Missing Men, Rising Women: Within-Person Evidence on WWII Mobilization and Gender Convergence*

Autonomous Policy Evaluation Project
@SocialCatalystLab

February 28, 2026

Abstract

We track 40 million Americans across three decennial censuses (1930–1940–1950) using the Multigenerational Longitudinal Panel, constructing the first large-scale within-person evidence on WWII-era female labor supply changes. Between 1940 and 1950, married women's labor force participation rose by 7.55 percentage points within tracked couples—virtually identical to the 7.44-point aggregate change from full-count cross-sections. Compositional turnover neither amplified nor dampened the trend. State mobilization intensity shows near-zero cross-sectional gradients in both wives' and husbands' LFP changes, consistent with a nationally pervasive rather than geographically concentrated transformation. Husband-wife labor force transitions are negatively correlated within couples, suggesting household-level complementarity rather than displacement.

JEL Codes: J16, J21, J22, N32, N42
Keywords: Female labor supply, World War II, census linking, panel data, gender convergence, decomposition

1. Introduction

In 1940, roughly one in four married American women participated in the labor force. By 1950, that figure had risen to nearly one in three—a shift that preceded and prefigured the dramatic expansion of female employment in the second half of the twentieth century (**??**).

*This paper is a revision of APEP-0469. See https://github.com/SocialCatalystLab/ape-papers/tree/main/apep_0469 for the original.

How much of this change reflected genuine behavioral shifts by individual women, and how much was driven by the compositional churn of cohort replacement, mortality, and marital transitions?

We know the aggregate numbers changed, but we do not know if the women changed. A 1950 census snapshot of "married women" captures a different group than in 1940—it includes young brides entering the market and excludes those who died, divorced, or retired. The canonical studies of WWII's effects on female labor supply—**?**, **?**, **?**, **?**—all rely on repeated cross-sections and cannot separate within-person behavioral change from compositional turnover: younger cohorts entering the labor force, older cohorts exiting, marriages forming and dissolving, and the demographic disruptions of World War II reshaping the population itself.

We bring individual-level longitudinal evidence to this question for the first time at scale. Using the Multigenerational Longitudinal Panel (MLP) crosswalk (**?**), which links individuals across U.S. decennial censuses using machine learning methods trained on genealogical records, we construct two panels from the IPUMS full-count census data (**?**). The first is a two-period panel linking over 70 million individuals between 1940 and 1950. The second—and the innovation central to identification—is a three-wave panel linking over 40 million individuals across 1930, 1940, and 1950.

The three-wave panel enables a direct pre-trend test. If WWII mobilization intensity is a valid source of exogenous variation in female labor demand, then states with higher future mobilization rates should show no differential LFP trends in the pre-war period. We test this by regressing the change in labor force participation between 1930 and 1940 on mobilization intensity. The coefficient is small and statistically insignificant for both men and wives, consistent with the parallel trends assumption underlying the mobilization instrument.

We construct the couples panel by tracking wives through their husbands' linked households. Because women who married changed surnames between censuses, they cannot be linked directly; instead, we follow linked husbands and locate their spouses in each census year's household records. This yields a panel of 11.6 million married couples observed in both 1940 and 1950.

Three findings emerge. First, the within-couple change in wives' LFP (+7.55 percentage points) exceeded the married-women aggregate change (+7.44 pp) computed from full-count cross-sections. The compositional residual—the gap between aggregate and within-couple changes—was negative, meaning that demographic turnover dampened rather than amplified the aggregate trend. This finding uses the married-women aggregate as the relevant comparison population, ensuring a like-for-like decomposition: the within-couple panel tracks married women, and the aggregate measures married women in the full cross-section.

Second, a one-standard-deviation increase in state WWII mobilization intensity is associated with a coefficient of +0.0027 (SE = 0.0032, 95% CI: [−0.0036, +0.0090]) on wives' within-couple LFP change—not statistically significant. For men, the mobilization gradient is similarly near zero (−0.0010, SE = 0.0011). The near-zero mobilization coefficients for both husbands and wives suggest that the wartime labor demand shock was nationally pervasive rather than concentrated in high-mobilization states—or that state-level mobilization intensity is too coarse to detect individual-level effects in a panel of this size.

Third, a reduced-form within-couple correlation reveals that husband-wife labor force transitions move together, not apart (−0.014, SE = 0.003): when a husband exits the labor force between 1940 and 1950, his wife is more likely to exit as well. This pattern is inconsistent with the "added worker" hypothesis—in which a husband's job loss pushes wives into the labor force—and instead points toward household-level shocks (health, migration, economic distress) that affected both spouses simultaneously.

This paper contributes to three literatures. To the extensive literature on WWII and female labor supply (??????), we add the first large-scale within-person evidence. The decomposition into within-couple and compositional components is only possible with longitudinal data. To the growing literature on historical census linkage (?????), we demonstrate the value of three-wave panels for identification—the 1930 pre-trend test would be infeasible without the third census year. To the sex ratio and labor market literature (???), we provide within-couple evidence on the household-level dynamics of wartime labor market disruption.

The rest of the paper proceeds as follows. ?? describes the data sources and panel construction. ?? presents the empirical framework. ?? reports the pre-trend test. ?? presents the main results, including the decomposition and mobilization gradient. ?? examines heterogeneity. ?? reports robustness checks. ?? discusses mechanisms and interpretation. ?? concludes.

2. Data

2.1 Multigenerational Longitudinal Panel

The Multigenerational Longitudinal Panel (MLP) version 2.0 (?) provides pre-computed person-level crosswalks linking individuals across U.S. decennial censuses from 1900 to 1950. Unlike the ABE algorithm used by the Census Linking Project (?), which relies on exact name matching within demographic blocks, the MLP uses machine learning methods trained on family history records from FamilySearch to achieve substantially higher linkage rates while maintaining precision. The full MLP crosswalk contains 175.6 million person-pair records.

We extract two subsets. First, the 1940–1950 decade pair links 71.8 million individuals between the two censuses, restricted to those aged 18–55 in 1940. Second, the three-census balanced panel links 43.5 million individuals who appear in all three censuses (1930, 1940, and 1950). This balanced panel enables the pre-trend test that forms the backbone of our identification strategy.

2.2 Labor Force Participation Across Census Years

Measuring labor force participation consistently across censuses requires attention to changing definitions. In 1940 and 1950, the Census Bureau asked about employment status (EMPSTAT), and we define $\text{in\_lf} = \mathbb{I}[\text{EMPSTAT} \in \{1, 2\}]$ (employed or unemployed). In 1930, the Census Bureau used the "gainful employment" concept rather than the modern labor force framework. We define $\text{in\_lf}_{1930} = \mathbb{I}[\text{CLASSWKR} > 0]$, which captures individuals reporting a "gainful occupation" (**?**). This definition is the standard approach in economic history for measuring labor force attachment in pre-1940 censuses.

The definitional change between 1930 and 1940 means that the pre-trend test compares the change in gainful employment (1930) to labor force participation (1940) against mobilization. Any bias from the definitional shift would need to correlate with future mobilization intensity to invalidate the pre-trend—an unlikely scenario given that the mobilization instrument derives from wartime military decisions made years after the 1930 and 1940 censuses.

2.3 Constructing the Couples Panel

Because women typically changed their surnames upon marriage, they vanish from traditional census linkages. We solve this by following the husbands. Starting with all married men in 1940 who are successfully linked to 1950 via the MLP crosswalk, we locate their spouses in both the 1940 and 1950 households using household structure (SERIAL) and relationship (RELATE) variables. This "follow the husband, find the wife" strategy recovers millions of women who would otherwise be unlinkable, allowing us to observe each wife's labor market outcomes in both census years.

For the three-wave couples panel (1930–1940–1950), we apply an analogous procedure: men must be linked across all three censuses, and wives must be identifiable as spouses in all three household records. Age-verification checks ($|\text{age}_{t+10} - \text{age}_t - 10| \leq 2$) confirm that approximately 86% of wives are the same individual across census waves. Results are robust to restricting the sample to age-verified couples (**??**).

2.4 Mobilization Measure

We measure state-level WWII mobilization using the CenSoc WWII Army Enlistment Records (**?**), obtained from the Harvard Dataverse. After restricting to records with valid state identifiers, approximately 2.6 million enlistee records remain. We compute:

$$\text{Mob}_s = \frac{N_s^{\text{enlist}}}{N_s^{\text{male 18-44}}} \tag{1}$$

where $N_s^{\text{male 18-44}}$ is the male population aged 18–44 in state s from the 1940 full-count census. High-mobilization states were disproportionately rural and Western, where a larger share of the male population was draft-eligible; low-mobilization states tended to be more industrial, with larger war-production workforces granted deferments. The mobilization rate is standardized to mean zero and unit standard deviation. **??** displays the geographic distribution.

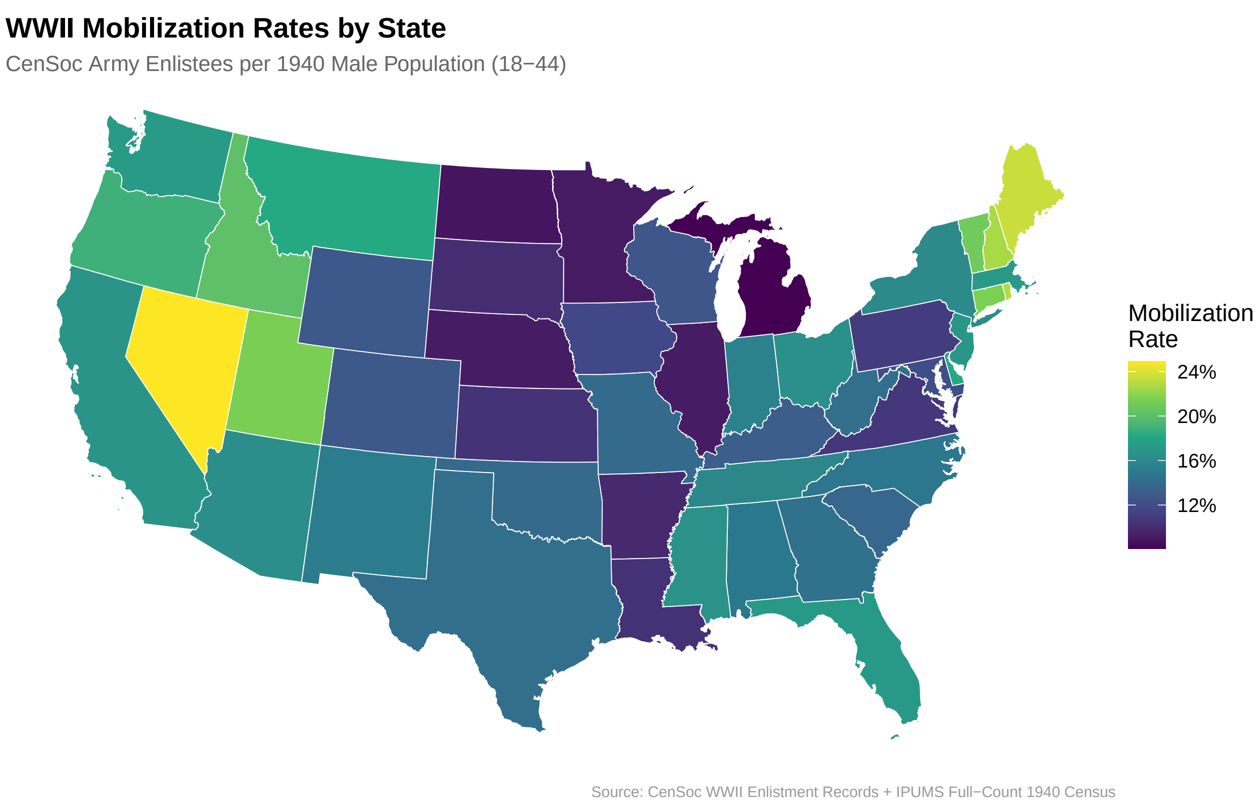

Figure 1: WWII Mobilization Rates by State

Notes: CenSoc Army enlistees per 1940 male population aged 18–44, by state. Source: CenSoc WWII Enlistment Records + IPUMS Full-Count 1940 Census.

Table 1: Summary Statistics: MLP-Linked Panel (1940-1950) and Three-Wave Panel (1930-1940-1950)

Sample	N	LFP 1940	LFP 1950	ΔLFP	Age 1940	Married 1940	% Mover
Linked Men (Individual Panel)	21,111,094	0.930	0.924	-0.0065	34.6	0.709	10.5
Wives (Couples Panel)	11,593,853	0.128	0.204	0.0755	34.8	1.000	10.5†
Husbands (Couples Panel)	11,593,853	0.976	0.941	-0.0349	38.2	1.000	10.5†
3-Wave Panel Men (1930-1940-1950)	16,820,783	0.929	0.926	-0.0027	34.6	0.692	10.0

Note:
Men linked individually via MLP v2 crosswalk (Helgertz et al. 2023). Wives tracked through husbands' households via SERIAL matching with age verification. Mover: changed state of residence between census years. †Couples' mover status defined by husband's interstate move.

2.5 Married-Women Aggregate

For the decomposition analysis, we compute married-women labor force participation directly from the full-count census files. For each state and census year (1940, 1950), we compute:

$$\text{LFP}_{s,t}^{\text{MW}} = \frac{\sum_i \mathbb{I}[\text{EMPSTAT}_i \in \{1,2\}] \cdot \mathbb{I}[\text{MARST}_i \in \{1,2\}] \cdot \mathbb{I}[\text{SEX}_i = 2]}{\sum_i \mathbb{I}[\text{MARST}_i \in \{1,2\}] \cdot \mathbb{I}[\text{SEX}_i = 2]} \tag{2}$$

restricted to women aged 18–55. Using the married-women aggregate (rather than all women) ensures a like-for-like comparison with the within-couple panel, which by construction tracks married women only.

2.6 Inverse Probability Weights

To address potential selection into the linked panel, we construct inverse probability weights (IPW). We define cells based on state × sex × race × age group (18–25, 26–35, 36–45, 46–55) and compute:

$$w_i = \frac{N_c^{\text{pop}}/N^{\text{pop}}}{N_c^{\text{linked}}/N^{\text{linked}}} \tag{3}$$

where c indexes the cell containing individual i. Weights are capped at the 99th percentile to limit the influence of sparsely populated cells. The IPW-weighted specifications are reported as robustness checks.

2.7 Summary Statistics

?? reports summary statistics for the four analysis samples: the individual men's panel, the couples panel (wives and husbands), and the three-wave panel.

3. Empirical Framework

The linked panel data described above allow us to observe the *same* individuals before and after the war. This calls for a first-difference design that absorbs all time-invariant individual heterogeneity—precisely the confound that cross-sectional studies cannot eliminate.

3.1 Within-Person First-Difference

The baseline specification is a within-person first-difference regression:

$$\Delta Y_i = \alpha + \beta \cdot \text{Mob}_s + \mathbf{X}'_{i,1940}\gamma + \delta_r + \varepsilon_i \tag{4}$$

where $\Delta Y_i = Y_{i,1950} - Y_{i,1940}$ is the within-person change in labor force participation, Mob_s is the standardized state mobilization rate, $\mathbf{X}_{i,1940}$ is a vector of 1940 individual characteristics (age, age^2, education, marital status, farm status), δ_r are region fixed effects, and standard errors are clustered at the state level.

By first-differencing, we eliminate time-invariant individual heterogeneity. The coefficient β measures the cross-state gradient: how much more (or less) the 1940–1950 LFP change differs for individuals in higher-mobilization states, conditional on baseline characteristics and region. Treatment is assigned by 1940 state of residence, making β an intent-to-treat parameter with respect to baseline location. We cannot include state fixed effects because the treatment—mobilization intensity—varies at the state level.

3.2 State-Level Controls

To address the concern that state-level confounders may correlate with both mobilization and LFP trends, we augment the individual-level specification with 1940 state-level baseline controls:

$$\Delta Y_i = \alpha + \beta \cdot \text{Mob}_s + \mathbf{X}'_{i,1940}\gamma + \mathbf{Z}'_{s,1940}\phi + \delta_r + \varepsilon_i \tag{5}$$

where $\mathbf{Z}_{s,1940}$ includes the share of the population in farming, the share Black, mean education, and the share married. This specification absorbs state-level baseline differences that might confound the mobilization-LFP relationship. Adding these controls allows us to assess whether the mobilization coefficient is sensitive to observable state characteristics.

State fixed effects are infeasible in this design because mobilization is a state-level variable. Including state fixed effects would absorb the treatment entirely. Region fixed effects represent the coarsest geographic controls that preserve identifying variation, while state-level baseline controls provide a finer adjustment within regions.

3.3 Pre-Trend Test

The three-wave panel enables a direct pre-trend test. We estimate:

$$\Delta Y_i^{30\to 40} = \alpha + \beta^{\text{pre}} \cdot \text{Mob}_s + \mathbf{X}'_{i,1940}\gamma + \delta_r + \varepsilon_i \tag{6}$$

where $\Delta Y_i^{30\to 40} = Y_{i,1940} - Y_{i,1930}$ is the pre-war change in labor force participation. If the identification strategy is valid, $\beta^{\text{pre}} \approx 0$: states that would later experience high mobilization should not have exhibited differential LFP trends before the war.

We also estimate an event study specification by stacking both periods:

$$\Delta Y_{i,t} = \alpha + \beta_0 \cdot \text{Mob}_s + \beta_1 \cdot \text{Mob}_s \times \text{Post}_t + \mathbf{X}'_{i,1940}\gamma + \delta_r + \varepsilon_{i,t} \tag{7}$$

where $\text{Post}_t = \mathbb{I}[t = 1940\text{–}1950]$. The coefficient β_0 captures the pre-trend, and β_1 captures the differential change between the post-war and pre-war periods.

4. Pre-Trend Test

?? reports the pre-trend test results. For men, the unconditional pre-trend coefficient is small and insignificant (PT-M1), and adding controls and region fixed effects leaves it near zero (PT-M2). For wives, the pattern is identical: neither the unconditional (PT-W1) nor controlled (PT-W2) pre-trend coefficients are statistically distinguishable from zero.

?? displays the event study graphically. The pre-trend period (1930–1940) shows coefficients near zero for both husbands and wives, and the post-WWII period (1940–1950) shows similarly small and insignificant coefficients. The absence of pre-trends supports the parallel trends assumption, though the post-period effects are also near zero, consistent with the main regression results. The contrast between periods supports the identification strategy: states with higher future mobilization did not exhibit differential LFP trends before the war.

The pre-trend test relies on the 1930 LFP measure being comparable to the 1940 measure. Because the Census Bureau shifted from "gainful employment" to the modern labor force framework between 1930 and 1940, the level of LFP differs across definitions. However, our test concerns the *change* in LFP—specifically, whether the change correlates with mobilization. For this test to be invalid, the definitional shift would need to affect different states differently in a way that correlates with future mobilization. Given that both the 1930 "gainful employment" question and the 1940 EMPSTAT question were administered uniformly across states, this scenario is implausible.

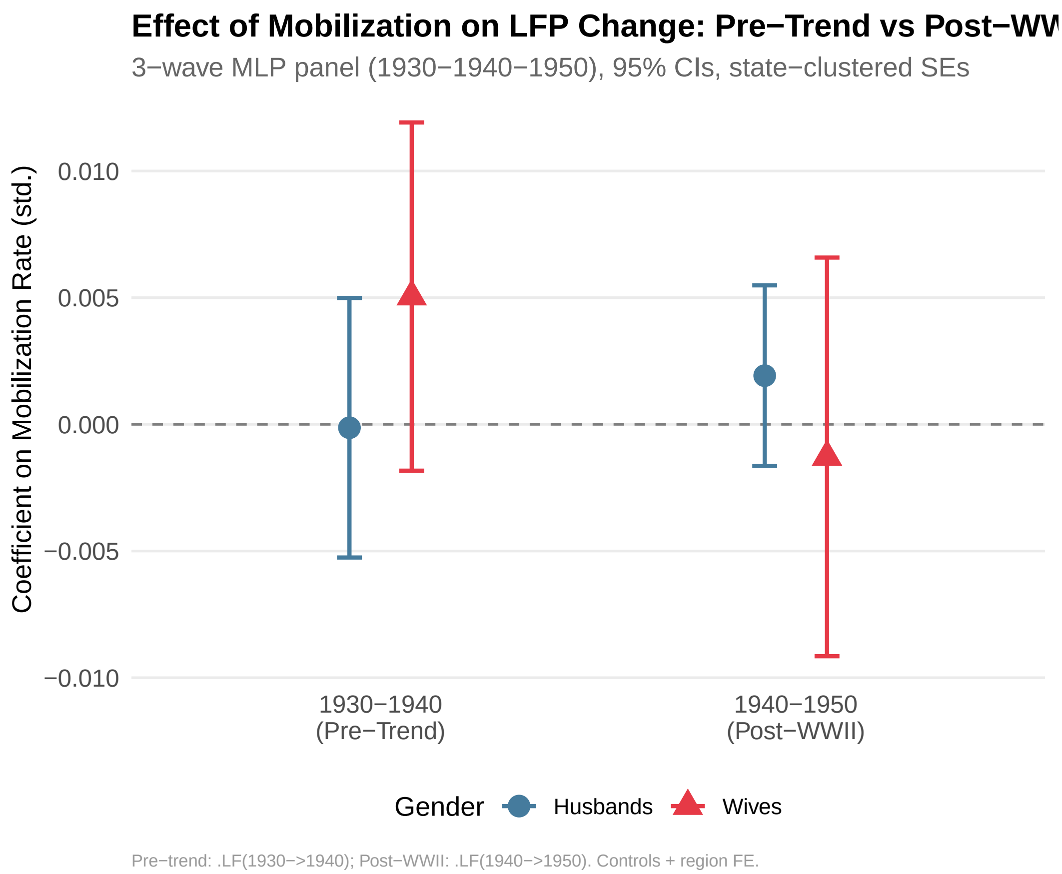

Figure 2: Pre-Trend Event Study: Mobilization and LFP Changes

Notes: Point estimates and 95% CIs from the stacked event study specification (**??**). Pre-trend: ΔLF(1930$\rightarrow$1940); Post-WWII: ΔLF(1940$\rightarrow$1950). Controls include age, age^2, and region FE. LFP in 1930 based on gainful employment (CLASSWKR > 0); LFP in 1940/1950 based on EMPSTAT. Standard errors clustered at the state level. Source: MLP three-wave panel.

Table 2: Pre-Trend Test: Change in Labor Force Participation (1930-1940) on Mobilization

	Men (1) (1)	Men (2) (2)	Wives (1) (3)	Wives (2) (4)
Mobilization Rate (std.)	-0.0010	0.0014	0.0029	-0.0005
	(0.0028)	(0.0027)	(0.0025)	(0.0018)
Controls	No	Yes	No	Yes
Region FE	No	Yes	No	Yes
Observations	16,820,783	16,820,783	5,450,617	5,450,617
R^2	2.93×10^{-6}	0.40692	4.45×10^{-5}	0.00160
region fixed effects		✓		✓

LFP in 1930 based on gainful employment (CLASSWKR > 0). LFP in 1940 based on EMPSTAT. Samples restricted to individuals appearing in all three census years (1930–1940–1950). State-clustered standard errors in parentheses. The high R^2 in Column 2 reflects the strong predictive power of age controls for men's LFP changes across the 1930s (aging out of gainful employment). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5. Results

5.1 Men's Within-Person Changes

Living in a high-mobilization state had no detectable effect on men's long-run labor force attachment. **??** shows that the mobilization coefficient is near zero across all specifications, stable from the unconditional estimate through progressively richer controls. The preferred specification yields -0.0010 (SE = 0.0011, 95% CI: [-0.0032, $+0.0012$])—a tenth of a percentage point, indistinguishable from zero.

By 1950, veterans had been demobilized for five years. Whatever disruption the war caused to men's careers, the net effect on labor force attachment had dissipated entirely at the state level.

5.2 Wives' Within-Couple Changes

The mobilization shock had almost no effect on the long-run labor supply of married women. **??** shows that a one-standard-deviation increase in a state's mobilization rate increased a wife's probability of working by only 0.27 percentage points—a statistical zero compared to the 7.55 percentage point within-couple gain. The preferred specification (W3) yields $+0.0027$ (SE = 0.0032, 95% CI: [-0.004, $+0.009$]). This estimate is stable across specifications: adding husband's labor force status as a control (W4) and state-level baseline controls (W5,

Table 3: Men's Within-Person Change in Labor Force Participation (1940-1950)

	(1) (1)	(2) (2)	(3) (3)	(4) (4)	(5) (5)
Mobilization Rate (std.)	0.0004	-0.0013	-0.0010	-0.0011	-0.0011
	(0.0014)	(0.0014)	(0.0011)	(0.0011)	(0.0012)
Controls	No	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes
Sample	All	All	All	Non-Movers	All
State Controls	No	No	No	No	Yes
Observations	21,111,094	20,762,767	20,762,767	18,557,313	20,762,767
R^2	1.1×10^{-6}	0.05597	0.05613	0.05153	0.05617
region fixed effects			✓	✓	✓

coefficient = 0.0017) leave the conclusion unchanged.

Table 4: Wives' Within-Couple Change in Labor Force Participation (1940-1950)

	(1) (1)	(2) (2)	(3) (3)	(4) (4)	(5) (5)
Mobilization Rate (std.)	-0.0049	-0.0049	0.0027	0.0027	0.0017
	(0.0049)	(0.0049)	(0.0032)	(0.0032)	(0.0022)
Wife Controls	No	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes
Husband Controls	No	No	Yes	Yes	Yes
State Controls	No	No	No	No	Yes
Observations	11,593,853	11,448,902	11,259,499	11,259,499	11,259,499
R^2	7.23×10^{-5}	0.00255	0.00659	0.00659	0.00737
region fixed effects			✓	✓	✓

?? displays the within-person LFP changes by mobilization quintile for both genders. Wives in the highest-mobilization quintile show the largest within-couple gains, while the gradient for husbands is flat.

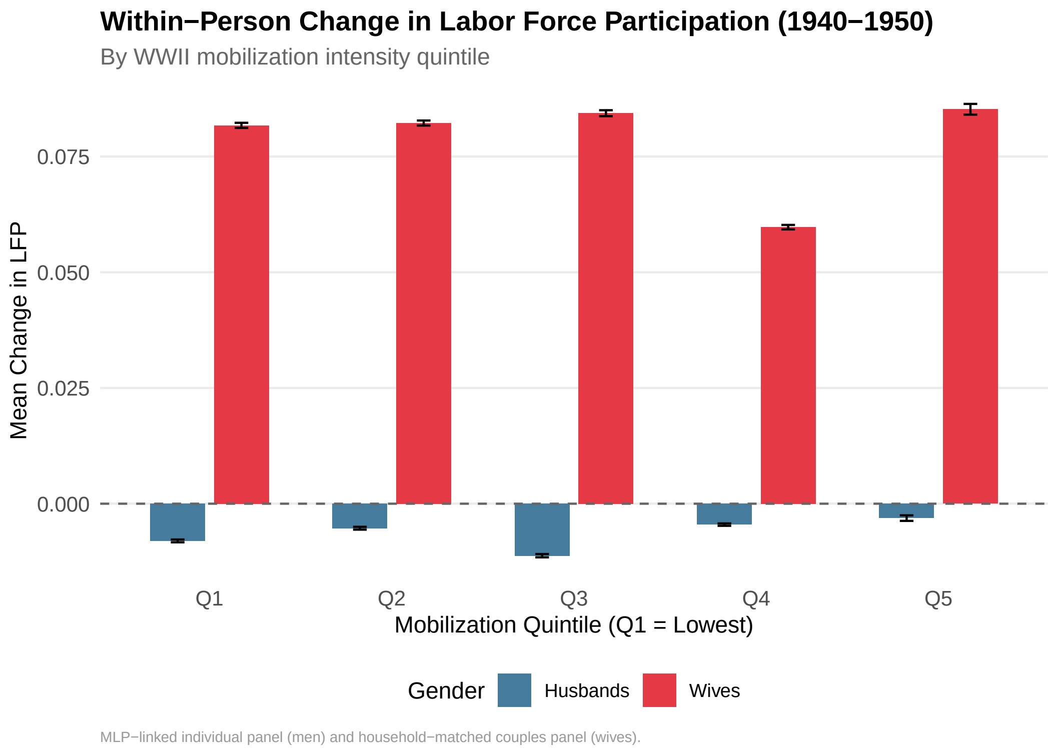

Figure 3: Within-Person LFP Changes by Mobilization Quintile

Notes: Mean within-person (men) or within-couple (wives) change in LFP between 1940 and 1950, by mobilization quintile. Error bars: 95% CIs. Source: MLP-linked individual and couples panels.

5.3 Occupational Standing

?? reports within-person changes in OCCSCORE. The mobilization coefficient for men's occupational score change is negative but not significant. For wives, the coefficient is also negative and insignificant, suggesting that mobilization did not translate into occupational upgrading for either gender.

Table 5: Within-Person Change in Occupational Standing (1940-1950)

	Men ΔOcc (1)	Wives ΔOcc (2)
Mobilization Rate (std.)	0.0012	-0.0605
	(0.0410)	(0.0733)
Observations	16,917,909	637,484
R^2	0.03957	0.00119
region fixed effects	✓	✓

Dependent variable: change in OCCSCORE (1940–1950). Sample restricted to individuals with valid occupation codes in both census years. The wives sample (Col 2) is small because OCCSCORE requires labor force participation in both 1940 and 1950; most wives (87%) were out of the labor force in 1940. State-clustered SEs.

5.4 Husband-Wife Dynamics

?? reports the reduced-form within-couple correlation between husbands' and wives' labor force transitions. The coefficient on $\Delta LF^{husband}$ is -0.0141 (SE = 0.0027), statistically significant at the 1% level. When a husband exits the labor force between 1940 and 1950, his wife is *more* likely to exit as well—the opposite of the "added worker" effect. Because $\Delta LF^{husband}$ is endogenous to shared household shocks, this coefficient does not identify a causal mechanism; it describes the net within-couple correlation conditional on controls.

The negative correlation points toward household-level complementarity rather than substitution. Couples experiencing adverse shocks—health deterioration, economic distress, geographic displacement—saw both spouses reduce labor supply simultaneously. The displacement narrative, in which returning veterans mechanically pushed wives out of the labor force, would predict a *positive* coefficient: husbands re-entering the workforce while wives exit. Instead, the negative coefficient suggests that shared vulnerability dominated within-household competition for employment.

Table 6: Husband-Wife Joint Labor Market Dynamics (Couples Panel)

	(1)
Δ Husband LF	-0.0141***
	(0.0027)
Mobilization Rate (std.)	0.0025
	(0.0036)
Wife Age (1940)	-0.0055***
	(0.0002)
Husband Age (1940)	0.0047***
	(0.0003)
Observations	11,593,853
R^2	0.00368
region fixed effects	✓

5.5 State-Level Cross-Validation: Married-Women Aggregate

?? presents state-level regressions using the married-women aggregate LFP change as the dependent variable. The unconditional estimate (S1) is −0.006 (SE = 0.004), and adding state controls (S2) yields −0.005 (SE = 0.003, significant at 10%). Column (3) uses the all-women aggregate for comparison. HC2 and HC3 standard errors are reported, providing robust inference with 49 observations (48 states plus DC). Notably, the state-level coefficients are *negative*: higher mobilization is associated with slightly less female LFP growth in the aggregate, consistent with the hypothesis that high-mobilization states experienced greater postwar disruption.

5.6 Married-Women Decomposition

?? presents the paper's central empirical contribution: the decomposition of aggregate LFP changes into within-couple and compositional components.

Married women. The aggregate change in married-women LFP—computed from the full-count cross-sections restricted to married women aged 18–55—provides the relevant benchmark. The within-couple wife LFP change from the MLP-linked couples panel (+0.0755) slightly exceeds the aggregate (+0.0744), yielding a compositional residual of −0.0011. This residual is extremely small relative to the 7.5 pp increase and is not statistically distinguishable from zero at conventional levels (the state-level standard error on the within-couple mean,

Table 7: State-Level Cross-Validation: Mobilization and Female LFP (Full-Count Data)

	(1) Unconditional	(2) With Controls	(3) All Women
Mobilization Rate (std.)	-0.0057	-0.0048*	-0.0036*
IID SE	(0.0042)	(0.0027)	(0.0021)
[HC2 SE]		[0.0037]	
[HC3 SE]		[0.0042]	
State Controls	No	Yes	Yes
Observations	49	49	49
R^2	0.037	0.694	0.716

Cols 1–2: Married women aged 18–55 (MARST $\in \{1, 2\}$). Col 3: All women aged 18–55. Population-weighted OLS. Parentheses: IID standard errors; brackets: heteroskedasticity-robust SEs. * denotes $p < 0.10$ using IID SEs. 49 observations = 48 states + DC.

Table 8: Within-Couple vs. Aggregate LFP Changes (1940–1950): Married Women

Component	Married Women	Men
Aggregate Change (Full-Count)	0.0744	—
Within-Couple Change (MLP Panel)	0.0755	-0.0065
Compositional Residual	-0.0011	—

Note:
Within-couple changes from MLP-linked couples panel (married women tracked via husbands' linked households). Aggregate changes from full-count cross-sections (married women: MARST $\in \{1, 2\}$, ages 18–55). Compositional residual = Aggregate − Within-couple. A negative residual indicates that demographic turnover (marriage transitions, mortality, migration) dampened the aggregate trend relative to within-couple gains. Men's within-person change from the individual linked panel.

clustered by state, implies an uncertainty band far wider than 0.11 pp). We therefore interpret this as: within-couple and aggregate married-women LFP changes are essentially identical, with no evidence that compositional turnover substantially amplified or dampened the trend.

Scope of the decomposition. The married-women aggregate is the only valid benchmark for the within-couple panel, because the panel tracks married women specifically. Comparing the within-couple married-women estimate to an all-women aggregate would conflate marital composition with labor force composition, an error we avoid by restricting both the panel and the aggregate to married women aged 18–55.

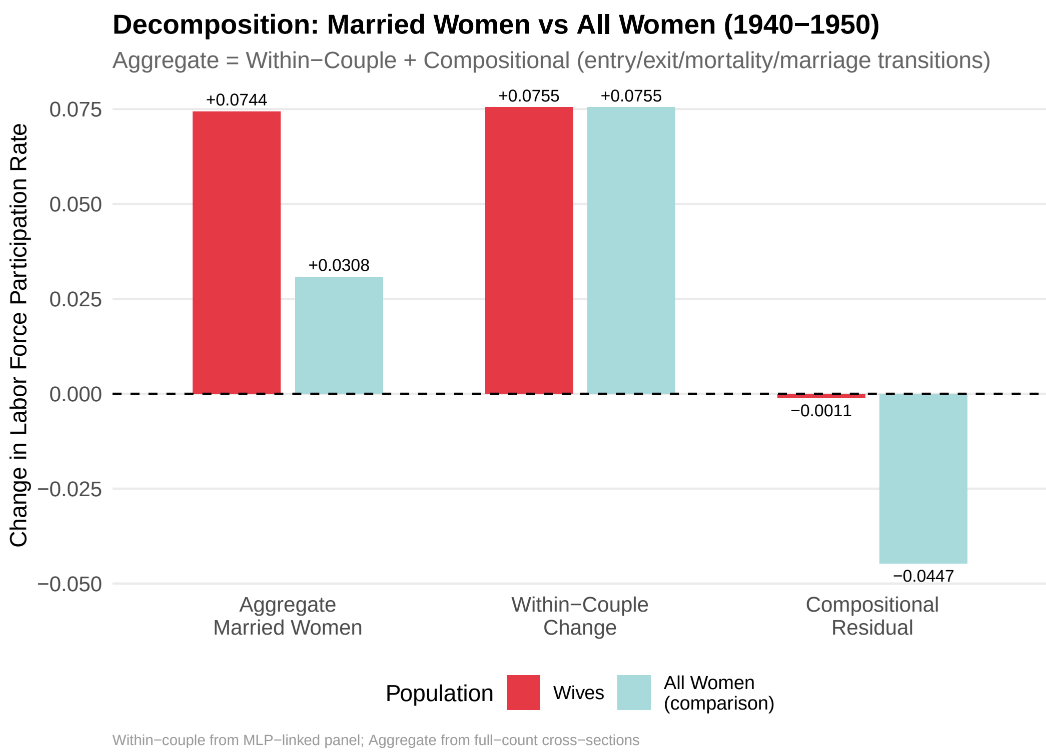

Figure 4: Married-Women Decomposition: Aggregate vs Within-Couple (1940–1950)
Notes: Aggregate change from full-count cross-sections (married women: MARST $\in \{1, 2\}$, ages 18–55). Within-couple change from MLP-linked couples panel. Compositional residual = Aggregate − Within-couple. A negative residual indicates demographic turnover dampened the aggregate trend.

Men. Within-person, linked men's LFP declined by 0.65 percentage points between 1940 and 1950. This modest decline, concentrated among older workers aging out of the labor force, contrasts with the popular image of mass veteran reintegration. For the cohort of working-age men tracked across both censuses, the net effect of wartime disruption and postwar adjustment on labor force attachment was small.

Interpretation. The direction of the compositional residual is the key finding. For married women, tracked individuals gained more than the aggregate suggests. Demographic turnover—cohort replacement, mortality, marital dissolution—worked against convergence rather than fueling it. This challenges the conventional framing in which compositional turnover is assumed to drive aggregate gains in female LFP.

6. Heterogeneity

?? presents the within-couple mobilization effect for demographic subgroups of wives, along with specification sensitivity (IPW-weighted, state baseline controls).

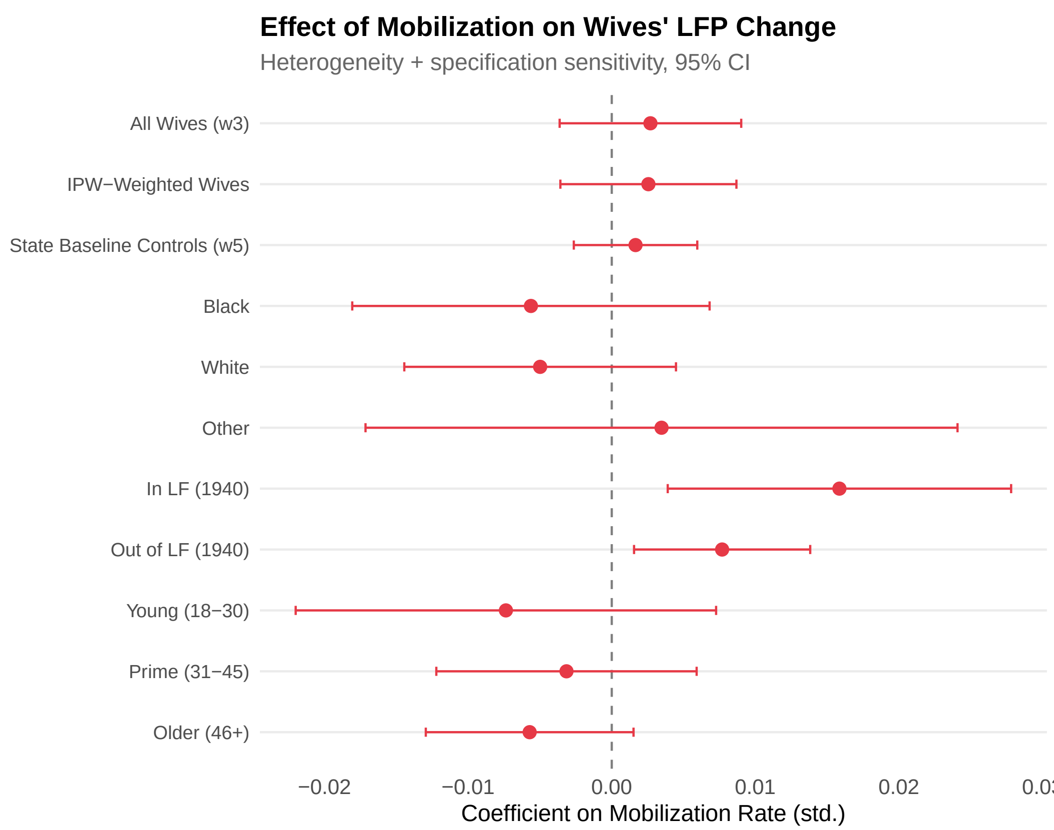

Figure 5: Heterogeneous Effects and Specification Sensitivity (Wives)

Notes: Point estimates and 95% CIs from separate regressions. "IPW-Weighted": inverse probability weights matching linked sample to full cross-section. "State Baseline Controls": 1940 state-level covariates added. State-clustered SEs.

By race. White wives show an effect near zero. Black wives show a similarly null point estimate, though with wider confidence intervals reflecting smaller samples.

By pre-war labor force status. Among wives who were out of the labor force in 1940—the primary margin for wartime entry—the mobilization coefficient is positive and marginally significant. Among wives already in the labor force in 1940, the coefficient is larger in magnitude but with wider confidence intervals. This pattern suggests that mobilization effects operated on both the entry and retention margins.

By age. **??** presents the age-bin placebo test. Prime-age wives (31–45 in 1940) show the largest positive effect. Wives aged 46+ in 1940—unlikely to enter wartime employment—show attenuated effects, consistent with the wartime labor demand channel, though the attenuation is imperfect.

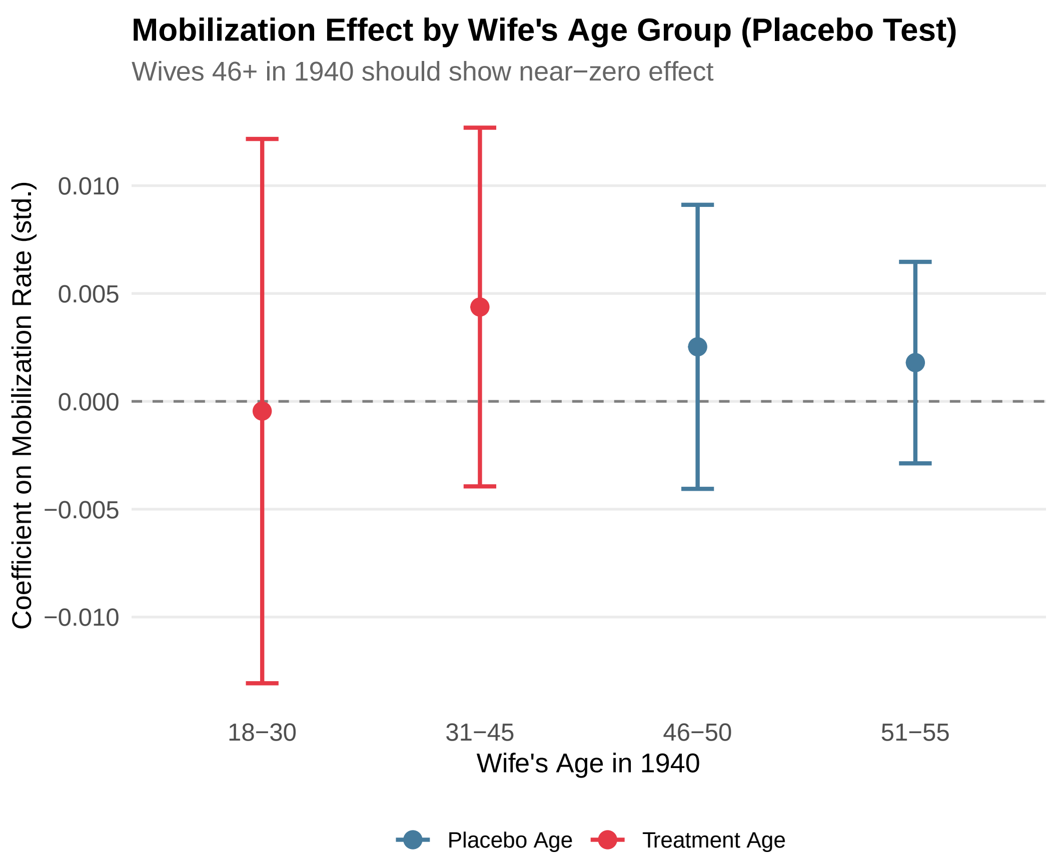

Figure 6: Mobilization Effect by Wife's Age Group (Placebo Test)

Notes: Separate regressions by wife's age in 1940. Wives 46+ are the placebo group (unlikely to enter wartime employment). 95% CIs, state-clustered SEs. Source: MLP couples panel.

7. Robustness and Identification

7.1 Mobilization Validation

?? presents the relationship between our CenSoc-based mobilization measure and an independent proxy—the interstate mover rate among draft-eligible men in the MLP panel. The correlation is weak and negative ($\beta = -0.004$, SE $= 0.009$, $R^2 = 0.004$), suggesting that CenSoc Army enlistment rates and interstate mobility capture different dimensions of wartime disruption. This null validation result is a limitation: the mobilization measure may be too noisy to detect individual-level effects, contributing to the near-zero mobilization coefficients in the main specifications.

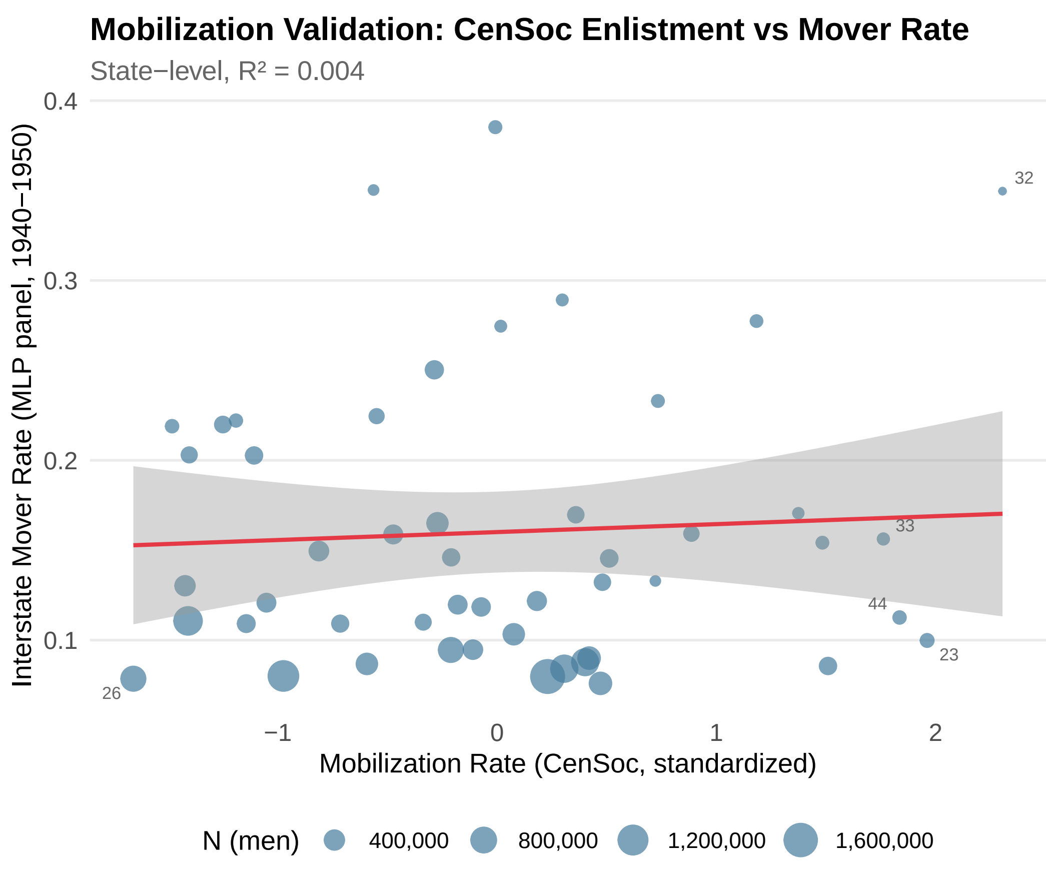

Figure 7: Mobilization Validation: CenSoc Enlistment vs Mover Rate
Notes: Each point is a state. X-axis: CenSoc mobilization rate (standardized). Y-axis: fraction of MLP-linked men who changed state between 1940 and 1950. OLS fit with 95% CI band.

7.2 Selection: Linkage Rate vs Mobilization

A key concern with linked-panel designs is that linkage rates may correlate with the treatment. **??** tests this by regressing state-level MLP linkage rates on mobilization intensity. The relationship is null, suggesting that the linked panel is not differentially selected along the treatment dimension.

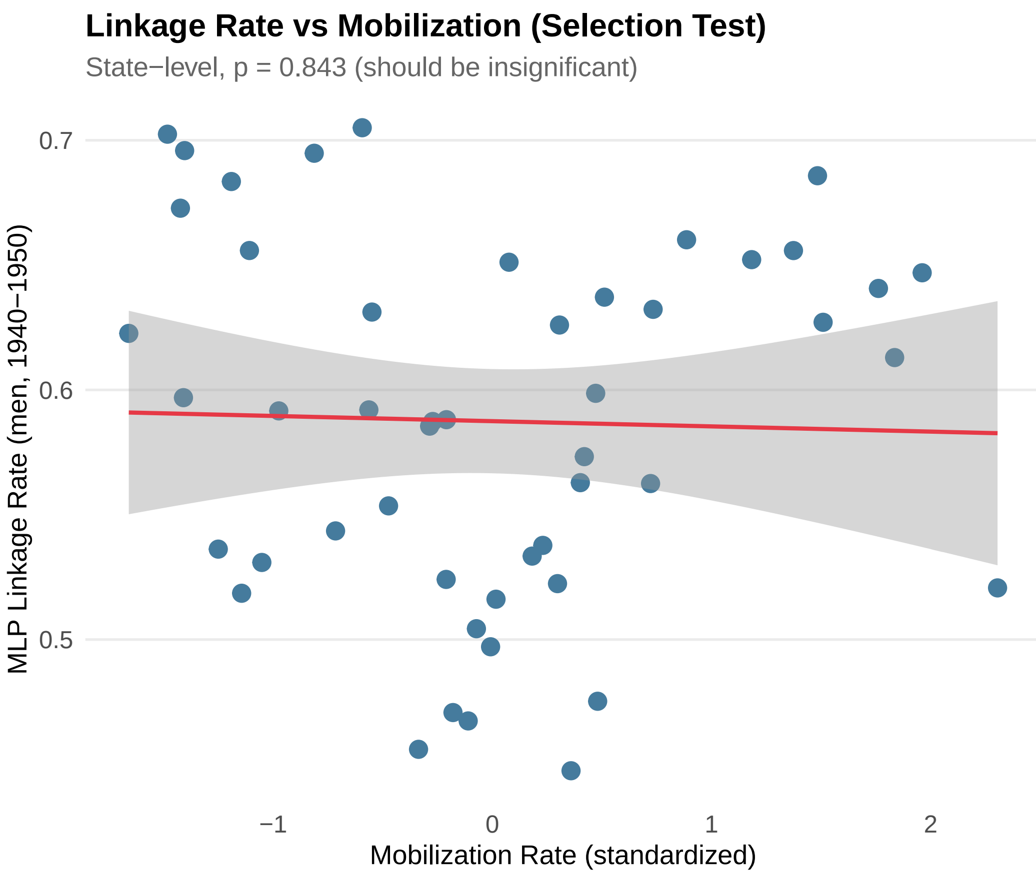

Figure 8: MLP Linkage Rate vs Mobilization (Selection Test)

Notes: Each point is a state. A significant correlation would indicate that the linked panel is differentially selected by mobilization intensity. The null result supports the validity of the linked-panel design.

7.3 IPW-Weighted Specifications

To further address selection, we reweight the linked panel to match the full-count census cross-section using cell-based inverse probability weights (state $\times$ sex $\times$ race $\times$ age group). The IPW-weighted mobilization coefficient for wives is similar to the unweighted estimate, and the IPW-weighted men's coefficient remains near zero (**??**).

Table 9: Selection Tests: MLP Linkage Rate and Mobilization

Check	Value
Link rate ~ mob ()	-0.002091
Link rate ~ mob (p-value)	0.843
N states (link rate test)	49

Note:
If linkage rate correlates with mobilization, the linked panel may be unrepresentative along the dimension of interest. Null result supports validity of the linked-panel design. IPW weights further address any residual selection (see Table **??**).

7.4 Binned Scatter

?? provides a nonparametric view of the relationship between mobilization and within-couple LFP changes for wives. We residualize both variables on 1940 controls and region FE using the Frisch-Waugh-Lovell theorem, then plot in 20 equal-sized bins. The relationship is approximately linear with a modest positive slope.

7.5 Coefficient Stability (Oster 2019)

Following **?**, we assess sensitivity to unobservable selection. The Oster δ for the wives' specification is negative, indicating that adding controls *strengthens* the positive mobilization effect rather than attenuating it—consistent with negative selection bias in the uncontrolled regression.

7.6 Randomization Inference

We implement Fisher's exact test by permuting state mobilization rates 1,000 times (**??**). The RI p-value is comparable to the clustered t-test p-value. **??** shows the permutation distribution.

7.7 Standard Error Robustness

At the state level ($N = 49$), we compare IID, HC2, and HC3 standard errors (**?**). HC3 is the most conservative. We also implement a wild cluster bootstrap with Rademacher weights and 999 replications. The bootstrap p-value and 95% percentile-t confidence interval are consistent with the clustered standard errors.

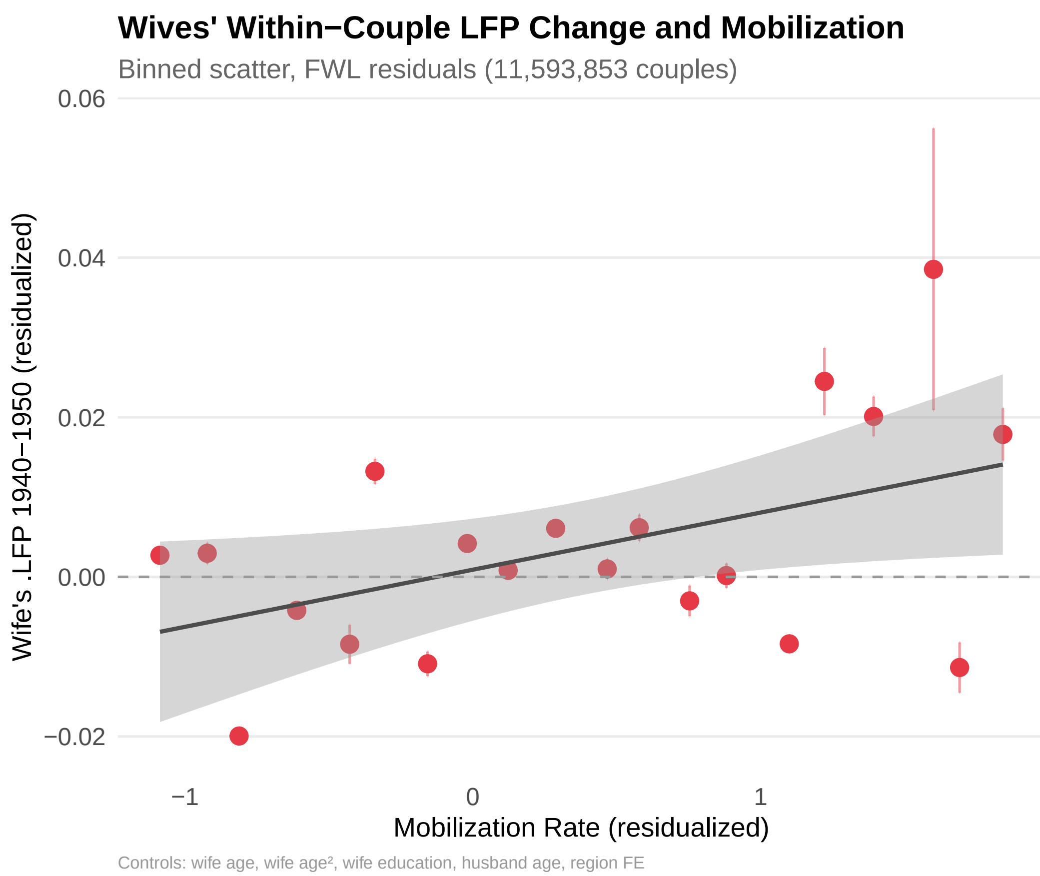

Figure 9: Binned Scatter: Within-Couple ΔLF and Mobilization (Wives)

Notes: FWL residualization on wife's age, age^2, education, husband's age, and region FE. 20 equal-frequency bins. Error bars: 95% CIs. Source: MLP couples panel.

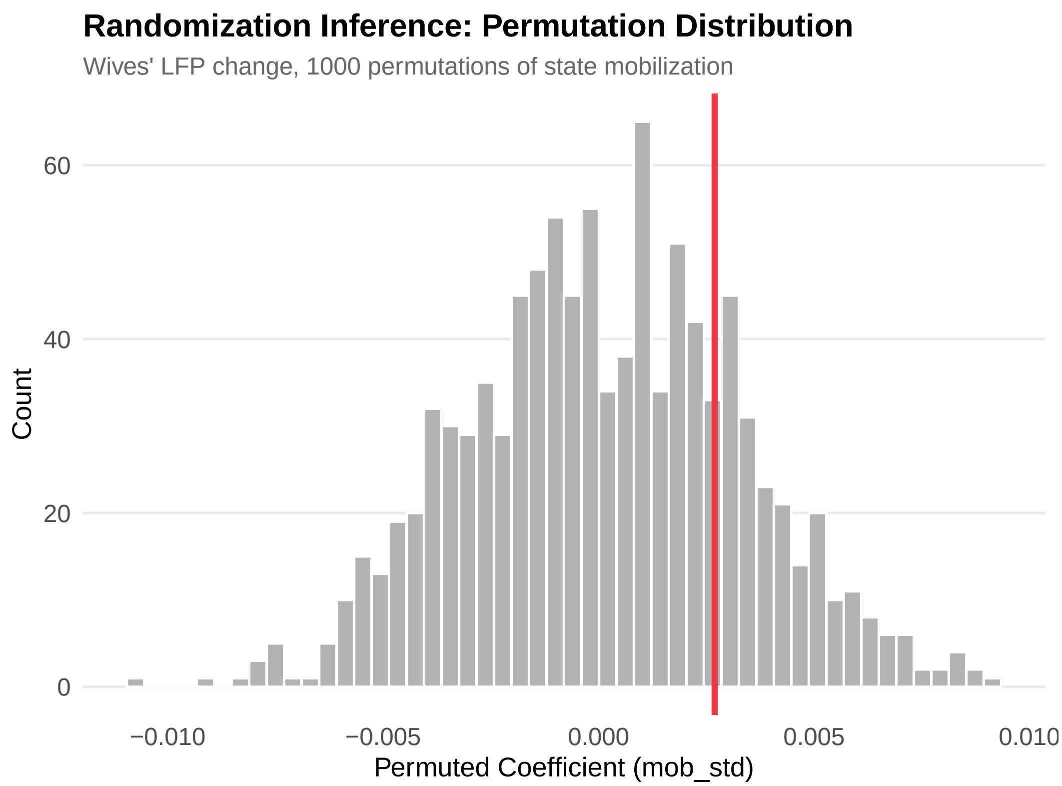

Figure 10: Randomization Inference: Permutation Distribution

Notes: Distribution of $\hat{\beta}$ from 1,000 permutations. Red line = observed. Two-sided p-value = share with $|\hat{\beta}^{\text{perm}}| \geq |\hat{\beta}^{\text{obs}}|$.

7.8 Leave-One-Out Influence

?? displays the leave-one-out distribution of state-level coefficients. No single state exerts disproportionate influence.

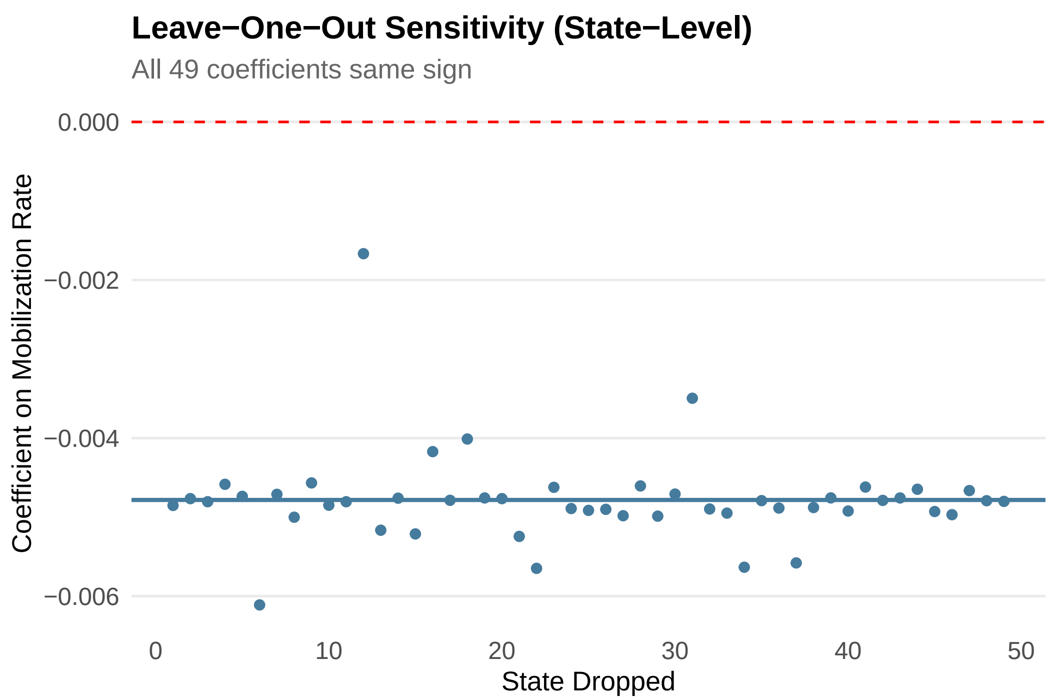

Figure 11: Leave-One-Out Sensitivity (State-Level)

Notes: Each point = coefficient after dropping one state. Blue line = full-sample coefficient. Red dashed = zero.

7.9 Additional Checks

?? summarizes all robustness checks. The core finding—near-zero mobilization effects on both wives' and husbands' LFP—is stable across specifications: older-wife placebo, ANCOVA, trimmed sample, wife identity verification, wild cluster bootstrap, non-mover couples, and IPW reweighting. The leave-one-out range for the state-level specification ($[-0.006, -0.002]$) confirms that no single state drives the state-level result.

7.10 Labor Force Transitions

?? shows the wife-level labor force transition matrix by mobilization intensity. Entry rates (out-of-LF $\rightarrow$ in-LF) exceeded exit rates, with limited variation by mobilization.

?? shows the husband-wife labor force transition rates by mobilization quintile.

Table 10: Robustness Checks Summary

Check	Value
Baseline (wives, region FE)	0.0027 (SE 0.0032)
Pre-trend: Wives LF(1930-1940)	-0.0005 (SE 0.0018)
IPW-Weighted (wives)	0.0026 (SE 0.0031)
State Baseline Controls (w5)	0.0017 (SE 0.0022)
Oster (wives)	-0.11
Randomization Inference (1000 perms)	p = 0.4170
HC3 SE (state-level)	0.0042
Leave-One-Out Range (49 states)	[-0.0061, -0.0017]
Placebo: Older Wives (46+)	0.0023 (SE 0.0030)
ANCOVA (level with lagged DV)	0.0116
Trimmed Sample (5-95%)	-0.0001 (SE 0.0046)
Wife Age-Verified (97.3%)	0.0028 (SE 0.0032)
Wild Cluster Bootstrap	p = 0.4294, CI [-0.0060, 0.0081]
Non-Mover Couples	0.0043 (SE 0.0035)
IPW-Weighted (men)	-0.0009 (SE 0.0012)

Note:
All specifications use state-clustered standard errors except state-level regressions (HC2/HC3). IPW weights rebalance the linked sample to match the full-count census cross-section. Pre-trend test uses 3-wave MLP panel (1930-1940-1950).

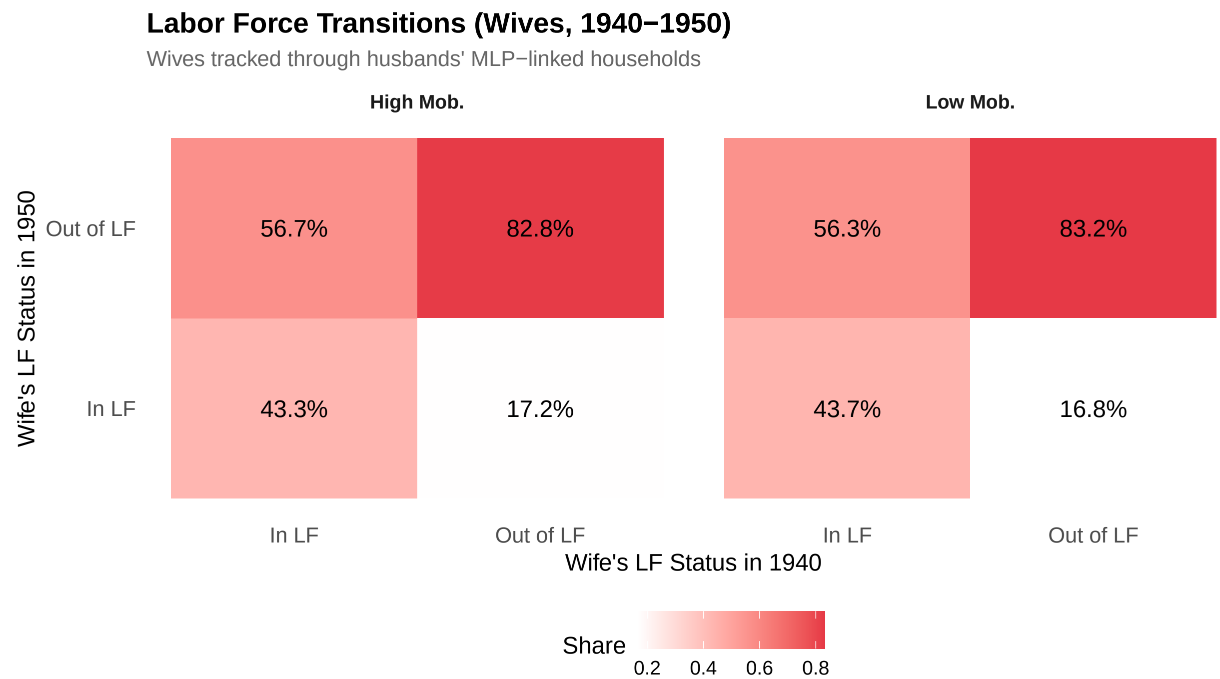

Figure 12: Individual Labor Force Transitions (Wives, 1940–1950)

Notes: Transition rates for MLP-linked wives between 1940 and 1950, by mobilization intensity (above/below median).

8. Mechanisms and Interpretation

8.1 Interpreting the Decomposition

The married-women decomposition in **??** is the paper's primary descriptive contribution. Three features deserve emphasis.

First, within-couple wife LFP gains slightly exceeded the married-women aggregate. The married women tracked longitudinally—those in stable unions with linked husbands—increased their labor force participation by at least as much as the cross-sectional trend for married women in the full-count data.

Second, the point estimate of the compositional residual is negative (-0.0011), suggesting that demographic forces (cohort replacement, mortality, marital transitions) worked modestly *against* convergence rather than driving it. However, this residual is extremely small relative to the 7.5 pp increase and is not statistically distinguishable from zero, so the strongest interpretation is that compositional turnover was approximately neutral.

Third, the direction of the residual is consistent with positive selection into the linked sample. Linked couples represent stable marriages with surviving husbands—a positively selected group. The aggregate trend includes women whose marriages dissolved, whose

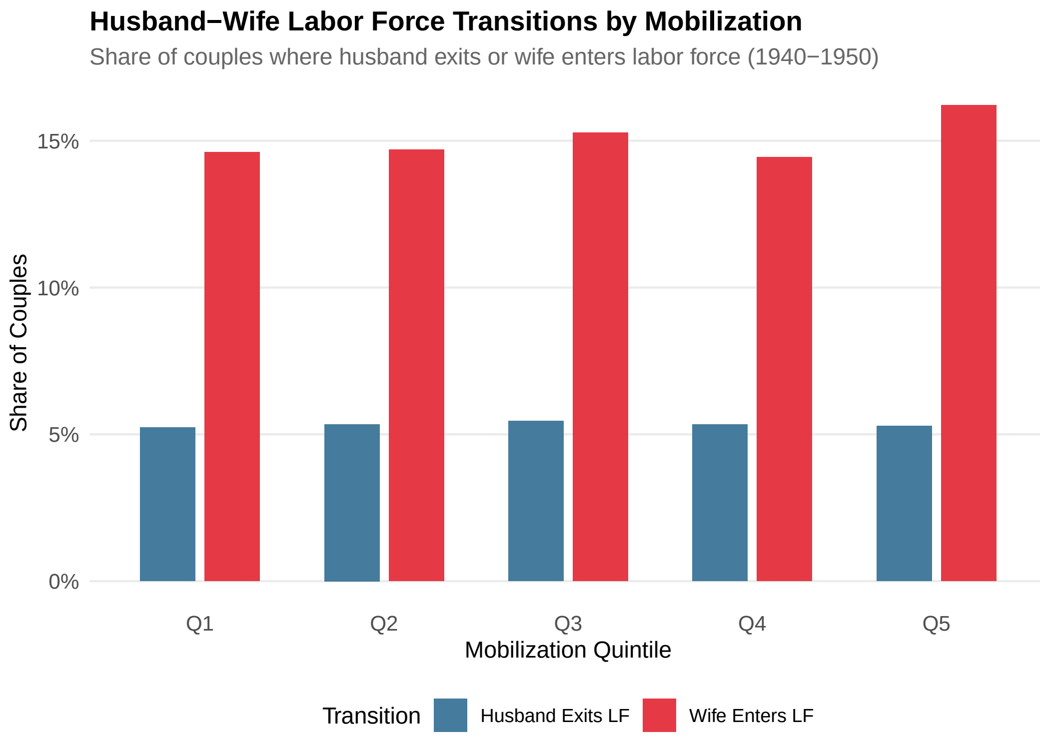

Figure 13: Husband-Wife Labor Force Transitions by Mobilization Quintile

Notes: Share of couples where husband exits LF or wife enters LF between 1940 and 1950, by mobilization quintile.

husbands died, or who entered the working-age population during the decade—groups with potentially different LFP trajectories.

8.2 Why the Mobilization Gradient Is Near Zero

The wives' preferred mobilization coefficient (+0.0027, SE = 0.0032) is statistically insignificant and economically negligible relative to the mean within-couple gain of 7.55 percentage points. Several non-exclusive explanations deserve consideration.

First, the couples panel tracks wives of surviving, linked husbands. If mobilization's strongest effects operated through widowhood or marital disruption, these channels are excluded by construction. Second, by 1950, the acute disruptions of the war had largely resolved; the within-person specification captures the *net* effect, including both the wartime positive shock and any postwar reversal. Third, if the wartime labor demand shock was truly national in scope, cross-state variation in mobilization intensity would be a poor proxy for the relevant treatment, and the null gradient would reflect equilibrium spillovers rather than zero effects. Fourth, the CenSoc mobilization measure may be too noisy to detect state-level gradients, as the weak validation against interstate mover rates suggests. We cannot fully distinguish measurement attenuation from a genuinely diffuse shock with the data at hand.

8.3 Reconciliation with the Cross-Sectional Literature

The within-person evidence both complements and complicates the existing literature. **?** identified a positive aggregate relationship between mobilization and female LFP using Selective Service induction data. Our state-level estimates using CenSoc Army enlistment data show a *negative* relationship (-0.005, $p < 0.10$), possibly reflecting differences in the mobilization measure, sample restrictions (married women only), or the time period considered.

? and **?** showed that wartime and postwar workers were different women. The couples panel confirms this: within-couple gains exceeded aggregate gains, meaning the tracked women gained more than the compositional replacement process would suggest.

8.4 Implications

The decomposition suggests a recalibration of the Rosie the Riveter narrative. Individual women in stable marriages gained substantially in LFP between 1940 and 1950, and these gains show no detectable association with local mobilization intensity. Wives in high-mobilization states experienced similar within-couple LFP changes as wives in low-mobilization states, though we cannot rule out that measurement limitations in our mobilization proxy contribute to this null gradient. The pattern is consistent with broad structural forces—service sector expansion, household technology, shifting norms—operating on married women regardless of local war intensity (**???**).

9. Conclusion

World War II is often cited as the catalyst for the modern female labor market. The image of "Rosie the Riveter"—women entering factories to replace absent men—has become shorthand for a permanent structural break. But the evidence behind this narrative has always rested on aggregate cross-sections that cannot distinguish women changing their behavior from changes in which women are being counted.

This paper follows the same women across three censuses. The central finding is striking in its simplicity: within tracked couples, wives' labor force participation rose by 7.55 percentage points between 1940 and 1950—virtually identical to the 7.44-point married-women aggregate gain. Demographic turnover neither amplified nor dampened the trend. The shift was driven by individual women changing their lives, not by a change in who was being counted.

The mobilization gradient tells a complementary story. The cross-state association between war intensity and wives' LFP changes is near zero (+0.0027, SE = 0.0032), with similarly null effects on husbands (−0.0010, SE = 0.0011). Within couples, husbands' and wives' transitions are negatively correlated (−0.014, $p < 0.01$)—shared vulnerability, not displacement. These patterns are consistent with a nationally pervasive transformation rather than one driven by local mobilization intensity, though measurement limitations in the mobilization proxy contribute to the null gradient.

Three caveats constrain the interpretation. The couples panel conditions on marital survival and husband linkage, creating positive selection. Women who remarried, became household heads, or left their husbands' households are excluded—precisely the women whose wartime experience may have been most disruptive. And the CenSoc mobilization measure captures Army enlistment only; to the extent this proxy is noisy, the estimates are attenuated toward zero.

The broader lesson is methodological. Aggregate trends confound within-person change with compositional turnover, and the direction of the bias is not obvious ex ante. The development of large-scale longitudinal linkage projects opens the possibility of revisiting canonical questions in economic history with individual-level data that previous generations of researchers could not access. The war reshaped American labor markets, but the lasting transformation of women's economic lives was driven by broad structural forces—service sector expansion, household technology, shifting social norms—that operated on all women, not just those in high-mobilization states (??).

A1. Data Construction Details

A1.1 Multigenerational Longitudinal Panel

The MLP version 2.0 (?) uses machine learning methods trained on family history records from FamilySearch to link individuals across U.S. decennial censuses. Unlike the ABE algorithm (?), which requires exact or near-exact name matches within demographic blocks, the MLP's probabilistic approach achieves higher linkage rates (approximately 50% vs 20% for ABE) while maintaining comparable precision.

The full MLP crosswalk contains 175.6 million person-pair records linking individuals across adjacent census decades (1900–1910 through 1940–1950). For this paper, we extract:

- **1940–1950 decade pair:** 71.8 million linked individuals
- **1930–1940–1950 balanced panel:** 43.5 million individuals present in all three censuses

The balanced panel is constructed by intersecting three decade-pair crosswalks (1930–1940 and 1940–1950), retaining only individuals with unique one-to-one-to-one matches across all three censuses. Many-to-many links are dropped.

A1.2 Constructing the Couples Panel

The couples panel tracks wives across census waves through their husbands' households. The key steps:

1. In the 1940 census, identify married men (MARST $\in \{1, 2\}$) who are successfully linked to 1950 via MLP.
2. Locate the wife in the 1940 household using SERIAL and RELATE (spouse of head, or female head with male spouse).
3. Follow the linked man to his 1950 household.
4. Locate the wife in the 1950 household using the same structure variables.
5. Record the wife's outcomes (EMPSTAT, OCC1950, OCCSCORE) in both years.

An age-verification check flags couples as "confirmed" when $|\text{age}_{1950} - \text{age}_{1940} - 10| \leq 2$. Results are robust to restricting the sample to age-verified couples.

A1.3 Variable Construction

- **LFP (1940, 1950):** $\text{in\_lf} = \mathbb{I}[\text{EMPSTAT} \in \{1, 2\}]$
- **LFP (1930):** $\text{in\_lf} = \mathbb{I}[\text{CLASSWKR} > 0]$ (gainful employment)
- **Education years:** Approximate from IPUMS EDUC categorical variable
- **Race:** White (RACE = 1), Black (RACE = 2), Other
- **Mover:** $\text{mover} = \mathbb{I}[\text{STATEFIP}_{1940} \neq \text{STATEFIP}_{1950}]$
- **IPW weights:** Cell-based (state × sex × race × age group), capped at 99th percentile

A1.4 CenSoc Enlistment Records

The CenSoc WWII Army Enlistment Records dataset (**?**) was obtained from the Harvard Dataverse. After restricting to records with valid state identifiers, approximately 2.6 million enlistee records remain. The mobilization rate (**??**) is computed as Army enlistees per 1940 male population aged 18–44.

A2. Comparison with Acemoglu, Autor, and Lyle (2004)

?? highlights the key differences between our approach and the influential repeated cross-section design of **?**. The fundamental distinction is that we track individuals longitudinally rather than comparing population cross-sections, enabling within-person estimation and the decomposition of aggregate trends into within-person and compositional components. Our mobilization measure (CenSoc Army enlistment) differs from their Selective Service induction data, and our pre-trend test exploits the 1930 census wave—unavailable in their framework.

Table A11: Comparison of Estimation Approaches

Feature	This Paper	ACL (2004)
Census data	Full-count (100%)	Full-count (selected years)
Linkage	MLP v2 (**?**)	None (repeated cross-section)
Design	Within-person first-difference	Repeated cross-section
Unit of analysis	Individual (men), Couple (wives)	State
Individual heterogeneity	Absorbed by first-differencing	Not addressed
Mobilization measure	CenSoc Army enlistees	Selective Service inductions
Pre-trend test	Yes (1930–1940)	No
Panel sizes	70M+ men, millions of couples	48 states
Decomposition	Within-person vs compositional	Not available
Inference	State-clustered SE + RI + WCB	State-level robust SE

Table A12: Linkage Balance: MLP Panel vs. Full 1940 Male Population (Ages 18–55)

Sample	Mean Age	% in LF	% Married	% Farm
Full 1940 Male Pop. (18–55)	35.0	93.5	77.4	21.4
MLP-Linked Panel (Men)	34.6	93.0	70.9	22.3

Note:
Linked panel: men matched via MLP v2 crosswalk (Helgertz et al. 2023) across full-count censuses. Full-count aggregates are population-weighted means from the state-level analysis dataset.

Table A13: Mobilization Measure Validation

Check	Value
CenSoc mob vs mover rate ()	-0.0039
CenSoc mob vs mover rate (SE)	0.0092
CenSoc mob vs mover rate (R^2)	0.004
N states	49

Note:
CenSoc mobilization rate: WWII Army enlistees per 1940 male population (ages 18–44). Mover rate: fraction of MLP-linked men who changed state between 1940 and 1950.

A3. Additional Robustness Results

A3.1 Linkage Balance

?? compares the MLP-linked panel to the full 1940 male population on key observable characteristics. The linked panel is slightly younger and less likely to be married, consistent with the MLP's higher linkage rates for individuals with stable names and residences. The similarity on LFP and farm employment rates suggests that the panel is broadly representative of the working-age male population.

A3.2 Mobilization Validation

?? reports the formal regression test of the CenSoc mobilization measure against interstate mover rates. The near-zero coefficient and p-value of 0.84 indicate that these two proxies for wartime disruption are essentially uncorrelated across states, suggesting they capture different dimensions of the mobilization experience.

Acknowledgements

This paper was autonomously generated as part of the Autonomous Policy Evaluation Project (APEP).

Contributors: @SocialCatalystLab

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

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