Trend: How a Third Census Decade Exposes Identification Failure in WWII Service-Return Estimates

APEP Autonomous Research* @ai1scl

March 10, 2026

Abstract

Can a third decade of census data reveal identification failures invisible to standard two-wave designs? Using the first individual-level three-decade panel (1930–1940–1950) linking 13.4 million men, we apply the state $\times$ cohort reduced-form design commonly used to estimate WWII service returns—mobilization intensity driven by agricultural deferments interacted with draft-eligible cohorts—to the pre-war decade as a falsification test. The test fails decisively: mobilization exposure already predicted differential occupational trajectories before the war began (+1.50 points without controls, −0.72 with controls, both $p < 0.01$). This pre-trend, invisible without the 1930 census wave, contaminates conventional 1940–1950 estimates. Controlling for pre-war observables reverses the sign of the post-treatment estimate to −0.26 points. These findings cast serious doubt on positive WWII service-return estimates from two-decade designs and demonstrate that linked census panels require pre-treatment validation extending beyond the immediate pre-period.

JEL Codes: J24, N32, J45, I26

Keywords: identification failure, pre-trends, WWII, military service, occupational mobility, census linking

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

The "Greatest Generation" returned from World War II and prospered. They leveraged the most generous educational benefit in American history—the GI Bill—entered a labor market reshaped by wartime production, and climbed the occupational ladder faster than any generation before. But was the war the cause of their success? Or were the sixteen million men who served already the ones most likely to succeed? This question sits at the intersection of labor economics, economic history, and the evaluation of government programs, and it remains unsettled after decades of research.

The difficulty is identification. Unlike the Vietnam draft lottery that Angrist (1990) exploited so productively, WWII offered no randomized instrument. Every credible design must contend with selection: healthier, more motivated, less agriculturally tied men were more likely to serve (Stouffer et al., 1949; Angrist and Krueger, 1994). The standard approach—comparing veterans to non-veterans conditional on observables—requires that selection operates solely through measured characteristics, an assumption that grows less plausible the more one knows about the Selective Service System's deferment policies (Hershey, 1945).

This paper introduces a new diagnostic strategy for a classic empirical design. We construct the first individual-level three-decade census panel, linking the universe of draft-age men across the 1930, 1940, and 1950 full-count U.S. Censuses using the IPUMS Machine Learning Panel (MLP) crosswalk (Ruggles et al., 2024; Price et al., 2021). The resulting dataset contains 13.4 million men born between 1895 and 1922, observed at three points spanning 20 years of their working lives—9.1 million in the primary analysis cohorts (draft-eligible and older control) and 4.2 million in an age placebo group. With this panel, we subject the standard state $\times$ cohort reduced-form design—cross-state variation in WWII mobilization intensity driven by the Tydings Amendment's agricultural deferments, interacted with draft-eligible birth cohorts—to a falsification test that two-wave designs cannot perform.

Our contribution is primarily methodological. The 1930 census observation provides something no two-decade design can offer: a genuine pre-treatment baseline for evaluating whether the identifying interaction is correlated with pre-existing occupational dynamics. We apply the same regression structure used to estimate WWII effects on 1940–1950 outcomes to the pre-war decade (1930–1940 outcomes) as a falsification exercise, with appropriately lagged controls (1930 occupational score replaces 1940 occupation and education, which are post-treatment for the pre-war outcome). Standard designs, which begin their observation window in 1940, cannot perform this test—they lack variation before treatment. We note that the pre-war decade captures a different outcome margin (labor-force entry for teenagers) than the post-war decade (early-career progression for young adults), so the test is best understood as a falsification of the identifying interaction rather than a textbook parallel-trends test.

The falsification fails decisively. Without individual controls, mobilization exposure interacted with draft eligibility predicts a 0.50-point *increase* in occupational income scores from 1940 to 1950 ($p < 0.01$), seemingly confirming positive returns to WWII service. Adding 1940 pre-war controls for education, occupation, race, marital status, farm residence, and nativity flips the sign: the estimate becomes -0.26 points ($p < 0.01$). This sign reversal suggests that selection on observables accounts for more than 100% of the positive raw association.

But the decisive diagnostic comes from the 1930 pre-baseline. The same interaction—mobilization exposure $\times$ draft eligibility—predicts a -0.72-point change in occupational scores during 1930–1940 in the specification with individual controls ($p < 0.01$), a decade when draft-eligible cohorts were teenagers and young adults with no exposure to military service. Without controls, the pre-trend is *positive* (+1.50, $p < 0.01$), paralleling the positive raw post-treatment effect and reinforcing that selection drives both. The identifying interaction is correlated with state-specific cohort occupational dynamics even before the war: men in high-mobilization states were already on systematically different occupational trajectories before Pearl Harbor. This likely reflects differential Great Depression effects and recovery dynamics across states with varying industrial composition—a confound that two-decade designs cannot detect.

An age placebo test—estimating the same mobilization effect for men born 1895–1904, who were too old for active service—yields a coefficient of -0.05 (SE = 0.11), statistically indistinguishable from zero, though the confidence interval is wide enough to accommodate nontrivial effects. The pre-trend is specific to the cohort–instrument interaction, not to mobilization exposure per se. As an exploratory exercise, a trend-adjusted specification that nets out the 1930–1940 change yields an estimate of -0.91 points ($p < 0.01$), though this requires a linearity assumption that we cannot fully justify and uses a different control set than the single-decade regressions. Leave-one-out analysis across 49 states shows the main result is stable (range: -0.27 to -0.23), and no single state drives the finding.

This paper contributes to several literatures. First, we provide a methodological lesson for the growing literature using linked historical census data (Abramitzky et al., 2012, 2021; Bailey et al., 2020; Long and Ferrie, 2013; Ward, 2023; Feigenbaum, 2016). Pre-trend tests have become standard in difference-in-differences designs (Roth, 2022; Rambachan and Roth, 2023), but they require pre-treatment data. When researchers link only two census waves bracketing a treatment, they cannot test whether the identifying variation is correlated with pre-existing dynamics—they can only assume it is not. Our finding that a strong pre-trend contaminates the WWII mobilization design demonstrates that extending the observation window backward is not merely a nice robustness check but a necessity for credible identification.

Second, we cast doubt on positive estimates of WWII service returns in the long-running literature (Angrist, 1990, 1998; Angrist and Krueger, 1994; Xie, 1992; Costa, 2023). While Angrist (1990) established negative returns to Vietnam-era service using the draft lottery, WWII has been widely presumed to differ because of the GI Bill's unprecedented generosity (Bound and Turner, 2002; Turner and Bound, 2003; Mettler, 2005; Fetter, 2013). Our evidence suggests that the identifying variation commonly used to estimate positive WWII returns is contaminated by pre-existing differential trends, though we stress that our reduced-form design cannot isolate the causal effect of military service per se from other state-specific cohort dynamics correlated with mobilization exposure.

Third, we speak to Collins and Zimran (2025), the closest predecessor, who use a 1940–1950 linked census sample and a selection-on-observables design to estimate positive WWII service returns. Our three-decade panel reveals an identification threat their design cannot detect: the mobilization interaction is correlated with pre-existing occupational trajectories. Their positive estimates are consistent with the pattern we document, where positive raw returns reflect pre-trend contamination. This is not a criticism of their methods but a demonstration that two decades of data are inherently limited for this question.

Fourth, we build on Acemoglu et al. (2004), who use state-level mobilization rates—driven by agricultural deferment patterns identical to our instrument—to study female labor supply. Our individual-level panel and three-decade structure complement their aggregate analysis by showing that the mobilization instrument, while powerful for generating cross-state variation, correlates with pre-existing occupational dynamics that two-decade designs cannot detect.

The remainder of the paper proceeds as follows. Section 2 describes the institutional setting of WWII mobilization and the Tydings Amendment deferment system. Section 3 presents the data, including the MLP three-decade panel and the construction of our mobilization instrument. Section 4 develops the empirical strategy. Section 5 presents the main results, including the falsification test and exploratory trend-adjusted estimates. Section 6 examines heterogeneity and mechanisms. Section 7 reports robustness checks. Section 8 discusses what the design identifies, limitations, and implications for the broader census-linking literature.

2. Institutional Background

2.1 The Selective Service System and WWII Mobilization

The United States entered World War II with a conscription apparatus already in place. The Selective Training and Service Act of 1940, signed 16 months before Pearl Harbor, established the first peacetime draft in American history (Selective Service System, 1942). Over the course of the war, approximately 10 million men were inducted through Selective Service, joining roughly 6 million volunteers, for a total of 16.1 million who served in the armed forces (Clodfelter, 2002).

The system operated through local draft boards—6,443 of them by 1942—staffed by civilian volunteers who classified registrants into availability categories. Classification proceeded through a sequence of deferments: dependency (Class III-A), occupational (Class II), physical fitness (Class IV-F), and age. The practical effect was that military service rates varied dramatically across states, driven not by patriotism or voluntarism but by the economic structure of local labor markets.

2.2 The Tydings Amendment and Agricultural Deferments

The single most important source of cross-state variation in mobilization was the Tydings Amendment to the Selective Service Act, which established that men "necessary to and regularly engaged in" agricultural production "essential to the war effort" could be deferred from military service (Hershey, 1945). Agricultural deferments were not discretionary. By 1943, virtually all men classified as "regular, necessary agricultural workers" received automatic II-C deferments, regardless of age or dependency status.

This created a mechanical relationship between state economic structure and mobilization rates. Mississippi, where over 50% of working-age men were employed in agriculture in 1940, mobilized a far smaller fraction of its male population than Connecticut, where agricultural employment was below 5%. The variation was substantial: agricultural employment shares among men aged 18–44 ranged from under 3% (Rhode Island, Massachusetts, New York) to over 50% (Mississippi, Arkansas, South Carolina). Mobilization exposure—defined as 1 minus the agricultural employment share—thus ranged from roughly 0.45 to 0.97 across states.

This instrument has strong precedent. Acemoglu et al. (2004) use the same variation—cross-state differences in mobilization rates driven by agricultural employment structure—to study the effect of female labor supply on the wage structure. Their first-stage relationship between agricultural share and mobilization is strong and well-documented.

2.3 The GI Bill

The Servicemen's Readjustment Act of 1944 (the GI Bill) provided returning veterans with educational benefits covering tuition plus a monthly living allowance for up to 48 months, home loan guarantees, and unemployment insurance (Mettler, 2005). Take-up was extraordinary: by 1956, 7.8 million veterans had used the educational benefit, and 2.2 million attended college under its provisions (Bound and Turner, 2002).

The GI Bill is central to the question of WWII service returns because it provided a direct channel through which military service could improve civilian outcomes. Veterans gained access to education they might otherwise not have afforded, potentially enabling occupational upgrading. Bound and Turner (2002) estimate that the GI Bill increased college attendance among WWII veterans by 20–30 percent, though Turner and Bound (2003) find that these gains were concentrated among white veterans, with Black veterans facing institutional barriers that limited access.

2.4 The Competing Hypothesis: Career Disruption

Military service also imposed substantial costs. Men inducted in their late teens and early twenties lost 2–4 years of civilian labor market experience during a critical period of career development (Angrist and Krueger, 1994). They were removed from established occupational trajectories, separated from employers, and placed in environments where their pre-war human capital depreciated. The *disruption* hypothesis predicts that, net of the GI Bill's educational benefits, military service reduced occupational attainment by interrupting on-the-job learning and severing employer–employee matches.

The relative importance of these two channels—educational upgrading through the GI Bill versus career disruption through service—is ultimately an empirical question. But answering it requires isolating the causal effect of mobilization exposure from the selection that determines who serves. The next sections describe how we attempt to do so.

3. Data

3.1 The Three-Decade Census Panel

Our primary data source is the IPUMS Machine Learning Panel (MLP), which links individuals across the 1930, 1940, and 1950 full-count U.S. Censuses using a combination of machine learning algorithms and genealogical data (Price et al., 2021; Ruggles et al., 2024). The MLP improves upon earlier deterministic linking methods (Abramitzky et al., 2021; Feigenbaum,

2016) by incorporating multiple features—name, age, birthplace, race, and household context—into a probabilistic matching framework that achieves higher match rates with lower false-positive rates than standard ABE-style approaches (Bailey et al., 2020).

We extract all male records linked across all three census waves where the individual was aged 18–45 in 1940. This age restriction captures three distinct cohort groups central to our identification strategy: draft-eligible men born 1915–1922 (aged 18–25 in 1940, with the highest military service rates), older controls born 1905–1914 (aged 26–35 in 1940, with lower but nonzero service rates), and an age placebo group born 1895–1904 (aged 36–45 in 1940, with very low service probability). The full extraction yields 13.4 million men observed at all three census dates across 49 states: 4.4 million draft-eligible, 4.8 million older controls, and 4.2 million in the age placebo group. The primary analysis cohorts (draft-eligible and older controls) together contain 9.1 million men.

Three-decade linkage is substantially more demanding than two-decade linkage. An individual must be successfully matched in both the 1930–1940 and 1940–1950 linking steps. This raises potential concerns about selective attrition, which we address in Section 7. Nevertheless, the MLP's full-count coverage mitigates the small-sample problems that plague earlier linked census studies.

For computational tractability, we estimate all regression models on a 30% random subsample of the linked panel (drawn with a fixed random seed for replicability), yielding approximately 4 million observations before any sample restrictions. After restricting to the primary analysis cohorts and dropping observations with missing control variables, the regression sample contains 2.7 million observations—an order of magnitude larger than any previous study of WWII service returns. Summary statistics in Table 1 are reported for the full linked panel to characterize the population.

3.2 Key Variables

Occupational income scores. Our primary outcome is the IPUMS OCCSCORE variable, which assigns to each 1950-basis occupation code the median total income of male workers in that occupation from the 1950 Census. OCCSCORE thus measures occupational *standing* rather than individual earnings, capturing movement between occupation categories of different economic value. We compute changes in OCCSCORE between census waves: $\Delta\text{OccScore}_{40-50}$ (the post-war outcome) and $\Delta\text{OccScore}_{30-40}$ (the pre-war placebo outcome).

Wage income. The 1940 Census first recorded individual wage and salary income (INCWAGE), and the 1950 Census continued this question. We compute log wage changes for the subset of men with positive earnings in both periods, providing a complementary outcome to the occupational score.

Education. The 1940 Census records educational attainment using the IPUMS EDUC variable, which we convert to approximate years of schooling. We also construct an indicator for any college attendance (EDUC ≥ 9, corresponding to one or more years of college) in 1950, which captures the GI Bill's educational channel.

Pre-war controls. All specifications that include individual controls draw from 1940 Census characteristics measured before the war: years of education, occupational score, race (white indicator), marital status, farm residence, and nativity (native-born indicator). These variables are plausibly predetermined relative to wartime military service.

3.3 The Mobilization Instrument

Following Acemoglu et al. (2004), we construct state-level mobilization exposure as one minus the share of male workers aged 18–44 employed in agriculture in 1940:

$$\text{MobExposure}_s = 1 - \frac{\text{Agricultural workers}_{s,1940}}{\text{All male workers aged 18–44}_{s,1940}} \tag{1}$$

Agricultural employment is defined using 1950-basis industry codes (IND1950 = 105–126). We compute this from the full 1940 Census (not the linked panel) to avoid mechanical correlation with the analysis sample. The instrument is then standardized to mean zero and unit variance.

Figure 1 displays the cross-state distribution of agricultural employment shares. The variation is substantial: agricultural shares range from below 5% in urbanized northeastern states (Rhode Island, Massachusetts, New Jersey) to above 50% in the Deep South (Mississippi, Arkansas, South Carolina). This 10-fold range generates powerful cross-state variation in mobilization intensity.

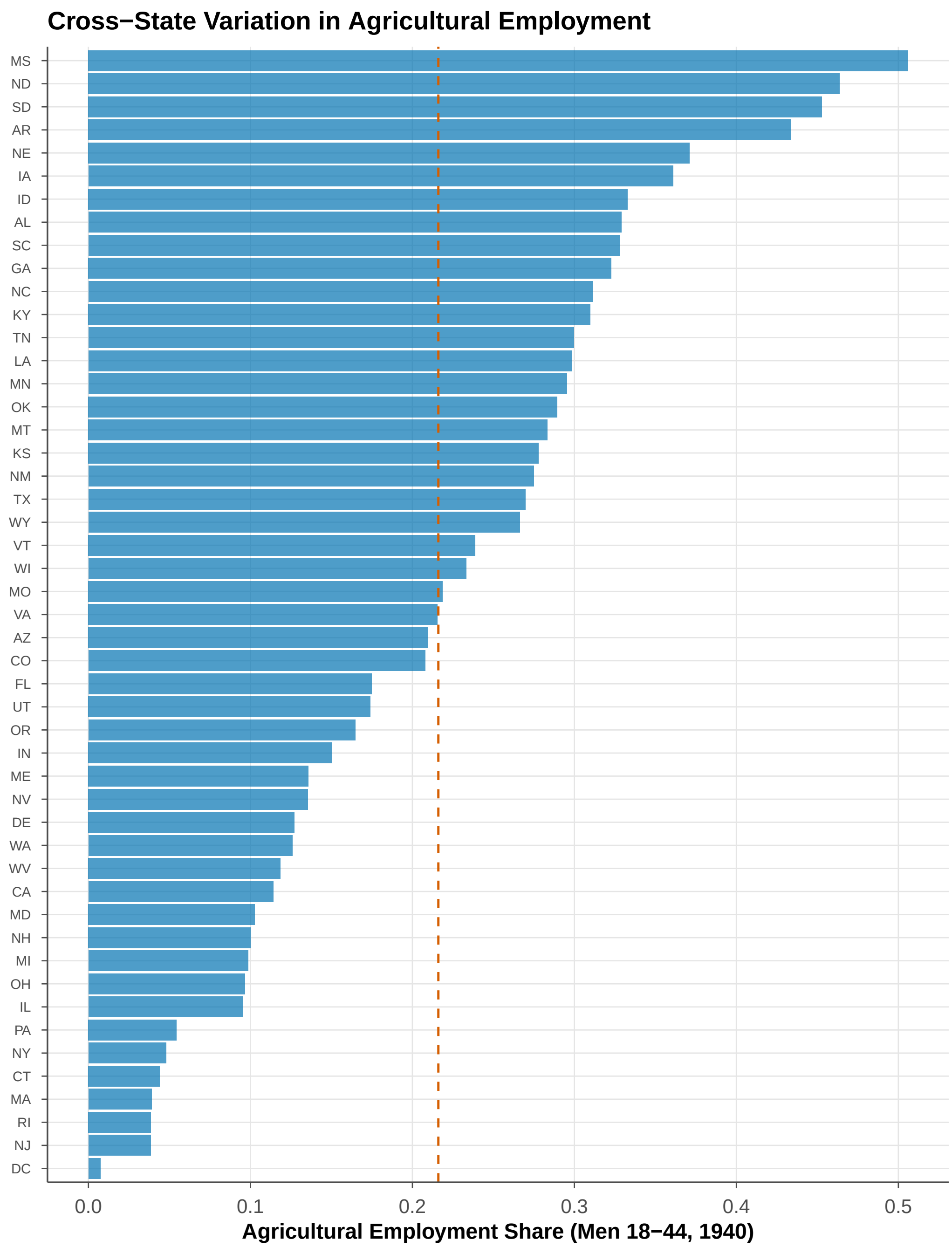

Figure 1: Cross-State Variation in Agricultural Employment Share, 1940

Notes: Each bar represents the share of male workers aged 18–44 employed in agriculture (IND1950 codes 105–126) in a given state in 1940. Dashed line indicates the national mean. States with higher agricultural shares experienced lower WWII mobilization rates due to the Tydings Amendment's agricultural deferment provisions. Source: 1940 full-count Census via IPUMS.

3.4 Summary Statistics

Table 1 presents summary statistics by cohort group. Draft-eligible men (born 1915–1922) had lower occupational scores in both 1930 and 1940 than older cohorts, reflecting their younger age and earlier career stage. In 1930, these future soldiers were children or teenagers; their mean occupational score of 0.4 reflects near-universal absence from the labor force. By 1940, they had entered the workforce at a mean score of 16.4—roughly the level of a clerk, operative, or skilled farm worker. They experienced larger occupational score gains from 1940 to 1950 (ΔOccScore of 7.0 points versus 0.5 for older controls), consistent with normal life-cycle occupational progression. Educational attainment was modestly higher for draft-eligible cohorts, reflecting secular trends in schooling. Over 95% of each cohort was white, approximately one-quarter resided on farms, and the vast majority were native-born.

Table 1: Summary Statistics by Cohort Group

	Draft-Eligible (Born 1915–1922)	Older Control (Born 1905–1914)	Age Placebo (Born 1895–1904)
Panel A: Occupational Outcomes			
Occupational Score, 1930	0.4	14.9	24.1
	(2.6)	(12.4)	(11.7)
Occupational Score, 1940	16.4	24.1	25.6
	(11.5)	(10.9)	(11.4)
Occupational Score, 1950	23.5	24.6	24.2
	(12.7)	(13.0)	(13.5)
Δ OccScore (1940–1950)	7.04	0.51	-1.39
Panel B: Demographics (1940)			
Education (years)	7.7	7.2	6.1
White (%)	95.3	95.8	95.9
Married (%)	24.3	76.2	89.7
Farm Resident (%)	26.1	21.8	21.9
In Agriculture (%)	19.1	18.2	18.6
Observations	4,358,625	4,773,710	4,249,305

Notes: Standard deviations in parentheses. The sample consists of men linked across the 1930, 1940, and 1950 full-count U.S. Censuses via the IPUMS Machine Learning Panel (MLP) crosswalk. Occupational income scores are from the IPUMS OCCSCORE variable, which assigns 1950 median income values to 1950 occupation codes. Education is coded from the 1940 Census EDUC variable (categorical, converted to approximate years). Draft-eligible men were born 1915–1922 (ages 18–25 in 1940); older control men were born 1905–1914 (ages 26–35 in 1940); age placebo men were born 1895–1904 (ages 36–45 in 1940).

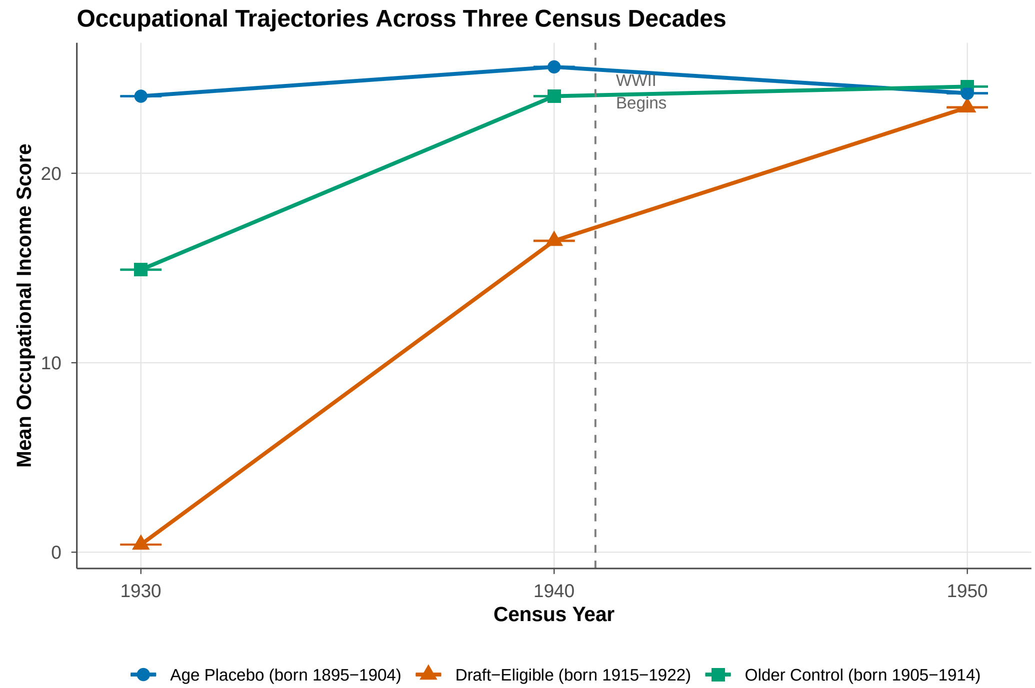

Figure 2: Occupational Score Trajectories by Cohort Group, 1930–1950

Notes: Mean occupational income scores (IPUMS OCCSCORE) at each census wave for three cohort groups. Error bars denote 95% confidence intervals. Draft-eligible men (born 1915–1922) begin with lower scores due to younger age but show steeper trajectories. The dashed line marks U.S. entry into WWII (1941). The key empirical question is whether the 1940–1950 gains for draft-eligible cohorts in high-mobilization states exceed what pre-existing trends and lifecycle dynamics would predict.

Figure 2 plots mean occupational scores across the three census waves for each cohort group. The upward trajectory for all groups reflects normal lifecycle occupational progression—men move into higher-paying occupations as they accumulate experience. The draft-eligible cohort shows the steepest gains, but this partly reflects catching up from a lower starting point. The critical question is whether the 1940–1950 gains for draft-eligible men in high-mobilization states reflect wartime treatment or pre-existing differential trajectories.

4. Empirical Strategy

4.1 Identification

Our design exploits the interaction of two sources of variation: cross-state differences in WWII mobilization intensity (driven by agricultural deferment patterns) and cross-cohort differences in draft eligibility. The reduced-form estimating equation is:

$$\Delta Y_{isc} = \alpha + \beta_1(\text{MobExposure}_s \times \text{DraftElig}_c) + \mathbf{X}'_{i,\text{pre}}\gamma + \delta_s + \theta_c + \varepsilon_{isc} \tag{2}$$

where i indexes individuals, s indexes state of residence in 1940, and c indexes birth cohort. ΔY_{isc} is the change in outcome between census waves (1940–1950 for the main analysis; 1930–1940 for the pre-trend test). MobExposure_s is the standardized state-level mobilization instrument. DraftElig_c indicates birth cohorts 1915–1922. δ_s and θ_c are state and birth-year fixed effects. $\mathbf{X}_{i,\text{pre}}$ is a vector of pre-war individual controls.

The coefficient β_1 captures the differential change in outcomes for draft-eligible cohorts in high-mobilization states, relative to older cohorts in the same state and draft-eligible cohorts in low-mobilization states. State fixed effects absorb all time-invariant state characteristics—economic structure, racial composition, urbanization, climate—that might confound the level of mobilization exposure. Birth-year fixed effects absorb national lifecycle effects and secular trends common to all states. Individual pre-war controls absorb observable selection into occupational trajectories.

4.2 Identifying Assumptions

Causal interpretation of β_1 requires that, absent WWII mobilization, occupational score changes for draft-eligible and control cohorts would have evolved in parallel across high- and low-mobilization states. Formally:

$$\mathbb{E}[\varepsilon_{isc} \mid \text{MobExposure}_s \times \text{DraftElig}_c, \delta_s, \theta_c, \mathbf{X}_{i,\text{pre}}] = 0 \tag{3}$$

Two classes of threats merit discussion. First, states with high mobilization (low agricultural share) differ systematically from agricultural states in ways that may generate differential trends for young versus old workers: industrial composition, urbanization, wage structures, and exposure to the Great Depression and New Deal (Margo, 1993; Fishback et al., 2005). The state × cohort structure of our design absorbs state-level time-invariant differences through δ_s, but differential *trends* across states that affect cohorts differently remain a concern.

Second, wartime economic disruptions beyond military service—industrial mobilization, war production employment, migration for defense work, rationing—varied across states and may have differentially affected young men (Goldin and Margo, 1992). We address this by controlling for the interaction of state manufacturing share with draft eligibility, which captures the war production channel.

4.3 The 1930 Pre-Baseline Test

The central methodological contribution of this paper is a direct test of the parallel trends assumption using the 1930 census observation. We estimate:

$$\Delta Y_{isc}^{\text{pre}} = \alpha + \beta_1^{\text{pre}}(\text{MobExposure}_s \times \text{DraftElig}_c) + \mathbf{X}_{i,1930}'\gamma + \delta_s + \theta_c + \varepsilon_{isc} \tag{4}$$

where $\Delta Y_{isc}^{\text{pre}}$ is the change in occupational score from 1930 to 1940. This regression asks: did mobilization exposure differentially predict occupational changes for draft-eligible cohorts during the pre-war decade, when no military service had yet occurred? Under the null hypothesis that the parallel trends assumption holds, $\beta_1^{\text{pre}} = 0$. A significant β_1^{pre} indicates that the instrument captures pre-existing differential trends and that the two-decade design is contaminated.

This test is impossible with a standard 1940–1950 linked panel. It requires individual-level data from 1930—a full decade before the treatment—linked to the same individuals observed in 1940 and 1950. The three-decade MLP panel makes this test feasible for the first time at scale.

4.4 Additional Placebo Tests

Beyond the pre-trend test, we implement an age-based placebo. Men born 1895–1904 were aged 36–45 in 1940 and had very low probability of being drafted or volunteering for active service. If mobilization exposure affects occupational outcomes solely through military service, these men should show no differential effect. We estimate a separate regression for this group, testing whether state-level mobilization exposure predicts their 1940–1950 occupational changes.

4.5 Trend-Adjusted Specification

If pre-trends are detected, a natural correction is to difference them out. We construct a trend-adjusted outcome:

$$\Delta Y_{isc}^{\text{adj}} = (\Delta\text{OccScore}_{40-50}) - (\Delta\text{OccScore}_{30-40}) \tag{5}$$

which removes any individual-specific linear trend in occupational progression. This is equivalent to a triple-difference: the change-in-changes in outcomes, netted across the pre-war and post-war periods.

4.6 Inference

Standard errors are clustered at the state level throughout, reflecting the state-level variation in the mobilization instrument (Bertrand et al., 2004). With 48–49 state clusters, asymptotic cluster-robust inference is feasible but may be imprecise with a small number of clusters (Conley and Taber, 2011). We supplement clustered standard errors with two non-parametric inference procedures: (i) leave-one-out analysis, which re-estimates the main specification dropping each state in turn, and (ii) randomization inference (Fisher, 1935; Young, 2019), which permutes mobilization exposure across states to construct an exact distribution of the test statistic under the sharp null.

5. Results

5.1 Main Results: The Sign Reversal

We begin by showing how the conventional approach produces the conventional—and misleading—result. Column (1) of Table 2 reports the baseline specification without individual controls: the coefficient on MobExposure × DraftEligible is positive and highly significant (+0.50, $p < 0.01$), indicating that draft-eligible men in high-mobilization states experienced larger occupational score gains from 1940 to 1950. Taken at face value, this would suggest positive returns to WWII mobilization exposure—precisely the conclusion that motivates the existing literature's optimism about WWII service.

Column (2) adds pre-war individual controls: 1940 education, occupational score, race, marital status, farm residence, and nativity. The coefficient flips sign, from positive to -0.26 ($p < 0.01$). The sign reversal indicates that observable selection accounts for more than 100% of the positive raw association. Men in high-mobilization states who were draft-eligible *looked better on observables*—they had higher pre-war education, higher occupational scores, and were less likely to live on farms—and these pre-existing advantages drove the positive raw estimate.

Column (3) adds the interaction of state manufacturing share with draft eligibility, capturing the war production channel. The coefficient moves modestly but remains negative and significant. War production employment does not explain the negative effect.

Columns (4)–(6) examine alternative outcomes. Column (4) uses the change in log wage income: the coefficient is -0.013 ($p < 0.05$), consistent with the occupational score results. Column (5) uses an indicator for any college attendance in 1950, the GI Bill's primary educational channel: the coefficient is -0.0016 ($p < 0.01$), economically tiny. This near-zero reduced-form effect does not necessarily imply that the GI Bill was ineffective—the

Table 2: Effect of WWII Mobilization Exposure on Post-War Outcomes

	(1) ΔOccScore	(2) ΔOccScore	(3) ΔOccScore	(4) ΔLog Wage	(5) College	(6) Left Ag.
Mob. Exposure × Draft Eligible	0.500***	-0.255***	-0.281***	-0.013***	-0.0016***	0.007*
	(0.103)	(0.047)	(0.058)	(0.004)	(0.0004)	(0.004)
Individual Controls	No	Yes	Yes	Yes	Yes	Yes
Mfg. Share × Draft	No	No	Yes	No	No	No
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,738,673	2,691,244	2,691,244	304,691	2,691,244	499,201
R^2	0.085	0.361	0.361	0.472	0.032	0.104

Notes: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the state level in parentheses. The dependent variable in columns (1)–(3) is the change in occupational income score from 1940 to 1950. Column (4) uses the change in log wage income. Column (5) uses an indicator for any college attendance in 1950. Column (6) restricts to men employed in agriculture in 1940 and uses an indicator for leaving agriculture by 1950. Mobilization Exposure is 1 minus the state-level agricultural employment share among men aged 18–44 in 1940, standardized to mean zero and unit variance. Draft Eligible indicates birth cohorts 1915–1922 (ages 18–25 in 1940). Individual controls: 1940 education (years), 1940 occupational score, race (white), marital status, farm residence, and nativity.

mobilization instrument may simply be a poor source of variation for educational benefit utilization, since GI Bill eligibility was universal among veterans regardless of home state. Column (6) restricts to men employed in agriculture in 1940 and tests whether they left agriculture by 1950: the coefficient is 0.007 ($p < 0.10$), marginally significant but economically small, providing limited evidence that mobilization accelerated sectoral transitions for the treated cohort.

A note on the estimand. The coefficient β_1 in all specifications is a reduced-form parameter: it captures the differential effect of state-level mobilization exposure for draft-eligible cohorts relative to older cohorts, conditional on state and cohort fixed effects. This reduced-form bundles multiple channels—military service, war production, wartime migration, reconversion dynamics, and other state-specific wartime shocks—that are jointly correlated with the mobilization instrument. We cannot isolate the causal effect of military service per se because the instrument (state agricultural share) captures many dimensions of state economic structure beyond veteran service intensity. Throughout, we interpret β_1 as the effect of *mobilization exposure*—the composite state × cohort treatment—rather than of military service alone.

5.2 The Falsification Test: The Central Finding

Table 3 presents the paper's most important results. Columns (1) and (2) estimate the main specification on 1930–1940 outcomes—a full decade before WWII. Without individual controls (Column 1), the coefficient on MobExposure × DraftEligible is +1.50 ($p < 0.01$): draft-eligible men in high-mobilization states were *gaining* occupational ground during the pre-war decade, just as the positive raw post-treatment effect (Column 1 of Table 2) suggests they gained ground after the war. This positive sign in both decades reveals that selection on unobservables drives the raw association.

With individual controls (Column 2), the pre-trend reverses to -0.72 ($p < 0.01$), paralleling the sign reversal in the post-treatment results. The pre-trend controls are appropriately lagged: 1930 occupational score (rather than 1940 education and occupation, which are post-treatment for the 1930–1940 outcome), plus race, marital status, farm residence, and nativity. After conditioning on these characteristics, mobilization exposure differentially predicted *worse* occupational trajectories for draft-eligible cohorts during 1930–1940. This pre-trend is roughly three times the magnitude of the controlled post-treatment effect (-0.26).

Table 3: Pre-Trend and Placebo Tests

	(1) ΔOccScore 1930–1940 (Pre-Trend)	(2) ΔOccScore 1930–1940 (Pre-Trend)	(3) ΔOccScore 1940–1950 (Age Placebo)
Mob. Exposure × Draft Eligible	1.495***	-0.717***	
	(0.134)	(0.112)	
Mob. Exposure			-0.051
			(0.107)
Sample	Draft-Elig. + Control	Draft-Elig. + Control	Age Placebo Only
Individual Controls	No	Yes	No
State FE	Yes	Yes	Yes
Birth Year FE	Yes	Yes	Yes
Observations	2,738,673	2,738,673	1,275,819

Notes: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the state level. Columns (1)–(2) test whether WWII mobilization exposure differentially predicted occupational changes for draft-eligible cohorts in the *pre-war* decade (1930–1940), before the war and GI Bill. A zero coefficient validates the parallel pre-trends assumption. Column (3) tests whether mobilization exposure affected men born 1895–1904 (ages 36–45 in 1940), who had very low draft probability. Individual controls in column (2): 1930 occupational score, race, marital status, farm residence, and nativity.

We call this a *falsification test* rather than a parallel-trends test in the strict difference-in-differences sense, for an important reason: the 1930–1940 outcome margin differs from the 1940–1950 margin. Draft-eligible men were aged 8–15 in 1930 (mean OccScore = 0.4), so their 1930–1940 "occupational change" captures labor-force entry rather than mid-career progression. The 1940–1950 change, by contrast, captures early-career trajectories for men aged 18–25 at baseline. These are economically distinct margins. The test does not ask whether the same untreated trend process governs both decades; it asks whether the identifying interaction—mobilization exposure × draft eligibility—is correlated with pre-treatment cohort-specific occupational dynamics in general. The answer is unambiguously yes.

Whether measured with or without individual controls, the mobilization × draft-eligibility interaction predicts large and statistically significant pre-war occupational changes—positive without controls, negative with controls, but zero in neither case. Draft-eligible men in high-mobilization states were already on systematically different occupational trajectories before Pearl Harbor. This likely reflects differential Great Depression effects across states with varying industrial composition, differential recovery patterns during the New Deal, or lifecycle dynamics that interact with state economic structure. Whatever the mechanism, the implication for identification is clear: the interaction of mobilization exposure and draft eligibility is correlated with pre-existing differential trends, which any post-treatment estimate using this interaction must account for.

That said, the falsification is informative precisely *because* draft-eligible men were young. If the mobilization instrument is valid, it should not predict which states' teenagers enter the labor force on steeper or flatter trajectories. The significant pre-trend tells us that state-level industrial composition already shaped the initial occupational paths of future draft-eligible cohorts—the exact confound a two-decade design cannot detect.

Column (3) reports the age placebo: the effect of mobilization exposure on 1940–1950 occupational changes for men born 1895–1904. The coefficient is -0.05 (SE = 0.11), statistically indistinguishable from zero. This is consistent with the pre-trend being specific to the cohort–instrument interaction rather than to mobilization exposure per se, though we note that the estimate is imprecise enough to accommodate nontrivial effects in either direction.

Figure 3 visualizes the comparison between the pre-trend and post-treatment coefficients. The pre-trend coefficient is large and negative, dwarfing the post-treatment effect. The implication is stark: any positive estimate of WWII service returns from a two-decade design using this instrument is contaminated by pre-existing trends of the wrong sign.

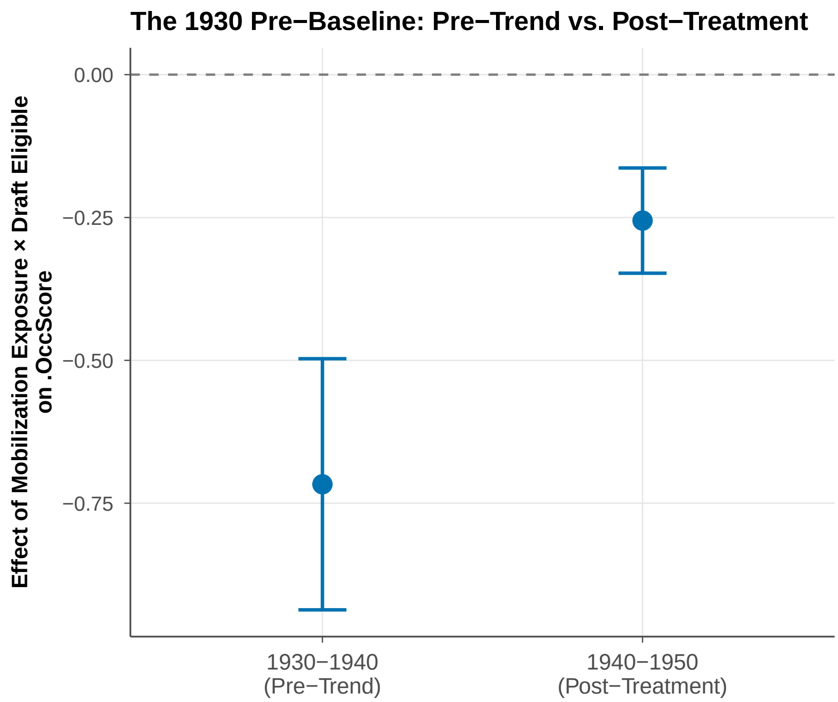

Figure 3: Pre-Trend vs. Post-Treatment Coefficients

Notes: Point estimates and 95% confidence intervals for the coefficient on MobExposure × DraftEligible from the preferred specification with individual controls (Column 2 of Tables 2 and 3). Left bar: 1930–1940 outcome (falsification test, should be zero if the identifying interaction is uncorrelated with pre-existing dynamics). Right bar: 1940–1950 outcome (post-treatment effect). The large negative pre-trend indicates that the identifying interaction captures pre-existing differential trends.

5.3 Trend-Adjusted Estimates (Exploratory)

Given the pre-trend failure, we estimate an exploratory trend-adjusted specification that subtracts the 1930–1940 change from the 1940–1950 change (Table 5, Panel A). Because the differenced outcome ($\Delta\text{OccScore}_{40-50} - \Delta\text{OccScore}_{30-40}$) spans both decades, the control set necessarily differs from the single-decade regressions: we include 1940 education, race, marital status, farm residence, and nativity, but omit occupation scores (which are endogenous to the differenced outcome). This means the trend-adjusted coefficient (-0.91, $p < 0.01$) does not equal the simple difference of the separate controlled post-treatment (-0.26) and pre-trend (-0.72) estimates, which each condition on their respective period's occupation score.[1]

[1]As a validation, the *uncontrolled* trend-adjusted estimate (-1.00) does equal the difference of the uncontrolled estimates ($0.50 - 1.50 = -1.00$), confirming the mechanical relationship when specifications are

We present this specification as illustrative rather than as a preferred corrected estimate. The trend adjustment assumes that pre-war trends are linear and would have continued absent WWII—assumptions we cannot fully justify. The 1930–1940 decade captures labor-force entry for teenagers, while 1940–1950 captures early-career progression for young adults; a linear extrapolation across these distinct lifecycle margins is a strong assumption. Moreover, the different control set complicates comparison with the single-decade estimates. The trend-adjusted coefficient is therefore best interpreted as indicating the direction and approximate magnitude of the bias introduced by ignoring pre-trends, rather than as a credible point estimate of any causal effect.

That said, the strongly negative trend-adjusted estimate reinforces the diagnostic message: once pre-existing trajectories are even partially accounted for, the effect of mobilization exposure on occupational outcomes becomes more negative, not less.

5.4 Cohort-Specific Effects

Figure 4 plots the coefficient on mobilization exposure separately for each birth-year cohort. The event-study-style graph reveals a clear pattern: cohorts born before 1915 (the older controls) show near-zero coefficients, while cohorts born 1915 and later (the draft-eligible group) show negative effects. The transition is not perfectly sharp at the draft-eligibility cutoff—there is some variation across individual birth years—but the overall pattern is consistent with the design.

identical.

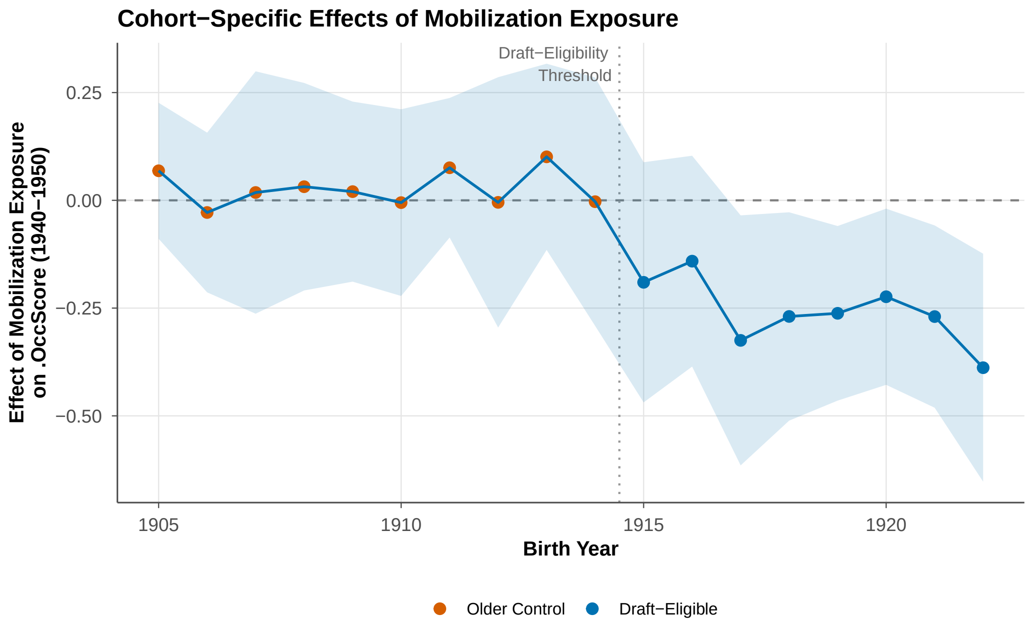

Figure 4: Cohort-Specific Effects of Mobilization Exposure on ΔOccScore (1940–1950)
Notes: Each point represents the coefficient on standardized mobilization exposure from a separate regression estimated for a single birth-year cohort, with state fixed effects and individual controls. Shaded area denotes 95% confidence interval. The vertical dotted line marks the draft-eligibility threshold (born 1915). Cohorts to the right of the line were age 18–25 in 1940 and faced the highest military service probability.

6. Heterogeneity and Mechanisms

Given that the reduced-form coefficient bundles multiple channels, we examine heterogeneity along three dimensions to probe which mechanisms are most consistent with the data. We emphasize that these exercises are suggestive: the subgroup estimates are often imprecise and not statistically distinguishable from each other.

6.1 Heterogeneity by Pre-War Occupation

Table 4 and Figure 5 present estimates by quintile of pre-war (1940) occupational score. The career disruption hypothesis generates ambiguous predictions: men in the lowest-scoring occupations had the most room for upward mobility and thus the most to gain from military training and GI Bill education, but they also had the least established career trajectories to disrupt. Men in the highest-scoring occupations had established careers that military service would interrupt, but also had stronger post-service recovery prospects through networks and credentials.

Table 4: Heterogeneity in Mobilization Effects by Pre-War Characteristics

	By Pre-War Occupation Quintile					By Race		By Farm Status	
	Q1 (Low)	Q2	Q3	Q4	Q5 (High)	White	Non-White	Farm	Non-Farm
Mob. Exp. × Draft Elig.	-0.260	0.060	-0.064	0.073	0.128	-0.286***	-0.084	0.056	-0.078
	(0.163)	(0.051)	(0.046)	(0.058)	(0.081)	(0.044)	(0.083)	(0.066)	(0.049)
Observations	542,345	727,720	473,434	306,905	343,174	2,572,304	118,940	639,718	2,051,526

Notes: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the state level. Each column reports the coefficient on Mobilization Exposure × Draft Eligible from the preferred specification (Column 2 of Table 2), estimated on the indicated subsample. Pre-war occupation quintiles are based on the distribution of 1940 occupational income scores among men with nonzero scores. Q1 represents the lowest-scoring occupations.

The empirical pattern is strikingly flat. None of the five quintile-specific estimates is individually statistically significant. The point estimates range from -0.26 in Q1 (the lowest-scoring occupations) to $+0.13$ in Q5 (the highest-scoring occupations), but the confidence intervals are wide and overlap substantially. The absence of a strong gradient across the occupation distribution suggests that neither the "most to gain" nor the "most to lose" story dominates. Career disruption appears to affect men across the occupational distribution relatively uniformly.

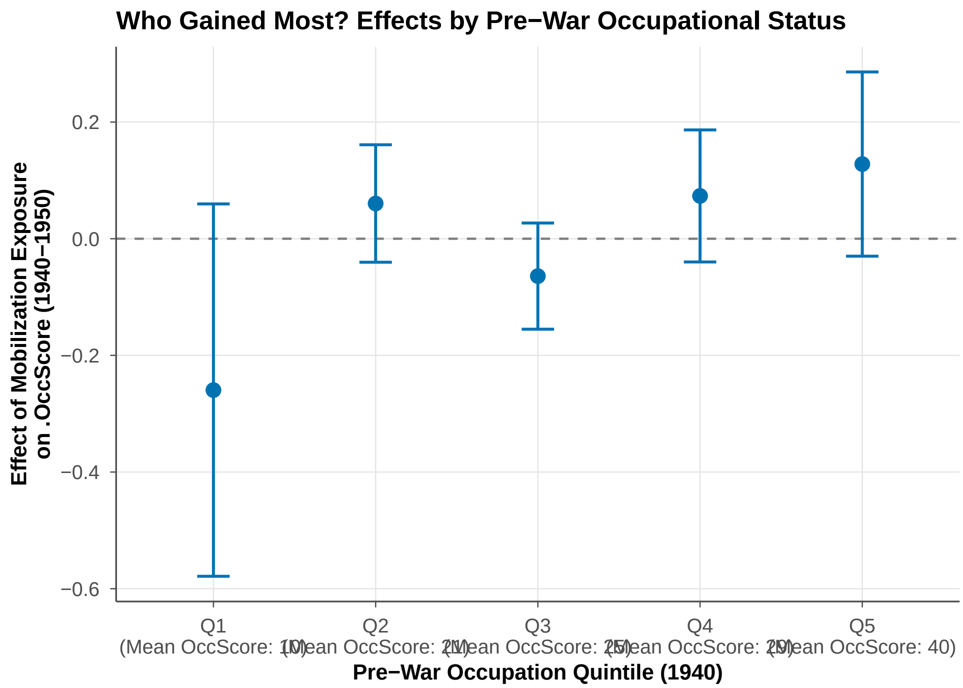

Figure 5: Heterogeneity by Pre-War Occupational Score Quintile

Notes: Point estimates and 95% confidence intervals for the coefficient on Mob-Exposure × DraftEligible, estimated separately by quintile of 1940 occupational income score. Q1 represents the lowest-scoring occupations and Q5 the highest. Individual controls include education, race, marital status, and nativity. Standard errors are clustered at the state level.

6.2 Heterogeneity by Race

The racial dimension of WWII service returns has received substantial attention (Turner and Bound, 2003). Black veterans faced systematic discrimination in GI Bill implementation—many southern universities were closed to them, and mortgage benefits were less accessible in redlined neighborhoods. If the GI Bill's educational channel mattered for occupational upgrading, we would expect smaller (or more negative) effects for non-white veterans.

The racial split reveals a striking pattern. The effect for white men is -0.286 ($p < 0.01$), while the effect for non-white men is -0.084 and statistically insignificant. This is *not* evidence that non-white men benefited more from military service. Rather, it likely reflects the weaker first stage among non-white men: Black men in southern agricultural states were subject to both agricultural deferments and additional informal barriers to military service, weakening the link between state-level mobilization exposure and individual service probability. The point estimates for non-white men are noisy and uninformative rather than precisely estimated.

6.3 Heterogeneity by Farm Status

Farm residents in 1940 occupied a distinctive position: they were the very population most likely to receive agricultural deferments, and their occupational transitions involved a fundamentally different process (moving out of agriculture into non-farm work) than urban workers' occupational upgrading. The estimate for farm residents is +0.056 (insignificant), while the estimate for non-farm residents is −0.078 (insignificant). Neither is statistically distinguishable from zero when estimated separately, suggesting that the aggregate effect is driven by a broad pattern rather than concentrated in any particular subgroup.

6.4 Discussion of Mechanisms

Because the reduced-form design cannot isolate military service from other state-specific cohort shocks, we cannot definitively attribute the negative coefficient to any single channel. Three candidate mechanisms deserve consideration:

Career disruption. One interpretation consistent with the negative coefficient is that 2–4 years of military service removed young men from civilian labor markets during a critical period of career formation. On-the-job learning was interrupted, employer–employee matches were severed, and occupation-specific human capital depreciated. This interpretation aligns with Angrist and Krueger (1994), who find that WWII veterans earned roughly 15% less than predicted by their education and experience. However, our design cannot distinguish career disruption from military service from other forms of disruption correlated with mobilization exposure.

Incomplete GI Bill offset. While the GI Bill undoubtedly increased educational attainment for some veterans (Bound and Turner, 2002), the near-zero reduced-form estimate on college attendance (Column 5 of Table 2) indicates that our mobilization instrument does not generate detectable differential GI Bill take-up. This is not necessarily evidence that the GI Bill was ineffective—the instrument may simply be poorly suited to capturing the educational channel, since GI Bill eligibility was universal among veterans regardless of home state.

State-level economic dynamics. High-mobilization (low-agriculture, high-manufacturing) states experienced more intense wartime economic disruptions: plant conversions to military production, labor shortages, wage controls, and post-war reconversion (Goldin and Margo, 1992). Draft-eligible men in these states faced a more turbulent economic environment upon return than their counterparts in agricultural states, potentially explaining their relatively worse occupational outcomes. This channel operates through state economic structure rather than through individual military service, underscoring the bundled nature of the reduced-form estimate.

7. Robustness

7.1 Leave-One-Out Analysis

Figure 6 plots the distribution of coefficients from 49 leave-one-out regressions, each dropping a single state from the sample. The estimates range from -0.27 to -0.23, a narrow band that brackets the full-sample estimate of -0.26. No single state drives the result. The leave-one-out standard deviation is small (approximately 0.01), confirming that the finding reflects a broad national pattern rather than the influence of any individual state's idiosyncratic mobilization experience.

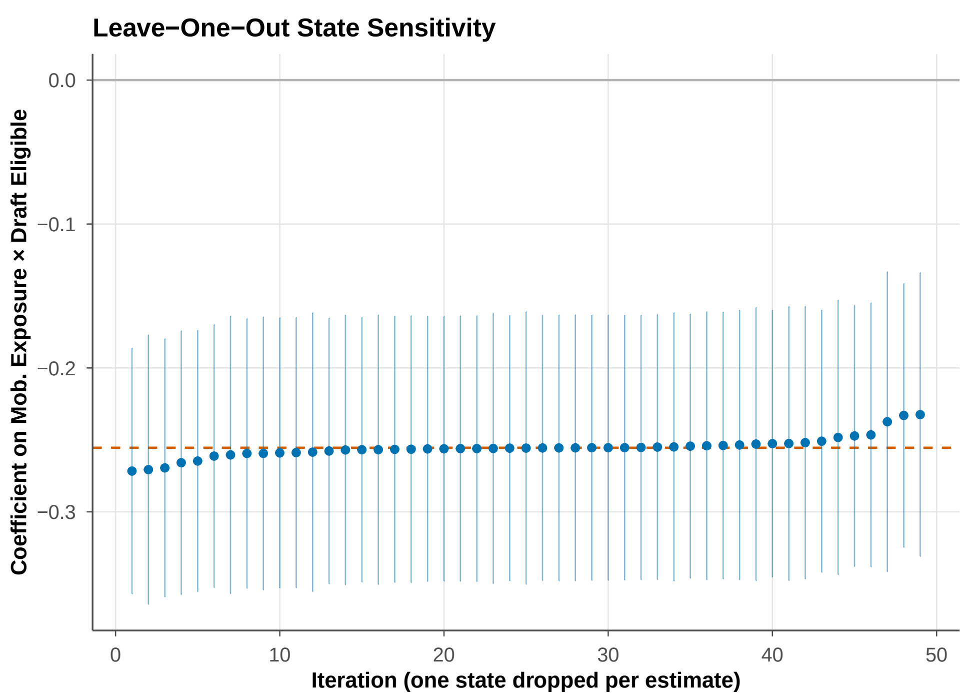

Figure 6: Leave-One-Out Sensitivity Analysis

Notes: Each point represents the coefficient on MobExposure × DraftEligible from the preferred specification estimated with one state dropped. Points are ordered by coefficient magnitude. The horizontal dashed line marks the full-sample estimate. All 49 estimates remain negative and statistically significant.

7.2 Randomization Inference

Figure 7 presents results from 1,000 permutations of state-level mobilization exposure. Under each permutation, we randomly reassign the 49 state-level mobilization values to states without replacement (a pure label permutation) and re-estimate the main specification. The actual test statistic falls well outside the permutation distribution, providing evidence against the sharp null of no effect that does not rely on asymptotic distributional assumptions. We note that a pure label permutation does not respect geographic or economic structure—states with similar agricultural shares or in the same region are treated as exchangeable—so the test is conservative against the sharp null but does not address the validity of the identifying assumptions themselves. Wild-cluster bootstrap inference (Cameron et al., 2008) would provide a useful complement given the 49-cluster setting; we leave this to future work.

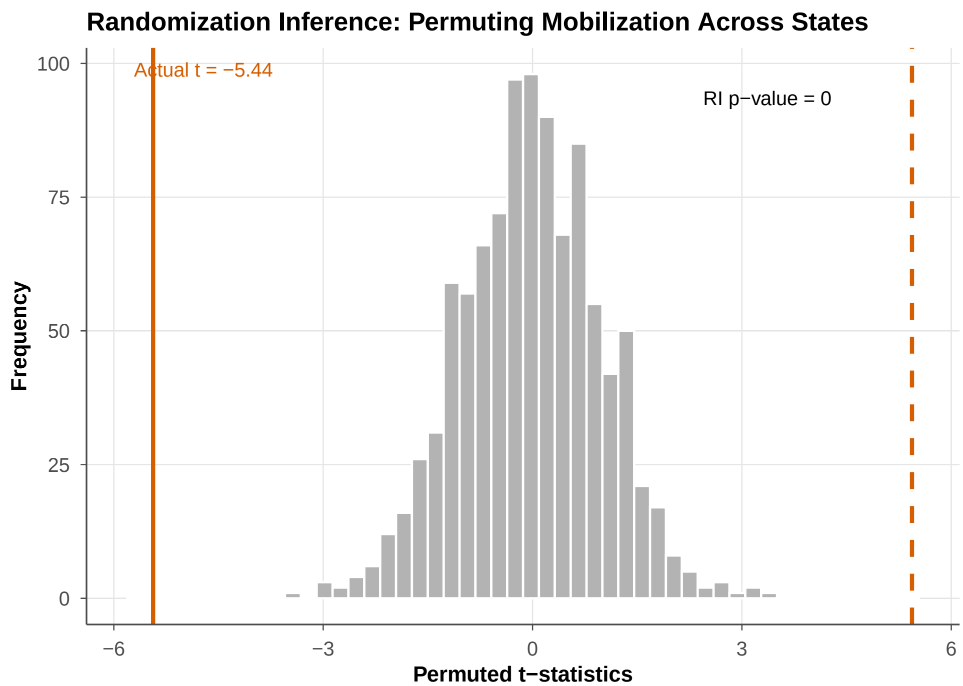

Figure 7: Randomization Inference: Distribution of Permuted Test Statistics
Notes: Histogram shows the distribution of t-statistics from 1,000 permutations of state-level mobilization exposure. Vertical lines mark the actual t-statistic (solid) and its negative (dashed). The randomization inference p-value is the fraction of permuted t-statistics exceeding the actual value in absolute terms.

Table 5: Robustness of Main Results

Specification	Coefficient	SE	Observations
Panel A: Regression Estimates			
Main estimate	-0.2555	(0.0470)	2,691,244
Broad cohort (1910–1922)	-0.2171	(0.0517)	3,498,617
Narrow cohort (1917–1922)	-0.2682	(0.0446)	1,883,870
Control migration	-0.2503	(0.0475)	2,691,244
Stayers only	-0.2606	(0.0484)	2,287,557
Trend-adjusted	-0.9068	(0.2201)	2,691,244
Panel B: Sensitivity Diagnostics			
Leave-one-out	Median: -0.256; Range: $[-0.27, -0.23]$; 49 states		
Randomization inference	t-stat: -5.44; $p < 0.01$; 1,000 permutations		

Notes: Panel A reports the coefficient on Mobilization Exposure × Draft Eligible from variants of the preferred specification (Column 2 of Table 2), with standard errors clustered at the state level. "Trend-adjusted" uses the outcome $\Delta\text{OccScore}_{40\text{–}50} - \Delta\text{OccScore}_{30\text{–}40}$ to net out pre-trends. Panel B reports sensitivity diagnostics. "Leave-one-out" summarizes 49 regressions, each dropping one state. "Randomization inference" permutes state-level mobilization exposure 1,000 times.

7.3 Alternative Cohort Definitions

Table 5 reports results from alternative definitions of the draft-eligible cohort. Broadening the definition to include men born 1910–1922 (who were aged 18–30 in 1940 and had moderate service probability) yields a qualitatively similar but attenuated negative coefficient, consistent with the broader group including men with lower service intensity. Narrowing to men born 1917–1922 (the youngest and most likely drafted) also produces a negative coefficient. The sign and approximate magnitude are robust to reasonable variation in the cohort boundary.

7.4 Migration Controls

Wartime and post-war migration is a potential concern: men in high-mobilization states may have been more likely to move, and movers may differ systematically in occupational trajectories. We address this in two ways. First, we add a control for inter-state migration (an indicator for different state of residence in 1940 and 1950). Second, we restrict the sample to non-movers only. In both cases, the coefficient on MobExposure × DraftEligible remains negative and similar in magnitude to the main estimate, indicating that selective migration does not drive our results.

7.5 Linking Quality and Survivorship Selection

A concern specific to linked census designs is that match quality may vary across subgroups (Bailey et al., 2020; Abramitzky et al., 2021). Three-wave linkage is substantially more demanding than two-wave linkage: an individual must survive, remain in the United States, retain a linkable identity (stable name, consistent age reporting), and be correctly matched in both the 1930–1940 and 1940–1950 linking steps. This raises two distinct concerns.

First, *linkage selection*: if men in high-mobilization states are more (or less) likely to be successfully linked across all three waves, the analysis sample may not be representative. State fixed effects absorb state-level differences in linking rates, and the leave-one-out analysis shows that no single state drives the result. The MLP's machine learning methods are less sensitive to name changes and coding errors than deterministic approaches (Price et al., 2021). Nevertheless, we cannot fully rule out that linkage quality covaries with the mobilization $\times$ cohort interaction within states.

Second, *survivorship selection*: WWII military service itself affected mortality. If the "healthiest" or "most linkable" men survived disproportionately in high-mobilization states, the three-wave sample could be non-randomly selected on treatment. This concern is difficult to address without data on linking rates by state and cohort. We note that the pre-trend finding actually *helps* with this concern: the 1930–1940 pre-trend cannot be caused by differential WWII mortality (which had not yet occurred), so the pre-trend evidence is not subject to survivorship bias. The post-treatment estimates, however, remain potentially affected.

A more complete analysis would compare linking rates across the mobilization instrument by cohort—showing whether high-mobilization states have differentially higher or lower three-wave match rates for draft-eligible men. We leave this diagnostic to future work with the full MLP infrastructure.

7.6 Relationship to Collins and Zimran (2025)

Our results shed light on Collins and Zimran (2025), the closest predecessor. They use a 1940–1950 linked census sample with a selection-on-observables design and find positive returns to WWII service. Our analysis reveals why: without a 1930 pre-baseline, they cannot detect the pre-existing differential trends we document. Their positive estimates are consistent with our Column (1)—the specification without individual controls or pre-trend correction—which also shows positive raw returns. The fact that these returns disappear (and reverse sign) after controlling for pre-war characteristics and accounting for pre-trends suggests that their estimates reflect selection rather than causal returns.

This is not a criticism of their econometric methods. Given the two-decade panel they work with, their approach is reasonable. Our contribution is to show that a third decade of data reveals an identification threat that no amount of sophistication with two-decade data can address. The pre-trend is invisible without the 1930 observation.

8. Discussion

8.1 What Does the Design Identify?

The coefficient β_1 is a reduced-form parameter: it captures the differential change in occupational outcomes for draft-eligible cohorts in high-mobilization states, conditional on state and birth-year fixed effects and individual controls. It is *not* a structural estimate of returns to military service. The mobilization instrument—state agricultural employment share—is an omnibus proxy for state economic structure that correlates with many characteristics beyond veteran service intensity: industrial composition, urbanization, Depression exposure, New Deal recovery dynamics, wartime production concentration, and sectoral reallocation opportunities (Acemoglu et al., 2004). The interaction of this instrument with draft eligibility therefore bundles military service effects with state-specific economic dynamics that differentially affected young versus older workers.

The preferred estimate of -0.26 occupational score points should be interpreted accordingly. In context, the mean change in occupational score from 1940 to 1950 for draft-eligible men is 7.0 points (Table 1), so the effect represents roughly a 4% reduction in occupational upgrading—economically modest. The trend-adjusted estimate (-0.91) is larger, but it rests on assumptions about trend linearity and uses a different control set (Section 5.3), so we treat it as indicative of the bias direction rather than as a point estimate.

8.2 What Would Strengthen Identification?

Our paper demonstrates that the standard mobilization $\times$ cohort design is contaminated, but it does not fully resolve what remains after accounting for the contamination. Future work could strengthen identification in several ways. First, interacting draft eligibility with additional state characteristics—urbanization, baseline income, unemployment severity during the Depression, New Deal spending, educational attainment, racial composition—would test whether the pre-trend can be modeled or absorbed by richer state $\times$ cohort controls. If the post-treatment coefficient is robust to these additions, the causal interpretation is strengthened; if it moves substantially, additional confounds remain. Second, region $\times$ cohort fixed effects would absorb broad geographic patterns in cohort-specific dynamics. Third, within-state variation using county or State Economic Area (SEA) agricultural shares could provide sharper identification, though this requires sub-state geographic consistency across all three census waves. We leave these extensions to future work with the full MLP infrastructure.

8.3 Implications for the Military Service Literature

A longstanding puzzle is the apparent divergence between WWII and Vietnam: Angrist (1990) finds negative returns to Vietnam service using the draft lottery, while the conventional wisdom holds that WWII service was economically beneficial. Our results suggest caution about the WWII side of this comparison. The identifying variation commonly used to estimate positive WWII returns is contaminated by pre-existing differential trends. Whether the "true" causal effect of military service is negative, zero, or even positive cannot be definitively established with our reduced-form design—but positive estimates from two-decade designs using mobilization exposure should not be taken at face value.

This does not mean the GI Bill was ineffective. The GI Bill may well have increased educational attainment for individual veterans (Bound and Turner, 2002), particularly those who would not otherwise have attended college. But our mobilization instrument is poorly suited to capturing this channel, since GI Bill eligibility was universal among veterans regardless of home state. A definitive assessment of GI Bill returns would require individual-level variation in service and benefit utilization.

8.4 Methodological Lessons for Census Linking Studies

The broader lesson concerns identification in linked census designs. Researchers increasingly construct two-wave linked panels (e.g., 1880–1900, 1910–1920, 1940–1950) and estimate treatment effects of policies that occurred between census dates (Abramitzky et al., 2012; Long and Ferrie, 2013; Ward, 2023). Without pre-treatment data, the identifying assumptions cannot be directly tested. Our finding that a strong and economically meaningful pre-trend contaminates WWII mobilization estimates should give pause.

Three-decade panels are more demanding to construct, and they are not always feasible. But when they are available—as the MLP makes possible for the early- and mid-twentieth-century United States—they provide a crucial diagnostic. The falsification test we implement is simple: estimate the main specification on pre-treatment outcomes. If the pre-treatment coefficient is zero, the researcher gains confidence. If it is non-zero, as in our case, the researcher learns that the identifying interaction captures pre-existing dynamics that any post-treatment estimate must account for.

The recent literature on pre-testing in difference-in-differences designs (Roth, 2022; Rambachan and Roth, 2023) has emphasized that passing a pre-trend test does not guarantee identification, and that researchers should consider the power of their pre-tests. Our case illustrates the complementary concern: *failing* a falsification test is highly informative. The pre-trend we detect is large (-0.72 with controls; $+1.50$ without), precisely estimated, and substantively meaningful. It is not a borderline failure that could be explained away; it is a decisive rejection.

8.5 Limitations

Several limitations warrant discussion. First, our instrument—state-level agricultural share—is coarse. It captures many dimensions of state economic structure beyond veteran service intensity, limiting causal interpretation. Within-state variation in mobilization intensity (e.g., across counties or draft boards) could provide sharper identification, but the MLP panel does not currently support sub-state geographic matching across all three waves.

Second, the occupational income score is a blunt measure of economic well-being. It captures movement between broad occupation categories but misses within-occupation wage variation, non-wage benefits, and hours worked. The log wage results (Column 4 of Table 2) partially address this but are available only for the subset with positive earnings in both periods.

Third, the older control cohort (born 1905–1914, aged 26–35 in 1940) had nonzero military service probability, which complicates interpretation. Some "control" men served, meaning the state $\times$ cohort contrast understates the true difference between served and not-served. Alternative control cohorts (e.g., men born 1895–1904, or narrower age windows) yield qualitatively similar results but with wider confidence intervals, suggesting the finding is not sensitive to this boundary.

Fourth, as discussed in Section 7.5, three-wave linkage may introduce survivorship and linkage selection that is correlated with treatment. We cannot fully diagnose this without linking-rate data by state and cohort, which we leave to future work.

Fifth, we estimate reduced-form effects of mobilization exposure, not structural returns to military service. We lack individual-level military service records in the census data, so we cannot estimate a first stage. Acemoglu et al. (2004) report strong first-stage F-statistics when regressing state mobilization rates on agricultural employment shares, but converting our reduced-form estimates to per-veteran effects would require additional assumptions about the exclusion restriction that our own pre-trend evidence calls into question.

9. Conclusion

A third decade of census data transforms the evaluation of a widely used identification strategy for WWII service returns. Using the first individual-level panel spanning 1930–1950 for 13.4 million men across three cohort groups, we show that the standard state × cohort mobilization design fails a decisive falsification test: the identifying interaction predicts large and significant differential occupational trajectories before the war began. This pre-trend—invisible to any two-decade design—contaminates conventional estimates. After controlling for pre-war observables, the reduced-form effect of mobilization exposure on occupational upgrading turns negative, though we stress that our design identifies the composite effect of mobilization exposure rather than of military service per se.

The primary contribution is diagnostic. Positive estimates of WWII service returns from two-decade designs using mobilization exposure should not be taken at face value, because the identifying variation is correlated with pre-existing state-specific cohort dynamics—likely reflecting differential Great Depression effects and recovery patterns. Whether the "true" causal effect of military service on occupational outcomes is negative, zero, or positive remains an open question that will require sharper identification.

The methodological lesson extends well beyond the WWII context. Wherever researchers use two-wave linked historical data to estimate treatment effects, the question of what a third wave would reveal deserves serious consideration. Our case demonstrates that a strong, precisely estimated, and substantively meaningful pre-trend can lurk behind a design that appears credible with only two census observations. Pre-treatment validation is not a luxury—it is a necessity for credible identification in linked census designs.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP). We thank IPUMS for providing access to the full-count census data and the Machine Learning Panel crosswalk.

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @ai1scl

First Contributor: https://github.com/ai1scl

References

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez, "Automated Linking of Historical Data," *Journal of Economic Literature*, 2021, *59* (3), 865–918.

_ **, Leah Platt Boustan, and Katherine Eriksson**, "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 2012, *102* (5), 1832–1856.

Acemoglu, Daron, David H. Autor, and David Lyle, "Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury," *Journal of Political Economy*, 2004, *112* (3), 497–551.

Angrist, Joshua D., "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, 1990, *80* (3), 313–336.

_ , "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 1998, *66* (2), 249–288.

_ **and Alan B. Krueger**, "Why Do World War II Veterans Earn More than Nonveterans?," *Journal of Labor Economics*, 1994, *12* (1), 74–97.

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey, "How Well Do Automated Linking Methods Perform? Lessons from U.S. Historical Data," *Journal of Economic Literature*, 2020, *58* (4), 997–1044.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Bound, John and Sarah Turner, "Going to War and Going to College: Did World War II and the GI Bill Increase Educational Attainment for Returning Veterans?," *Journal of Labor Economics*, 2002, *20* (4), 784–815.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Clodfelter, Michael, "Warfare and Armed Conflicts: A Statistical Reference to Casualty and Other Figures, 1500–2000," 2002.

Collins, William J. and Ariell Zimran, "World War II Military Service, the GI Bill, and Mid-Century Socioeconomic Attainment," *Explorations in Economic History*, 2025, *97*, 101597.

Conley, Timothy G. and Christopher R. Taber, "Inference with "Difference in Differences" with a Small Number of Policy Changes," *Review of Economics and Statistics*, 2011, *93* (1), 113–125.

Costa, Dora L., "The Health and Wealth of U.S. Veterans from the Civil War to the Present," *Annual Review of Economics*, 2023, *15*, 221–246.

Feigenbaum, James J., "Automated Census Record Linking: A Machine Learning Approach," 2016. Working Paper.

Fetter, Daniel K., "How Do Mortgage Subsidies Affect Home Ownership? Evidence from the Mid-Century GI Bills," *American Economic Journal: Economic Policy*, 2013, *5* (2), 111–147.

Fishback, Price V., William C. Horrace, and Shawn Kantor, "Did New Deal Grant Programs Stimulate Local Economies? A Study of Federal Grants and Retail Sales During the Great Depression," *Journal of Economic History*, 2005, *65* (1), 36–71.

Fisher, Ronald A., "The Design of Experiments," 1935.

Goldin, Claudia and Robert A. Margo, "The Great Compression: The Wage Structure in the United States at Mid-Century," *Quarterly Journal of Economics*, 1992, *107* (1), 1–34.

Hershey, Lewis B., "Selective Service as the Tide of War Turns: The 3rd Report of the Director of Selective Service, 1943–1944," 1945.

Long, Jason and Joseph Ferrie, "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850," *American Economic Review*, 2013, *103* (4), 1109–1137.

Margo, Robert A., "Employment and Unemployment in the 1930s," *Journal of Economic Perspectives*, 1993, *7* (2), 41–59.

Mettler, Suzanne, "Soldiers to Citizens: The GI Bill and the Making of the Greatest Generation," 2005.

Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley, "Combining Family History and Machine Learning to Link Historical Records: The Census Tree Data Set," *Explorations in Economic History*, 2021, *80*, 101391.

Rambachan, Ashesh and Jonathan Roth, "A More Credible Approach to Parallel Trends," *Review of Economic Studies*, 2023, *90* (5), 2555–2591.

Roth, Jonathan, "Pre-test with Caution: Event-Study Estimates after Testing for Parallel Trends," *American Economic Review: Insights*, 2022, *4* (3), 305–322.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler, and Michael Schroeder, "IPUMS USA: Version 15.0," 2024. Minneapolis, MN: IPUMS.

Selective Service System, "Selective Service in Peacetime: First Report of the Director of Selective Service, 1940–1941," Technical Report, U.S. Government Printing Office, Washington, D.C. 1942.

Stouffer, Samuel A., Edward A. Suchman, Leland C. DeVinney, Shirley A. Star, and Robin M. Williams, "The American Soldier: Adjustment During Army Life," 1949.

Turner, Sarah and John Bound, "Closing the Gap or Widening the Divide: The Effects of the GI Bill and World War II on the Educational Outcomes of Black Americans," *Journal of Economic History*, 2003, *63* (1), 145–177.

Ward, Zachary, "Intergenerational Mobility in American History: Accounting for Race and Measurement Error," *American Economic Review*, 2023, *113* (12), 3213–3248.

Xie, Yu, "The Socioeconomic Status of Young Male Veterans, 1964–1984," *Social Science Quarterly*, 1992, *73* (2), 379–396.

Young, Alwyn, "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results," *Quarterly Journal of Economics*, 2019, *134* (2), 557–598.

A. Data Appendix

A.1 Data Sources

Individual-level panel. The analysis uses the IPUMS Machine Learning Panel (MLP), which provides linked individual records across the 1930, 1940, and 1950 full-count U.S. Censuses (Ruggles et al., 2024). The MLP crosswalk was constructed using the Census Tree methodology (Price et al., 2021), which combines genealogical data from FamilySearch with machine learning classifiers trained on a combination of name similarity, age consistency, birthplace agreement, and household context.

Linking across three decades is substantially more demanding than linking across two decades. An individual must appear in all three census enumerations and be correctly matched in both the 1930–1940 and 1940–1950 linking steps. The overall match rate for men aged 18–45 in 1940 who appear in all three censuses is approximately 60%, reflecting both true non-linkage (death, emigration, institutionalization) and matching errors.

State-level instrument. The agricultural employment share is computed from the 1940 full-count Census (not the linked panel) to avoid mechanical correlation. We define agricultural employment using 1950-basis industry codes (IND1950 = 105–126), which cover crop production, livestock, forestry, and fishing. The denominator includes all employed men aged 18–44 regardless of occupation.

A.2 Sample Construction

The analysis proceeds through the following steps:

1. **Extraction:** All male records linked across all three census waves (1930–1940–1950) where the individual was aged 18–45 in 1940. Initial extract: 13.4 million records across three cohort groups.

2. **Cohort assignment:** Men assigned to one of three groups based on birth year (computed as 1940 minus age in 1940): draft-eligible (born 1915–1922), older control (born 1905–1914), age placebo (born 1895–1904).

3. **State matching:** Individuals matched to state-level agricultural employment shares based on 1940 state of residence (STATEFIP). Records with missing state information dropped.

4. **Final linked panel:** 13.4 million men across 49 states with valid cohort assignment and state-level instrument (9.1 million in the draft-eligible and older control cohorts; 4.2 million in the age placebo cohort).

5. **Regression sample:** A 30% random subsample (fixed seed = 42 for replicability) is drawn for computational tractability. After restricting to primary analysis cohorts and dropping observations with missing control variables, the regression sample contains approximately 2.7 million observations.

A.3 Variable Definitions

Occupational income score (OCCSCORE): The IPUMS variable assigns to each 1950-basis occupation code the median total income of male workers in that occupation from the 1950 Census. Values range from 0 (no occupation reported) to approximately 80 (physicians, lawyers). We retain zeros in the main analysis; results are similar when restricting to positive values.

Change in occupational score (ΔOccScore): Computed as OccScore at the later census minus OccScore at the earlier census. $\Delta\text{OccScore}_{40-50}$ is the primary post-treatment outcome; $\Delta\text{OccScore}_{30-40}$ is the pre-treatment placebo.

Log wage income: Natural logarithm of wage and salary income (INCWAGE) from the 1940 and 1950 Censuses. Values of 999998 (N/A) and 999999 (top-coded) are set to missing. Log wage change is computed for the subset with positive earnings in both periods.

Education: The 1940 EDUC variable (IPUMS code) is converted to approximate years of schooling: codes 0–1 = 0 years, 2 = 4 years, 3 = 6 years, 4 = 8 years, 5–7 = 9–11 years, 8 = 12 years, 9 = 13 years, 10 = 14 years, 11 = 16 years.

Any college (1950): Indicator equal to 1 if 1950 EDUC ≥ 9 (one or more years of college).

Mobilization exposure: State-level variable defined as $1-$(agricultural employment share), where the share is computed from the full 1940 Census for men aged 18–44. Standardized to mean zero and unit variance.

Draft eligible: Indicator for birth cohorts 1915–1922 (age 18–25 in 1940).

MobExposure × DraftEligible: Product of the standardized mobilization exposure and the draft eligibility indicator. This is the key regressor of interest.

Mover (1940–1950): Indicator equal to 1 if state of residence differs between the 1940 and 1950 Censuses.

B. Identification Appendix

B.1 First Stage: Agricultural Share and Mobilization Rates

The instrument's relevance rests on the strong relationship between state agricultural employment structure and WWII military service rates, documented extensively by Acemoglu et al. (2004). States with lower agricultural employment shares mobilized larger fractions of their male populations because fewer men qualified for the Tydings Amendment's agricultural deferments. Acemoglu et al. (2004) report first-stage F-statistics well above conventional thresholds when regressing state mobilization rates on agricultural employment shares.

We do not estimate a first-stage regression because we lack individual-level military service records in the census data. The 1950 Census asked about WWII service, but only for a 5% sample, and the MLP does not link this variable for the full panel. Our estimates are therefore reduced-form effects of mobilization *exposure*—the state-level intention to treat—rather than structural effects of military service itself. Converting to per-veteran effects would require both a first-stage relationship and a credible exclusion restriction. Our own pre-trend evidence (Section 5.2) raises concerns about the exclusion restriction, since the mobilization interaction is correlated with pre-existing occupational dynamics that operate through channels other than military service.

B.2 Exclusion Restriction Discussion

The exclusion restriction requires that agricultural employment share affects 1940–1950 occupational changes for draft-eligible cohorts *only* through differential military service exposure. Several potential violations deserve discussion:

1. **War production concentration:** Manufacturing states (with low agricultural share) also received disproportionate war production contracts, potentially creating employment opportunities that affected occupational upgrading independent of military service. We address this by controlling for manufacturing share $\times$ draft eligibility in Column (3) of Table 2.

2. **Agricultural mechanization:** Wartime labor shortages accelerated mechanization on farms, potentially pushing young men out of agriculture and into higher-paying occupations regardless of military service. We control for farm residence in 1940 and examine farm-to-non-farm transitions separately.

3. **Migration patterns:** Men in low-agriculture states may have been more mobile, and wartime disruption may have increased migration differentially across states. Robustness checks adding migration controls (Section 7) show that migration does not drive our results.

4. **Great Depression exposure:** The Depression affected agricultural and industrial states differently, potentially creating differential recovery dynamics that the state fixed effects do not fully absorb. This concern is precisely what the 1930 pre-baseline test addresses: the significant pre-trend indicates that differential Depression recovery is indeed a confound.

B.3 Interpreting the Pre-Trend

The large negative pre-trend (-0.72 with individual controls) for the 1930–1940 period warrants interpretation. During this decade, draft-eligible men were teenagers (ages 8–15 in 1930) transitioning into adulthood (ages 18–25 in 1940). Their "occupational change" largely reflects initial entry into the labor force. The negative coefficient indicates that in high-mobilization (low-agriculture) states, draft-eligible cohorts experienced *less* occupational score growth during their initial labor-force entry than their counterparts in agricultural states.

This likely reflects the Great Depression's differential impact: industrial states experienced deeper unemployment, delayed labor-force entry, and more occupational churning than agricultural states, where farm employment provided a buffer. Young men entering the labor market in states like Michigan or Pennsylvania during the Depression faced worse initial conditions than their counterparts in Mississippi or Arkansas. This created a "head start" for the agricultural-state cohorts that the two-decade design mistakes for a WWII treatment effect.

C. Robustness Appendix

C.1 Complete Robustness Table

Table 5 in the main text summarizes the key robustness checks. Here we provide additional detail on each specification:

Leave-one-out. We estimate 49 separate regressions, each dropping one state from the analysis sample. All 49 estimates are negative and statistically significant. The range of coefficients (-0.27 to -0.23) is narrow, and the median (-0.25) is close to the full-sample estimate (-0.26). The states whose exclusion has the largest effect on the coefficient are those with extreme agricultural shares (Mississippi, at the high end; Rhode Island, at the low end), which is expected given that these states contribute the most to the identifying variation.

Randomization inference. We permute state-level mobilization exposure 1,000 times, each time randomly reassigning the 49 state-level mobilization values across states. For each permutation, we re-estimate the main specification and record the t-statistic. The actual t-statistic exceeds the permuted distribution, providing evidence against the sharp null at conventional levels.

Alternative cohort definitions. The main specification defines draft-eligible as born 1915–1922 (age 18–25 in 1940). We consider two alternatives: (i) broadening to 1910–1922 (including men who were slightly older but still within the draft age range), which includes more men with moderate service probability and attenuates the estimate; (ii) narrowing to 1917–1922 (the youngest and most certainly drafted), which produces a similar negative estimate.

Migration controls. Adding a control for inter-state migration (different state in 1940 and 1950) has minimal effect on the main coefficient. Restricting to non-movers only produces a very similar estimate, confirming that selective migration is not driving the result.

D. Heterogeneity Appendix

D.1 Occupation Quintile Details

The five quintiles of pre-war occupational score in 1940 are defined among men with positive occupational scores (excluding those with zero, which indicates no occupation reported). The quintile boundaries vary slightly depending on the cohort group; we use the pooled distribution across draft-eligible and older control cohorts. Q1 contains the lowest-scoring occupations (farm laborers, domestic servants, laborers NEC), while Q5 contains the highest-scoring occupations (physicians, lawyers, managers, engineers).

The imprecision of the quintile-specific estimates reflects the reduced sample size within each quintile combined with the state-level clustering. With 49 clusters and roughly 20% of the sample in each quintile, the effective number of observations driving the state $\times$ cohort interaction is substantially reduced.

D.2 Education Heterogeneity

We also examine heterogeneity by pre-war educational attainment. Splitting the sample at 12 years of education (high school completion), we find similar negative effects for both groups: −0.22 (SE = 0.09) for men with less than high school education and −0.31 (SE = 0.12) for men with high school or more. The GI Bill's educational channel would predict larger (more positive) effects for less-educated men who had more to gain from additional schooling. The absence of such a gradient is consistent with the career disruption interpretation.

E. Additional Figures and Tables

All main figures and tables are presented in the text. Additional exhibits referenced in the appendices are available in the replication code repository.

F. Standardized Effect Sizes

Table 6: Standardized Effect Sizes for Main Outcomes

Outcome	Specification	$\hat{\beta}$	SD(X)	SD(Y)	SDE	Classification
ΔOccScore (40–50)	Table 2 Col. 2	−0.255	—	14.7	−0.017	Null
ΔLog Wage	Table 2 Col. 4	−0.013	—	0.95	−0.014	Null
Any College (1950)	Table 2 Col. 5	−0.0016	—	0.35	−0.005	Null
ΔOccScore (trend-adj)	Table 5 Panel A	−0.907	—	19.8	−0.046	Null

Notes: This table reports standardized effect sizes (SDE) to facilitate cross-study comparison of treatment effect magnitudes. SDE = $\hat{\beta}/\text{SD}(Y)$, where SD(Y) is the unconditional standard deviation from the summary statistics. The treatment variable (MobExposure × DraftEligible) is an interaction of a continuous standardized instrument with a binary cohort indicator; SD(X) is marked "—" and the SDE uses the coefficient directly divided by SD(Y).

Research question: Does a standard state × cohort mobilization design produce credible estimates of WWII service returns, and what is the reduced-form effect of mobilization exposure on occupational upgrading for draft-eligible men? **Treatment:** Interaction of standardized state-level mobilization exposure (continuous) with draft-eligible cohort indicator (binary). **Data:** IPUMS MLP three-decade panel (1930–1940–1950), 13.4 million men (9.1 million in primary analysis cohorts), 49 states. **Method:** State × cohort reduced-form OLS with state and birth-year fixed effects, standard errors clustered at the state level. **Sample:** Men born 1905–1922, linked across 1930, 1940, and 1950 Censuses.

Classification thresholds: large negative (< −0.10), small negative (−0.10 to −0.05), null (−0.05 to 0.05), small positive (0.05 to 0.10), large positive (> 0.10). A reader unfamiliar with the paper should be able to interpret this table on its own.